Characterization of *Mariner* transposons in *Rhus* gall aphids (Hemiptera: Aphididae: Eriosomatinae)

Aftab Ahmad

Shanxi University

Gabriel Luz Wallau

Centro de Pesquisas Aggeu Magalhaes

Zhumei Ren (✉ zmren@sxu.edu.cn)

Shanxi University <https://orcid.org/0000-0003-2813-4871>

Research

Keywords: Rhus gall aphid, Transposons, Mariner-like elements (MLEs), Genome skimming, Phylogeny

Posted Date: April 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-432617/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Transposable elements (TEs), also known as jumping genes, are widely spread in the genomes of insects and play a considerable role in genomic evolution. *Mariner* family belongs to class II transposable elements, were searched in the genomes of seven species of *Rhus* gall aphids belonging to six genera. *Mariner*-like elements were characterized for the first time in *Rhus* gall aphids and classified in to respective subfamilies.

Results: In total, one hundred twenty-one MLEs were detected in the genomes of the seven investigated species of *Rhus* gall aphids, which showed a wide distribution of MLEs in both close and distant related species. The sequences of MLEs ranged from 1kb to 1.4kb in length and the structural analysis of the MLEs showed that only five copies were potentially active with intact open reading frame (ORF) while the remaining were classified as inactive MLEs according to absence of single intact ORF or terminal inverted repeats (TIRs). Based on the MLEs in *Rhus* gall aphids as well as the well characterized MLEs in other organisms from GenBank, the phylogenetic analysis showed that all the one hundred twenty-one MLE sequences belonged to four subfamilies, i.e., thirty from *Maurutiana* subfamily, twenty-six from *Drosophila* subfamily, thirty-three from *Vertumana* subfamily and thirty-two from *Irritans* subfamily, among which *Drosophila* and *Vertumana* subfamilies were reported in aphids for the first time. Moreover, the phylogenetic relationship suggested possible horizontal transfer events of MLEs between aphids and other insects.

Conclusion: Our present report revealed the diversity and distribution of MLEs in *Rhus* gall aphid genomes sequenced by shotgun genome skimming method. This study further expanded our understandings on the characterization of transposable elements in aphid genomes, which might be useful as genetic markers and tools and would play an important role in genomic evolution and adaptation of aphids.

Background

Transposable elements (TEs) are DNA sequences (usually less than 15kb), which have the ability to jump and change its location within the genome, also known as genomic parasites [1, 2]. Once these elements exploit the host cellular machinery for their own replication, they may have a large negative impact on the host fitness [1, 2]. Transposable elements have considerable influence on the evolution of host genome due to their propagation and replication within host genome [3]. However, very small proportion of TE sequences are currently active and most of them are degraded as inactive remains of once active copies [4]. During transposition, they may disrupt coding or regulatory sequences, and the high similar copies, which dispersed in the genome, can serve as source of non-homologous recombination breaking points resulting in chromosomal rearrangement such as inversion, deletion, translocation and duplication. Moreover, TEs can also be co-opted to new host function and give rise to new host genes [3–5], through a phenomenon known as molecular domestication [6]. The influence of TEs on genome organization and evolution is not surprising and enough information are available about impact of TEs in the host genome evolution. The complexity and paraphyletic origin of TEs poses substantial challenges to the scientific community, including the detection, classification, assembly, annotations and mapping of genomic variants [7]. Although the recent advancements to the understanding of TE evolution, there are still considerable gaps of knowledge to completely understand the evolutionary interplay between host and genomic parasites [7,8]. Transposable elements comprised a considerable proportion of eukaryotic and prokaryotic genome [8], e.g., approximately 3–20% of the genomes in many filamentous fungi [9], 10%, 12%, 37%, 45% and 80% of the genome in fish, *Caenorhabditis elegans*, mouse, human and some plants, respectively [2, 3]. The abundance and widespread distribution of transposable elements required a unified classification to divide these sequences into different lineages though it is still a subject of debate [10–12]. There are many difficulties in classification of TEs, one of which is the analysis of the protein sequences of TEs, because some TEs do not possess any coding sequence while some contain many coding regions with different evolutionary histories due to recombination events [11, 12]. Wicker et al. (2007) proposed a unified system to rapidly classify transposable elements.

TEs, are classified into two major classes: Class I or retrotransposons (RTs) and Class II or DNA transposons based on their life cycle and molecular structure [10]. The former is transposed by RNA intermediate while DNA transposons are transposed

by typical cut and paste mechanism [13]. Based on their sequence compositions and some conserved features, TEs can further be divided into subclasses, orders, superfamilies and families [10–13].

Class I transposons are divided into two classes: LTR RTs flanked by long terminal repeats (LTRs) and non-LTR RTs with lacking terminal repeats [13], while Class II elements, or DNA transposons are further classified in to two subclasses: Subclass 1 elements transpose by the process, i.e., excision and integration, in which cleavage of both DNA strands occurs during excision while elements belong to subclass 2, duplicate before insertion. Subclass 1 elements further consist of two orders, in which *Terminal Inverted Repeats (TIR)* order is widely known. *TIR* order further contains nine superfamilies: *Tc1/Mariner*, *hAT*, *Mutator*, *Transib*, *Merlin*, *P*, *CACTA*, *PIF/Harbinger* and *PiggyBac*. Subclass 2 elements consist of two orders: *Maverick* and *Helitron* [6, 10].

Among class II TEs in eukaryotes, *Tc1/Mariner* is one of the most abundant superfamily, whose members share many common characteristics [10]. The autonomous copies contain a single ORF, which encodes a transposase of 282 to 350 amino acid residues with the insertion target TA [14]. Transposase enzyme has a conserved catalytic triad DDE/D motif and a DNA binding domain containing two helix-turn-helix (HTH) motifs [15]. The major characteristics to distinguish the different *Tc1/Mariner* families are their sequence length and DDE/D signature motif. The length of *Tc1/Mariner* ranges from one to five kb due to the length of terminal inverted repeats (TIRs) which varies from 13-34bp in *mariner*, while 20 to 600bp in *Tc1*. The DDE/D signature motif corresponds to DD34E for *Tc1* and DD34D for *Mariner* [15].

Abundant transposable elements were found in different insect genomes, where the proportion of TEs could also explain the variations of insect genome size [16,17]. So far, insect genome analysis revealed that *Belgica antarctica* had the smallest (99Mb) genome with TEs less than 1%, while *Locusta migratoria* (6.5Gb) had the largest one, which consisted of 60% TEs [18]. *Tc1/Mariner* Superfamily of Class II transposons are well characterized in many insect's genomes [14,15,17] but not studied enough in aphid species. *Mariner* like elements (MLEs) of *Tc1/Mariner* superfamily have a simple structure, including single gene flanked by untranslated sequences and terminal inverted repeats (TIRs) at both 5' and 3' ends [19]. *Mariner* transposons were characterized in only a few aphid species in previous studies and very little is known about transposons abundance, diversification and influence on genomic evolution in aphids. Partial sequences of *Irritants* and *Melifera* subfamilies in *Aphis glycine* were identified in vitro by PCR amplification [16, 18], and deleted copies of *Mauritiana* subfamily were detected in seven fruit tree aphid species [19], whereas, only three complete sequences were reported in pea aphid *Acyrtosiphon pisum* deposited in Repbase as *MarinerN-1_AP*, *1B* and *2* [20]. However, many lineages of *Mariner* were detected recently in the genomes of three aphid species: *Acyrtosiphon pisum*, *Diuraphis noxia*, and *Myzus persicae* [18], whose genomes are available at NCBI (<http://www.ncbi.nlm.nih.gov/genbank>) and in aphid database (<http://tools.genouest.org/is/aphidbase>) respectively and both complete and truncated copies of TEs were detected in them from different families of *Tc1/Mariner* superfamily [18, 20]. As, *Mariner* transposons was characterized in only a few aphid species and very little is known about its abundance, diversification and influence on genomic evolution in aphids. In this study, we examined *Mariner* family of *Tc1/Mariner* superfamily of Class II transposons in the genomes of seven species of *Rhus* gall aphids from six genera.

Rhus gall aphids (Aphididae: Eriosomatinae: Fordini) include six genera, in which five genera are from east Asia while one from east North America, and specially comprise a unique group [21, 22]. *Rhus* gall aphids feed on their primary host plant *Rhus* species (Anacardiceae) to form galls with rich tannins, which were produced as an important medical and industry raw material [21, 22, 24, 49]. There were some reports on the phylogenetic relationship of the *Rhus* gall aphids referring from different molecular sequence markers [21, 22, 36]. However, these studies have been largely constrained by limited sampling or have failed to find high support for relationships among the genera and species. Recently, Ren et al. [2019] investigated the evolutionary relationships within *Rhus* gall aphids by sampling 15 accessions representing all six genera and using 20 gene regions: five nuclear genes as well as 13 protein-coding genes and two rRNA genes of the complete mitochondrial genome, which obtained the backbone phylogeny to well support the monophyly of six genera and resolve the relationship of genera and species from *Rhus* gall aphids [22].

In case of the seven species in this study, their relationship was as following: The North America species *Melaphis* was original in East Asia; *Meitanaphis* is sister to *Kaburagia*, and then grouped with *Floraphis*; *Nurudea ibofushi* is nested in *Schlechtendalia* and suggested to be merged in the genus *Schlechtendalia*. As transposable elements may serve as genetic markers and tools and have impact on insect genome, adaptation and biology [23, 24], we are interested in detecting and characterizing *mariner*-like transposons from at least one *Rhus* gall aphid species from all the six genera, i.e., *Schlechtendalia*, *Nurudea*, *Melaphis*, *Meitanaphis*, *Kaburagia* and *Floraphis* known to feed on *Rhus* species. To our knowledge, this study would represent the first report on the *mariner* transposable elements and its implications in *Rhus* gall aphids.

Results

Search of homologous sequences, belonging to *Mariner* family, was performed by the approach (tBLASTn) using a set of 50 known MLEs as queries (see additional file 1). We found in total 121 sequences of MLEs in all the seven *Rhus* gall aphid species, i.e., thirty-three in *Schlechtendalia chinensis*, twenty-six in *Schlechtendalia peitan*, ten in *Kaburagia rhusicola*, ten in *Floraphis choui*, ten in *Meitanaphis flavogallis*, sixteen in *Melaphis rhois* and sixteen in *Nurudea ibofushi*. All the detected transposons were classified into four subfamilies of *Mariner* family transposable elements based on the phylogenetic analysis with already classified MLEs from previous studies (see Fig. 1a and 1b). The numbers and the classifications of MLEs detected in all seven *Rhus* gall aphid's species are shown in Table 1.

General features

Fifteen of the extracted MLEs were truncated at both or one end among the total 121 detected ones, which were mostly due to their presence at the end of contigs (see Tables in supplemental file 2). Only five complete copies of MLEs with intact ORF and two truncated copies with complete intact ORF for transposase protein were detected in the study. Sequences of intact ORF with no stop codon or frameshift mutation were considered potentially active [18]. Among 33 MLEs detected in *Schlechtendalia chinensis*, only two (*Scmar7* and *Scmar10*) were found to have intact ORF for transposase but truncated with missing TIRs at one end. While three MLEs from *Kaburagia rhusicola* (*Krmar2*, *Krmar4* and *Krmar5*), one from *Floraphis choui* (*Fcmar4*) and *Meitanaphis flavogallis* (*Mfmar4*), respectively, were potentially active with intact ORF and TIRs at both 5' and 3' end. All the other sequences had at least one or more premature stop codons. No active copy with single intact ORF was detected in *Schlechtendalia peitan*, *Nurudea ibofushi* and *Melaphis rhois*. Five of the MLEs with intact transposase ORF belonged to the subfamily *Drosophila*, i.e., *Krmar2*, *Scmar7*, *Scmar10*, *Mfmar4* and *Fcmar4*, while two belonged to *Maurutiana*, i.e., *Krmar4* and *Krmar5* (see Tables in Additional file 2). All the MLEs, belonging to *Vertumana* and *Irritans* subfamilies, were inactive with no intact ORF for transposase. All the detected MLEs in the *Rhus* gall aphids have been submitted to GenBank with accession numbers (see Table 1–4 in Additional file 2).

Structure analysis

Terminal inverted repeats (*TIRs*), which are necessary for the transposition, were analyzed in all the complete copies of MLEs and consensus *TIRs* of each subfamily are also given (see Table 1–4 additional file 2). Meanwhile, the TA target site duplication (*TSD*) were also found at both ends in the complete copies except *Krmar8* and *Krmar10* in which TA were found at 3' end only. All the complete copies detected were of variable length ranging from 1.2kb to 1.35kb and *TIRs* from 13bp to 32bp (see Table 1, and additional file 2).

Transposases of the complete MLEs were analyzed for the conserved domains and motifs of mariner transposons. Catalytic domain DD34D were highly conserved in most of the complete copies, while WVPHEL and YSPDL motif required for transposition were slightly modified in some MLEs. Helix-turn-helix DNA binding motifs were also conserved and found in all the complete copies. Nuclear localization sequence (NLS) was also present in some complete copies while absent or modified in others. Some of the detected MLEs became inactive due to presence of only single point mutation (single nucleotide substitution) which led to generate premature stop codon (see Fig. 4). Conserved catalytic domain DD34D, helix-turn-helix

(HTH) DNA binding motifs, WVPHEL motif, YSPDL motif and nuclear localization signal (NLS) of three of the complete MLEs detected in the study belonged to three different subfamilies of MLEs, which are shown in Figs. 2, 3 and 4, respectively.

Phylogenetic analysis

The subfamily classification of detected MLEs in *Rhus* gall aphids was done on the basis of DNA sequence similarities. *Mariner* sequences of other organisms mainly from the class *insecta* were downloaded from GenBank, which belonged to already reported major subfamilies, i.e., *Maurutiana*, *Mellifera*, *Irritans*, *Cecropia*, *Capitata*, *Vertumana*, *Drosophila*, *Vertuman*, *Marmoratus*, *Lineata* and *Elegans*. Phylogenetic relationship of all the 121 MLEs in *Rhus* gall aphids along with the MLE sequences of other organisms mainly insects from GenBank were analyzed by constructing ML phylogenetic tree with 1000 duplicates. All the detected MLEs of *Rhus* gall aphids were clustered into four subfamilies, i.e., *Maurutiana*, *Irritans*, *Vertumana* and the subfamily *Drosophila* (see Fig. 1a and 1b). The MLEs detected in this study were classified into subfamilies according to their groups and relatedness with already known MLEs from different subfamilies downloaded from GenBank (see Fig. 1a and 1b).

Discussion

The seven *Rhus* gall aphid species sampled in this study feed on the primary host plant *Rhus* species to form galls containing rich tannin, so they have many economic importance to be widely applied in various fields, e.g., medicine, food, dye, chemical and military industry [21, 22]. The sequences of MLEs in *Rhus* gall aphids were searched using BLAST in the aphid whole genomes sequenced by shotgun genome skimming method. Our study focused only on the existence of *mariner*-like elements in *Rhus* gall aphid species, and were irrespective of their total copies, number and percent contribution in the genome size. As a result, *Mariner* transposons from four different subfamilies, i.e., *Maurutiana*, *Vertumana*, *Irritans* and *Drosophila* were detected in all the seven species of *Rhus* gall aphids (see Table 1), which is in agreement with previous studies, i.e., MLEs were shown to have the widespread distribution of in Hexapoda [5, 18, 29, 30].

A total of 121 MLEs were detected in all the seven aphid species, among which the subfamily *Vetumana*, and *Mauritiana* were the most widely spread one in all the seven *Rhus* gall aphids, as the previous studies also reported the presence of MLEs from *Mauritiana* family in seven tree aphids [20], while no MLEs from *Maurutiana* were found in the genome mining of *Aphis glycine* [4], *Acyrtosipon pisum*, *Diuraphis noxia* [18], however, MLEs from *Vertumana* Subfamily were reported for the first time in aphids during this study. MLEs from *Drosophila* subfamily were also detected in all the *Rhus* gall aphid species in this study, while it was not reported in other aphid species in previous studies. Thus, we also reported this for the first time in *Rhus* gall aphids, while it was discovered for the first time in *Drosophila* species [32]. MLEs from the subfamily *Irritans* were found in three species of *Rhus* gall aphids, i.e., *Schlechtendalia chinensis*, *Schlechtendalia peitan* and *Nurudea ibofushi*, while absent in the other four studied species (see Table 1). In contrast, MLEs from the subfamily *Irritans* were found in all aphid genomes in previous studies [4, 18]. Absence of MLEs from subfamilies like *Mellifera*, *Capitata* and others might indicate variable distributions of MLEs in aphid genomes, or might be related to the fact that our sequenced genomes didn't cover the 100% genes and repeat regions of aphid species in the study.

A fewer of complete copies of each MLE, i.e., 1 to 3 (see Table 1–4 in additional file 2) was detected in this study as comparing to previously proposed studies [5]. Most of the MLEs detected previously in aphids were in Vitro by PCR cloning, which resulted in detection of a relatively large number of deleted copies of MLEs [20]. Our study mainly focused on the detection of complete copies of *Mariner*-like elements in *Rhus* gall aphids and very few truncated copies were detected and reported in this study in contrast to previous studies which reported mostly internally deleted and truncated MLEs mostly less than 1000bp in aphids [4, 18]. No Miniature Inverted-repeats transposable elements (MITEs) were detected in this study which were previously reported in aphids [4,18].

The relatively low number of different MLEs in aphid genomes from only four subfamilies in our study as compare to other insects agreed with the previous studies [20], which might be the special genetic characteristics of the aphids including the *Rhus* gall aphids. Also, this might be due to (i) the genome size sequenced in our study didn't completely cover the repeated regions in genome due to the sequencing Illumina platform [31]; (ii) more than 60% of the assembled contigs were < 1000 bp long (see table no.2), which didn't result in producing good hits by tBLASTn search in the genomes. Though *Tc1/mariner* is the most abundant superfamily in insect genome, it is poorly represented in aphid genomes [4,5], which was also supported by our study.

Structural analysis of the protein polypeptides of the detected MLEs in all the seven aphid species showed that the conserved catalytic domains DD34D in the third aspartate residue were mutated in many of the inactive copies, while highly conserved in active copies, which also supported the previous studies [33]. DNA binding helix-turn-helix HTH motif and two main conserved domains of MLEs, i.e., WVPHEL and YSPDL, required for transposase activity, which were conserved in most of the MLEs, whereas there was slight modification in the conserved regions in some of the MLEs as shown in Fig. 2, 3, 4 which were in agreement with the previous finding [20]. Nuclear localization sequence (NLS) motif, being required for the import of transposase to the nucleus, were analyzed and found in some active MLEs (Fig. 2), and were slightly modified in some (Fig. 3, 4) while absent in many sequences due to frequent mutation or inactivation events [20]. However, the previous studies also showed that some of the MLEs didn't have their own NLS, which depended on other proteins for their nuclear importation [34].

The current study showed the diversity of MLEs in aphid genomes, but most of the detected MLEs corresponded to inactive lineages, which was in agreement with previous findings [4, 18]. The absence of very few potentially active copies supported the phenomenon of vertical inactivation of *Mariner* transposons [5, 20]. Single nucleotide substitution which leads to premature stop codon (see Fig. 4) and nucleotides loss due to deletions reported in previous studies [2] appeared to play an important role in vertical inactivation of transposons, e.g., *Irritans* subfamily, had no active MLE copy, i.e., all the copies were inactive with no intact ORF though, they are widely spread in our studied species and in previously studied species [3, 18].

Like all other genes, MLEs are transmitted vertically from parents to offspring during the evolutionary course, so the relationship between MLEs sequences must reflect the evolutionary relationship of their hosts [35, 36]. Phylogenetic relationship of aphids based on the mitochondrial COI gene showed the consistency with the classical phylogenetic analysis based on molecular and morphological characteristics in previous studies [35–39], which proved COI to be effective tool for characterizing aphid species phylogeny [40]. While other studies reported significant inconsistency of transposable elements with their molecular phylogeny as compared to COI [20] and also with other single non-transposable genes from the same genome [41–44]. Our study also indicated that the phylogeny of MLEs was non concordant while previous studies also showed that MLEs had evolved independently of host speciation event [45].

Horizontal transfer events of MLEs within aphid species and with other insects was obvious in our sequence analysis. The presence of almost identical MLEs in phylogenetically distant-related species and obviously variable MLEs within the same species in this study reflected the occurrence of horizontal transfer events during aphid evolution, which supported the previous findings of horizontal transfer of transposons in other insects [43,47,48]. For instance, MLE from *Bactrocera tryoni* *Batmar11* (accession no. KX931004) had more than 90% DNA sequence similarity with *Scmar11*, *Scmar12*, *Spmar2*, *Krmar5*, *Krmar6*, and *Mfmar7*, which supported the occurrence of one or more horizontal transfer events between *Rhus* gall aphids and fruit fly (*Bactrocera tryoni*) as the previous studies revealed that TEs from distantly related species having sequence query coverage more than 90% and similarity more than 90% indicates the event of horizontal transfer [51]. While one MLEs of *irritans* subfamily (*Scmar15*) detected *Schlechtendalia chinensis* was 100% similar throughout the length to MLE (*Spmar5*) detected in *Schlechtendalia peitans*, which is a clear evident of horizontal transfer of transposon (HTT), though these both are sister species belongs to same genus *Schlechtendalia* but similar HTT were reported in different species of drosophila belong to same genus [52]. However, the detailed explanation of horizontal transfer is not discussed in this study, and we will thoroughly examine and explain this phenomenon by sampling more species and more MLEs from different subfamilies in our further research.

Conclusion

Our present study reported the diversity and structure composition of *mariner*-like transposons in *Rhus* gall aphids. All the 121 *mariner*-like elements (MLEs) belonged to four subfamilies: *Maurutiana*, *Drosophila*, *Irritans* and *Vertumana*, among which subfamily *Drosophila* and *Vertumana* were reported for the first time in *Rhus* gall aphid species. We only demonstrated the presence of full length MLEs including both the active and inactive lineages in aphid species. Further research needs to demonstrate the activity of potentially active MLEs and its transposition in *Rhus* gall aphids and its role in genome evolution and adaptations.

Methods

All the aphid genomes used in this study were sequenced by shotgun genome skimming method by an already ongoing project in our lab.

Sample collections

All the mature *Rhus* galls were collected on the host plant from different location in China except one species which was collected in North America [22]. There were thousands of aphids in one gall because of the parthenogenetic generations during the gall formation. Some individuals from one gall were placed in 75% alcohol for taxonomic identification using microscopy by following taxonomy protocol [24]. The remaining individuals were preserved in absolute alcohol for DNA extraction. Voucher specimen were deposited at the School of Life Sciences in Shanxi University, China. Sampling information and species taxonomy are shown in Table 2.

DNA extraction and Sequencing

Three individuals of the aphid samples stored in absolute alcohol were transferred into distilled water for thirty-six hours in 1.5ml Eppendorf tube, and then the water was removed and the aphids were grounded with the help of a small pestle. Genomic DNA of all samples were extracted using DNeasy extraction kit (QIAGEN, Valencia, CA), and the qualified DNAs were sent to the Genomic Sequencing and Analysis Facility (GSAF), University of Texas, Austin for library construction and next generation sequencing (NGS). A TruSeq Nano DNA library preparation kit (Illumina, FC-121-4003) was used to prepare DNA library and the Illumina NextSeq sequencer was used for the generation of paired-end reads 2x150 bp with an insert size of 400 bp. Trimmomatic v.0.35 was used to filter raw data with default settings [25]. *De novo* assembly of the trimmed data was performed by the program Spades v. 3.7.1 [50] and the whole genome was assembled into contigs with different length. Genome size, GC content and detailed information of the contigs of all the seven *Rhus* gall aphid species were shown in Table 2.

Data mining

Panel of complete copies of both active and non-active *mariner* transposable elements were downloaded from GenBank (<http://www.ncbi.nlm.nih.gov/GenBank>, see the Additional file 3). Most the downloaded sequences belonged to the class *Insecta*, mainly *Drosophila* and also the MLEs already reported in other species of aphids. Geneious prime 11.0.3 with default parameters was used for mining the transposable elements using the downloaded sequences as query by local BLASTn searches on genomic contigs of each species. Detailed information of the sequenced genomes of the seven *Rhus* gall aphid species are given in Table 2. The sequences with the best hits (similarity more than 60% and query coverage more than 60%) were extracted and manually analyzed for MLE signatures and terminal inverted repeats (*TIRs*). Each of the complete sequences extracted were used again as query to retrieve more similar sequences. Truncated sequences with similarity less than 60% and query coverage less than 60% were manually analyzed and were not included and reported in this study due to absence of *TIRs* and any MLEs signatures. No MITEs (Miniature Inverted-repeats Transposable Elements) were retrieved

during this study and most of the truncated and deleted copies retrieved with length less than 1kb were the exact copies of that complete MLEs.

Sequence analysis and identification

All the *Mariner* sequences extracted from each local database of genomic contigs were manually analyzed for its terminal inverted repeats (TIRs) and target site duplications (TSD). Potentially active and non-active copies from the sequences were determined by translating the sequences for transposase using ORF finder implemented in Geneious prime 11.0.3 by default setting. DDD/E catalytic domain and HTH DNA binding conserved domains were analyzed for potentially active and non-active copies of MLEs by NCBI conserved domain search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) [26] with default parameters, while nuclear localization sequence (NLS) motif for active copies of transposase was searched by cNLS mapper (<http://nls-mapper.iab.keio.ac.jp/cgi>) [27]. Multiple alignment was done using MAFFT version implemented in Geneious 11.0.3 with default parameters for the analysis of the conserved DDE/D signature in the transposase for potentially active copies.

ORF and conserved domains of the MLEs

The analysis of the potentially active copies with ORF ranging from 310 to 345 amino acids was performed by aligning them with transposases of *Mariner* family of other organisms downloaded from GenBank. Complete structure composition of the transposable elements, i.e., DNA binding domain (HTH), nuclear localizing motif (NLS) and catalytic domain DD34D of active and inactive copies from each species, was predicted, and the sequences having intact ORF with no stop codon or frameshift mutation were considered active [18]. Conserved catalytic domains DD34D were used to justify the classification of detected TEs into *Mariner* family of *Tc1/Mariner* superfamily. MLEs with no intact ORF and having one or more than one stop codon were also translated and analyzed for conserved domain and motifs, i.e., DD34D catalytic domain, HTH motif, nuclear localization motif, WVPHEL and YSPDL motif. MLEs having no intact ORF and conserved motif and domains due mutations like deletion, insertion or substitution were also classified in the same group based on sequence similarity $\geq 80\%$ in the complete sequence or TIRs as proposed by Wicker et al. 2007.

Phylogenetic analysis

The phylogenetic analysis was performed in order to reveal the relationship of MLEs in *Rhus* gall aphids and other insects. All the sequences detected in *Rhus* gall aphids along with forty-four MLEs of other organisms mainly from the class *Insecta* downloaded from GenBank were used to construct the phylogenetic tree (see additional file 3). As not all the detected MLEs had intact ORF in our study, so their whole nucleotide sequences of MLEs were used to align using MAFFT multiple alignment implemented in Geneious 11.0.3 with default parameters and construct the phylogram of 121 MLEs detected in *Rhus* gall aphids and other 44 MLEs from GenBank. The Maximum likelihood (ML) phylogenetic tree were constructed by GTRGAMMA model with 1000 replications (bootstraps) using the software RAXML [28]. An MLE from the *Drosophil virilis* (Accession no. DVU26938) belonging to *Tc1* family was used as outgroup.

Declarations

Ethics approval and consent to participate

“Not applicable”

Consent of publication

“Not applicable”

Availability of data and material

The genomic data (sequenced genomes) of species used and generated in this study are not publicly available yet and deposited in the School of life Sciences, Shanxi University under voucher numbers mentioned in the manuscript. While the *Mariner* transposons used and detected are included in this article and its supplementary files.

Competing Interest

The authors declared they have no competing interests.

Funding

This study was partially supported by The National Natural Science Foundation of China (31870366), Shanxi International Science and Technology Cooperation Project (201803D421051), Research Project Supported by Shanxi Scholarship Council of China (2020-018), the National High Technology Research and Development “863” Program (2014AA021802).

Author's contributions

AA and ZR designed the study, interpreted all the data and findings and write the manuscript, and have equal contributions as major authors, while GLW reviewed and edited the final manuscript. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Zhumei Ren

Acknowledgements

Not applicable

Abbreviations

TEs

Transposable elements

MLES

Mariner-like elements

TIRs

Terminal Inverted Repeats

HTH

Helix-turn-helix

NLS

Nuclear localization signal

SC

Schlechtendalia chinensis

SP

Schlechtendalia peitan

MF

Meithanaphis flavogallis

KR

Kaburagia rhusicola

NI

Nurudea ibofushi

FC

Floraphis choui

References

1. Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*. 1994;371(6494):215 – 20.
2. Brunet F, Giraud T, Godin F, Capy P. Do deletions of Mos1-like elements occur randomly in the Drosophilidae family?. *Journal of molecular evolution*. 2002;54(2):227 – 34.
3. Dechaud C, Volff JN, Scharl M, Naville M. Sex and the TEs: transposable elements in sexual development and function in animals. *Mobile DNA*. 2019;10(1):42.
4. Mittapalli O, Rivera-Vega L, Bhandary B, Bautista MA, Mamidala P, Michel AP, Shukle RH, Mian MA. Cloning and characterization of *mariner*-like elements in the soybean aphid, *Aphis glycines* Matsumura. *Bulletin of Entomological Research*. 2011;(101):697–704.
5. Petersen M, Armisen D, Gibbs RA, Hering L, Khila A, Mayer G, Richards S, Niehuis O, Misof B. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC evolutionary biology*. 2019;19(1):11.
6. Gao B, Wang Y, Diaby M, Zong W, Shen D, Wang S, Chen C, Wang X, Song C. Evolution of pogo, a separate superfamily of IS630-Tc1-*mariner* transposons, revealing recurrent domestication events in vertebrates. *Mobile DNA*. 2020;11(1):1–5.
7. Makiowski W, Gotea V, Pande A, Makiowska I. Transposable elements: Classification, identification, and their use as a tool for comparative genomics. *Evolutionary Genomics 2019* (pp. 177–207). Humana, New York, NY.
8. Feschotte C, Zhang X, Wessler SR, Craig NL, Craigie R, Gellert M, Lambowitz AM. *Mobile DNA II. Miniature inverted-repeat transposable elements and their relationship to established DNA transposons*. Edited by Craig NL, Craigie R, Gellert M, Lambowitz AM. 2002;1147:1158.
9. Marini MM, Zanforlin T, Santos PC, Barros RR, Guerra AC, Puccia R, Felipe MS, Brigido M, Soares CM, Ruiz JC, Silveira JF. Identification and characterization of Tc1/*mariner*-like DNA transposons in genomes of the pathogenic fungi of the *Paracoccidioides* species complex. *BMC genomics*. 2010;11(1):130.
10. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*. 2007;8(12):973 – 82.
11. Kapitonov VV, Jurka J. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics*. 2008;9(5):411-2.
12. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E. Reply: A unified classification system for eukaryotic transposable elements should reflect their phylogeny. *Nature Reviews Genetics*. 2009;10(4):276–276.
13. Rouault JD, Casse N, Chénais B, Hua-Van A, Filée J, Capy P. Automatic classification within families of transposable elements: application to the mariner Family. *Gene*. 2009; 448(2):227 – 32.
14. Plasterk RH, Izsvák Z, Ivics Z. Resident aliens: the Tc1/*mariner* superfamily of transposable elements. *Trends in Genetics*. 1999;15(8):326 – 32.
15. Shao H, Tu Z. Expanding the diversity of the IS630-Tc1-*mariner* superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. *Genetics*. 2001;159(3):1103-15.

16. Chalopin D, Naville M, Plard F, Galiana D, Volff JN. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome biology and evolution*. 2015;7(2):567 – 80.
17. Mérel V, Boulesteix M, Fablet M, Vieira C. Transposable elements in *Drosophila*. *Mobile DNA*. 2020;11(1):1–20.
18. Bouallègue M, Filée J, Kharrat I, Mezghani-Khemakhem M, Rouault JD, Makni M, Capy P. Diversity and evolution of *mariner*-like elements in aphid genomes. *BMC genomics*. 2017;18(1):494.
19. Kharrat I, Mezghani M, Casse N, Denis F, Caruso A, Makni H, Capy P, Rouault JD, Chénais B, Makni M. Characterization of *mariner*-like transposons of the mauritiana subfamily in seven tree aphid species. *Genetica*. 2015;143(1):63–72.
20. Jurka J. *Mariner* families from *Acyrtosiphon pisum*. *Rebase Reports*. 2008; 8(3):340.
21. Ren Z, Harris AJ, Dikow RB, Ma E, Zhong Y, Wen J. Another look at the phylogenetic relationships and intercontinental biogeography of eastern Asian–North American *Rhus* gall aphids (Hemiptera: Aphididae: Eriosomatinae): Evidence from mitogenome sequences via genome skimming. *Molecular phylogenetics and evolution*. 2017;117:102 – 10.
22. Ren Z, von Dohlen CD, Harris AJ, Dikow RB, Su X, Wen J. Congruent phylogenetic relationships of Melaphidina aphids (Aphididae: Eriosomatinae: Fordini) according to nuclear and mitochondrial DNA data with taxonomic implications on generic limits. *PloS one*. 2019;14(2):e0213181.
23. Skipper KA, Andersen PR, Sharma N, Mikkelsen JG. DNA transposon-based gene vehicles-scenes from an evolutionary drive. *Journal of biomedical science*. 2013;20(1):1–23.
24. Zhang GX, Qiao GX, Zhong TS, Zhang WY. Fauna Sinica Insecta 14. ♂274. In Homoptera: Mindaridae and Pemphigidae. Science Press, Beijing, China. 1999:256.
25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-20.
26. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic acids research*. 2017;45(D1):D200-3.
27. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proceedings of the National Academy of Sciences*. 2009;106(25):10171-6.
28. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312-3.
29. Yamada K, Yamada A, Kawanishi Y, Gurung RD, Sasaki T, Tokuda G, Maekawa H. Widespread distribution and evolutionary patterns of *mariner*-like elements among various spiders and insects. *Journal of Insect Biotechnology and Sericology*. 2015;84(2):2_029–41.
30. Quesneville H. Impact of transposable elements on insect genomes and biology. *Current opinion in insect science*. 2015;7:1–7.
31. Dunn P, Albury CL, Maksemous N, Benton MC, Sutherland HG, Smith RA, Haupt LM, Griffiths LR. Next generation sequencing methods for diagnosis of epilepsy syndromes. *Frontiers in Genetics*. 2018;9:20.
32. Wallau GL, 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):727.

33. Doak TG, Doerder FP, Jahn CL, Herrick G. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. *Proceedings of the National Academy of Sciences*. 1994;91(3):942-6.
34. Demattei MV, Hedhili S, Sinzelle L, Bressac C, Casteret S, Moiré N, Cambefort J, Thomas X, Pollet N, Gantet P, Bigot Y. Nuclear importation of *mariner* transposases among eukaryotes: motif requirements and homo-protein interactions. *PLoS One*. 2011;6(8):e23693.
35. Foottit RG, Maw HV, Von Dohlen CD, Hebert PD. Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. *Molecular Ecology Resources*. 2008;6(6):1189 – 201.
36. Yang ZX, Chen XM, Havill NP, Feng Y, Chen H. Phylogeny of *Rhus* gall aphids (Hemiptera: Pemphigidae) based on combined molecular analysis of nuclear EF-1a and mitochondrial COII genes. *Entomol. Sci.* 2010;13(3):351–357.
37. Kim H, Lee S, Jang Y. Macroevolutionary patterns in the Aphidini aphids (Hemiptera: Aphididae): diversification, host association, and biogeographic origins. *PLoS One*. 2011;6(9):e24749.
38. Blackman RL, Eastop VF. Taxonomic issues. Aphids as crop pests. 2007;1:1–29.
39. Leclant F. Aphids of cultivated crops: identification keys. III Fruit crops. Aphids of cultivated crops: identification keys. III Fruit crops. 2000.
40. Park DS, Foottit R, Maw E, Hebert PD. Barcoding bugs: DNA-based identification of the true bugs (Insecta: Hemiptera: Heteroptera). *Plos one*. 2011;6(4):e18749.
41. Lawrence JG, Hartl DL. Inference of horizontal genetic transfer from molecular data: an approach using the bootstrap. *Genetics*. 1992;131(3):753 – 60.
42. Clark JB, Maddison WP, Kidwell MG. Phylogenetic analysis supports horizontal transfer of P transposable elements. *Molecular biology and evolution*. 1994;11(1):40– 50.
43. Robertson HM, Lampe DJ. Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. *Molecular Biology and Evolution*. 1995;12(5):850 – 62.
44. Capy P, Anxolabéhère D, Langin T. The strange phylogenies of transposable elements: are horizontal transfers the only explanation? *Trends in Genetics*. 1994;10(1):7–12.
45. Zhou MB, Zhong H, Tang DQ. Isolation and characterization of seventy-nine full-length *mariner*-like transposase genes in the Bambusoideae subfamily. *Journal of plant research*. 2011;124(5):607 – 17.
46. Dupeyron M, Leclercq S, Cerveau N, Bouchon D, Gilbert C. Horizontal transfer of transposons between and within crustaceans and insects. *Mobile DNA*. 2014;5(1):4.
47. Han G, Zhang N, Xu J, Jiang H, Ji C, Zhang Z, Song Q, Stanley D, Fang J, Wang J. Characterization of a novel Helitron family in insect genomes: insights into classification, evolution and horizontal transfer. *Mobile DNA*. 2019;10(1):25.
48. Grace CA, Carr M. The evolutionary history of *mariner* elements in stalk-eyed flies reveals the horizontal transfer of transposons from insects into the genome of the cnidarian *Hydra vulgaris*. *PloS one*. 2020;15(7):e0235984.
49. Favret C. Aphid Species File. Version 5.0/5.0. URL: <http://Aphid.SpeciesFile.org>.
50. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–477.

51. Wallau GL, 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*. *Molecular biology and evolution*. 2016;1(4):1094 – 109.
52. Bartolomé C, Bello X, Maside X. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome biology*. 2009;10(2):1 – 1.

Tables

Table 1. The total number of MLEs in all the seven *Rhus* gall aphid species, number of MLEs detected in each species and its classifications into subfamilies of MLEs are shown. Note: A/I- Potentially Active/Inactive, C/T-Complete copies/Truncated copies.

Species	No. of MLEs	MLEs Subfamily				Length(kb)	A/ I	TIRs(bp)	C/T
		<i>Maurutiana</i>	<i>Drosophila</i>	<i>Vertumana</i>	<i>Irritans</i>				
<i>Schlechtendalia chinensis</i>	33	4	9	8	12	1.2-1.4	2/31	22-32	28/5
<i>Schlechtendalia peitan</i>	26	2	2	6	16	1.1-1.4	0/8	27-30	23/3
<i>Nurudea ibofushi</i>	16	3	5	4	4	1.2-1.4	0/13	25-30	15/1
<i>Meithanaphis flavogallis</i>	10	4	3	2	1	1.1-1.4	1/9	27-29	8/2
<i>Floraphis choui</i>	10	3	4	3	0	1.1-1.4	1/9	26-30	8/2
<i>Kaburagia rhusicola</i>	10	4	2	4	0	1.2-1.4	3/7	26-28	8/2
<i>Melaphis rhois</i>	16	9	1	6	0	1.2-1.4	0/17	13-30	16/0

Table 2. Species and genomes of all seven *Rhus* gall aphids included in this study, and detailed information of the sequenced genomes of all the species.

Species	Voucher No.	Genome size (Mbp)	GC content (%)	No. of contigs	Longest contig (bp)	Shortest contig (bp)
<i>Schlechtendalia chinensis</i>	Ren A1798	261	34	64683	660885	100
<i>Schlechtendalia peitan</i>	Ren A242	277	34.4	250994	550865	128
<i>Meithanaphis flavogallis</i>	Ren A2012	235	34.8	169097	537951	128
<i>Kaburagia rhusicola</i>	Ren A63	237	35.2	195776	371348	128
<i>Nurudea ibofushi</i>	Ren A1796	261	33.8	75473	297804	100
<i>Floraphis choui</i>	Ren A403	274	33.7	214731	616219	128
<i>Melaphis rhois</i>	Ren A3037	266M	34	29914	624579	100

Figures

Figure 1

ML phylogenetic tree of Mariner lineages constructed with RAXML using GTRGAMMA substitution model. Clade colours denote the different subfamilies of MLEs. (A) Different subfamilies of the mariner family with element from Tc1 family as outgroup. (B) 165 mariner-like elements, i.e., 44 from GenBank and 121 from seven species of Rhus gall aphids. The branch end showed the NCBI accession numbers of the MLEs downloaded from the database and the names of MLEs detected in this study. Note: * indicates 100% support by all bootstrap replicates.

Figure 2

Full length sequence of MLE (Kmar5) from Maurutiana subfamily in *Kaburagia rhusicola* with slightly modified WVPYE(D)L and YSPDL(V) motifs for transposase.

Figure 3

Full length sequence of MLE (Mfmar4) from Drosophila subfamily in Meithanaphis flavogallis with slightly modified W(L)VPY(K)E(V)L motif and NLS for transposase.

Figure 4

Full length sequence of MLE (Mrmar18) from Vertumana subfamily in Melaphis rhois with a premature stop codon and slightly modified WV(I)PY(H)E(N)L motif and NLS for transposase.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.docx](#)
- [Additionalfile3.fasta](#)