

RESEARCH ARTICLE

Open Access



Diversity and evolution of *mariner*-like elements in aphid genomes

Maryem Bouallègue^{1,2}, Jonathan Filée¹, Imen Kharrat², Maha Mezghani-Khemakhem², Jacques-Deric Rouault¹, Mohamed Makni² and Pierre Capy^{1*}

Abstract

Background: Although transposons have been identified in almost all organisms, genome-wide information on *mariner* elements in Aphididae remains unknown. Genomes of *Acyrtosiphon pisum*, *Diuraphis noxia* and *Myzus persicae* belonging to the *Macrosiphini* tribe, actually available in databases, have been investigated.

Results: A total of 22 lineages were identified. Classification and phylogenetic analysis indicated that they were subdivided into three monophyletic groups, each of them containing at least one putative complete sequence, and several non-autonomous sublineages corresponding to Miniature Inverted-Repeat Transposable Elements (MITE), probably generated by internal deletions. A high proportion of truncated and dead copies was also detected. The three clusters can be defined from their catalytic site: (i) *mariner* DD34D, including three subgroups of the *irritans* subfamily (*Macrosiphinimar*, *Batmar*-like elements and *Dnomar*-like elements); (ii) *rosa* DD41D, found in *A. pisum* and *D. noxia*; (iii) a new clade which differs from *rosa* through long TIRs and thus designated *LTIR*-like elements. Based on its catalytic domain, this new clade is subdivided into DD40D and DD41D subgroups. Compared to other *Tc1/mariner* superfamily sequences, *rosa* DD41D and *LTIR* DD40-41D seem more related to *maT* DD37D family.

Conclusion: Overall, our results reveal three clades belonging to the *irritans* subfamily, *rosa* and new *LTIR*-like elements. Data on structure and specific distribution of these transposable elements in the *Macrosiphini* tribe contribute to the understanding of their evolutionary history and to that of their hosts.

Keywords: Aphids, Comparative genomics, *Tc1-mariner*, Transposable elements, MITEs, Molecular evolution

Background

Genomes contain diverse repetitive DNA sequences of transposable elements (TEs), contributing to their plasticity, adaptability and evolution [1–3]. *Class II* TEs use a “cut and paste” mechanism. They are either autonomous transposons encoding their own transposase or non-autonomous transposons including truncated copies (i.e. copies with only one or no extremity) or copies with internal deletions, but with two intact extremities. Although not encoding for a functional transposase, these shorter copies or miniature inverted repeat transposable elements (MITEs) can be *trans*-mobilized and may reach high copy number with a size homogeneity

that distinguishes them from other non-autonomous elements [4].

The *Tc1/mariner* superfamily is ubiquitous and forms the largest group of eukaryotic *Class II* TEs [5]. Its members share several common characteristics and synapomorphies. In particular, the target insertion site is TA, the ORF of autonomous copies encodes a transposase of 282 to 350 amino-acid residues [6]; the transposase contains two helix–turn–helix (HTH) motifs in DNA binding domains and a catalytic triad DDE/D motif [5, 7].

Despite these similarities, two major differences can separate families of *Tc1/mariner*: (i) their complete length from 1 to 5 kb due to their TIR (i.e. the *mariner*-like element *MLE* 13–34 bp long, the *Tc1*-like element *TLE* ranging from 20 to 600 bp), (ii) the DDE/D signature motif in their catalytic domains which corresponds to DD34D for *mariner*, DD34E for *Tc1*, DD37D for *maT*, DD37E, DD39D, and DD41D for *rosa* [8–10].

* Correspondence: pierre.capy@egce.cnrs-gif.fr

¹Laboratoire Evolution, Génomes, Comportement, Ecologie CNRS, Université Paris-Sud, IRD, Université Paris-Saclay, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

Full list of author information is available at the end of the article



The *mariner* family, initially described in *Drosophila mauritiana* [11], is one of the best known elements belonging to this superfamily. This element is characterized by a patchy and large distribution among metazoans [12–14], which can be explained, in part, by horizontal transfer (HT), corresponding to its ability to transpose between genomes [15–17]. Due to the great diversity of this family, these elements are classified into several subfamilies based on phylogenetic studies. Five major distinct subfamilies including *irritans*, *mauritiana*, *cecropia*, *mellifera/capitata*, and *elegans/briggsae* were reported [12]. However, 16 minor subfamilies also exist with a more limited distribution [18–20]. Otherwise, the *rosa* monophyletic group, first identified in *Ceratitis rosa* and other Tephritid flies, is closely related to the *mariner* subfamilies [9, 16]. Its main characteristic is a transposase with a DD41D motif, and the nucleotide identity between *MLE* subfamilies is about 40 to 56% [12, 21].

While *MLE* is characterized by a high proportion of inactive copies due to independent accumulation of substitution and indels, known as vertical inactivation [22], three elements, namely *mos1*, found in the fruit fly *Drosophila mauritiana* (*mauritiana* subfamily), *Famar1* discovered in the common earwig *Forficula auricularia* (*mellifera* subfamily) and *Mboumar9* isolated from the ant *Messor bouvieri* (*mauritiana* subfamily) are still naturally active, and thus able to be mobilized [12, 23–27]. Furthermore, the *Himar1* element from the horn fly *Haematobia irritans* (*irritans* subfamily) has been reconstructed by *in vitro* mutagenesis to restore a potential activity [28, 29]. Due to their wide distribution and ability to successfully invade new genomes by horizontal transmission, naturally and artificially active *mariner* transposons are used as powerful molecular tools in transgenesis and insertional mutagenesis, *inter alia* leading to genetic control strategies of pests [29–32].

In aphid species, only a few studies have described the presence of *mariner* elements. For instance, (i) internal partial sequences of *irritans* and *mellifera* subfamilies were identified *in vitro* by a Polymerase Chain Reaction (PCR) amplification in the soybean aphid *Aphis glycines* [33], (ii) deleted sequences of *mauritiana* subfamily were characterized in seven fruit tree aphid species [34], (iii) in the first version of pea aphid *Acyrtosiphon pisum* genome [35], only three complete sequences, namely *Mariner-Ap_1*, 2 and 3, were published in RepBase [36]. However these sequences shared catalytic motif DD34E and should be more related to *Tc1*-elements.

Nowadays, three aphid's genomes are available in public databases. Indeed, the recent sequencing of the Russian wheat aphid *Diuraphis noxia* genome (Dnoxia_1.0 reference annotation release 101, <http://www.ncbi.nlm.nih.gov>) [37], the green peach aphid *Myzus persicae* genome (AphidBase, <http://tools.genouest.org/tools/myzus/>), and

the new annotation of *A. pisum* genome (Acyr_2.0, new reference Annotation Release 102, <http://www.ncbi.nlm.nih.gov/>) offer an opportunity to investigate the diversity of the *mariner* family within and between aphid species, along with the evolutionary history and dynamics of these elements.

These species belong to the Macrosiphini tribe of the Aphididae family and diverged approximately 42.5 Mya [38]. They are found on different host plants: while *M. persicae* is generalist and found on peach trees or Solanaceae, *A. pisum* and *D. noxia* are specialist, infesting Fabaceae and cereals, respectively. In this paper, we explored these three genomes in order to identify *mariner*-related transposons and their non-autonomous derivatives through a homology-based method using as queries a panel of transposases from databases. Eleven TE clusters from *A. pisum*, seven from *D. noxia* and four from *M. persicae* have been detected. Classification and phylogenetic analysis suggested (i) that these lineages are divided into three groups: the *irritans* subfamily DD34D, *rosa* DD41D and a new group DD40/41D close to *rosa* and characterized by a long TIR, (ii) an evidence of vertical transfer with stochastic losses and several putative HT events. All these data provide new informations about the evolutionary history of these transposable elements in aphids.

Methods

Supporting data

The genome of *Acyrtosiphon pisum* and *Diuraphis noxia* are available at NCBI (<http://www.ncbi.nlm.nih.gov/>). The first contains 541 Mb covering 23,925 scaffolds and the second includes 393 Mb covering 5641 scaffolds [35, 37]. The genome of *Myzus persicae*, presenting 398 Mb and spanning 34,598 scaffolds, is published in aphidbase (The International Aphid Genomics Consortium <http://tools.genouest.org/tools/myzus/>).

Data mining

A panel of 18 transposases sequences belonging to the five major *mariner* subfamilies DD34D and to the *rosa* DD41D group (Additional file 1) were used as queries in tBLASTN searches on the three aphid genomes, with default parameters. In order to determine the full sequence of each copy, the best hits were extracted with 5 kb flanking sequences and were manually investigated for TIR searches. Each new complete sequence was then used to retrieve more elements. Truncated copies located at the end of scaffolds and sequences less than 250 bp were further discarded. The sequences closer to DDxE catalytic motif were excluded after a BLASTX search against transposases from this family. Finally, 115 sequences from *A. pisum*, 45 from *D. noxia* and 23 from *M. persicae* were obtained and used in this work.

Sequence analyses

The nucleotide sequence analyses, including alignment, were done with the Aliview 1.18 [39]. USEARCH6.0 [40] was performed to cluster repetitive sequences using a threshold of 75% identity. Shorter copies flanked by two TIRs and with evidence of transposition (at least 2 copies) were considered as MITEs [4, 41]. Consensus sequences were derived using the relative majority rule.

The putative amino acid sequences were deduced by ExpasyTool (<http://web.expasy.org/translate/>) and then manually optimized. The nuclear localization sequence (NLS) and the helix-turn-helix (HTH) domain were searched using PSORTII [42] and GYM2.0 [43, 44], respectively (Additional file 2).

Mining of available eukaryote genomes

The complete nucleotide sequences previously identified were used in BLASTn searches against the nr (non-redundant nucleotide) and WGS (whole genome sequence) databases available on the NCBI. Sequences with more than 60% of nucleotide identity over more than 65% of the length of the query were extracted. These thresholds have been chosen to avoid recovering small fragments and sequences phylogenetically far from the subfamilies here considered. Cases of potential horizontal transfers between aphids and other taxa are considered when elements present more than 90% of identity covering more than 90% of the query sequences as proposed by several authors [17, 20].

Classification and phylogenetic analysis

The classification is based on the Unweighted Pair Group Method with Variation of Metric UPGM-VM [19], an ascending hierarchical classification analogous to the UPGMA method, with two main differences: (i) there is no arithmetical mean, the nucleotide sequences are aligned by pairs, (ii) the metric varies with the ascending classification and gap is considered as a fifth nucleotide. This variation allows a complete sequence to be gathered in the same group with the corresponding truncated and/or deleted sequences such as MITEs. Thus, the 183 elements extracted from aphid genomes were added to a set of 96 already known complete sequences from the *Tc1-mariner-IS630* superfamily published in GenBank and to 50 sequences found in eukaryote genomes (Additional file 3). MITE classification is based on identity of TIRs, internal sequences of complete transposable elements and on the breakpoints of deletions.

For phylogenetic analysis, the amino acid sequences were aligned with Aliview1.18 [39] and the best-fitting ML model (AIC, matrix WAG + F + I + G) was selected using Protest 2.4 server [45]. Then, the phylogenetic analysis of transposases was computed using MEGA6 [46] with 1000 bootstrap replicates.

Results

Distribution and diversity of *mariner* and *rosa* elements within the Macrosiphini tribe

Search of sequences belonging to the main *mariner* subfamilies DD34D and to the *rosa* DD41D group was based on a homology approach (tBLASTN) using a set of 18 known transposases as queries (Additional file 1). We found a total of 115 copies from *A. pisum* clustered in 11 lineages, 45 from *D. noxia* clustered in seven lineages and 23 copies from *M. persicae* distributed in four lineages. A lineage corresponds to a group of sequences that is more than 75% similar and to clear phylogenetic clades (see below).

While 183 copies were extracted, 23 complete and potential autonomous sequences, representing 12.57% of all copies, have been identified in aphid genomes. A low copy number, ranging from one to six, per lineage and per species is observed. More precisely, only ten sequences distributed into nine lineages are found in *A. pisum* genome. All these sequences are named *Apismar*. For *D. noxia*, seven complete copies (*Dnomar*) are grouped into five lineages and only six copies from *M. persicae* (*Mpmar*) are gathered in the same group.

For most of these clusters (14 out of 15), the terminal inverted repeats (TIRs) necessary for transposition have been identified, as well as the TA target site duplication (TSD). The *Apismar4.2* does not display a TSD. Interestingly, the whole nucleotide sequences appear heterogeneous in length. Some clusters with a short TIR (15-32 bp) have a full length of approximately 1.3 kb (i.e. *Apismar1.2*, *Apismar4.1*), while others (i.e. *Apismar5.1*, *Apismar5.2*) showed sizes longer than 2 kb due to long TIR sequences about 460 bp (Table 1).

Classification of the 183 aphid sequences, based on the 146 nucleotide sequences of the *Tc1/mariner* superfamily, was performed using a UPGM-VM method. This allows all sequences to be dealt with whatever their length, including the distantly related *Tc1* and *Tc3* sequences of animals, plants, fungi and bacteria like *IS630* (Fig. 1, Additional file 3).

Results reveal that 75 copies (18 complete elements and 57 deleted/truncated sequences) belong to the *irritans* subfamily. They can be subdivided into three tribes: the first is widespread in aphids, namely *Macrosiphinimar* (*Apismar1*, *Dnomar1* and *Mpmar1*). The second is close to known *Batmar*-like elements found in the bat *Rhinolophus ferrumequinum* genome. This group includes complete (*Apismar2* and *Dnomar2*) and shorter sequences (deleted or truncated) from the three aphids species. The last tribe, namely *Dnomar*-like element, contains a complete copy from *D. noxia* (*Dnomar3*) and deleted/truncated sequences from *D. noxia* and *M. persicae*.

Furthermore, two other groups can be identified: *rosa* DD41D and a new one close to the latter (Fig. 1,

Table 1 Characteristics of 15 lineages corresponding to complete elements. The copy number, clade, length of the element, TIR and ORF, as well as the presence of potentially active copies, are specifically indicated for each complete sequence. The number of copies not truncated by “N” is mentioned in the fifth column. Potentially active copy = existence of at least copy with a complete ORF, with no frameshift or codon stop. In TIR sequences, the mirror sites are mentioned in bold

Clade	Tribe	Species	Lineage name	Complete copy number	Length (bp)	TIR		ORF Length (aa)	Potentially active copy
						Length (bp)	Sequences		
<i>irritans</i> DD34D	Macrosiphinimar	<i>A. pisum</i>	Apismar1.1	1	1334	28	CGAGGCGTGTCCAGAAAGTAAGTGACT	354	Yes
			Apismar1.2	1	1317	15	TTCGAAAAGTAAGGG	355	No
		<i>D. noxia</i>	Dnomar1.1	1	1347	28	CGAGGCGTGTCCAGAAAGTAAGTGACT	354	No
			Dnomar1.2	1	1300	20	TWCGAAAAKTAAGGGCCGTT	347	No
		<i>M. persicae</i>	Mpmar1.1	6	1334	28	CGAGGCGTGTCCAGAAAGTAAGTGACT	355	Yes
	Batmar-like	<i>A. pisum</i>	Apismar2.1	1	1323	30	CGAGGTATG CAATAAAATAAY GAGACTTT	354	Yes
			Apismar2.2	2	1280	22	AAYACCCAGACAAMAWKTATTA	354	No
		<i>D. noxia</i>	Dnomar2.1	2	1326	27	YGAKGTGWS AMATAAAATAAA CGAGAC	357	No
	Dnomar-like	<i>D. noxia</i>	Dnomar2.2	2	1344	24	CSWGGTGTGTCAAAAAGWACYCG	339	No
			Dnomar3.1	1	1360	26	CGAGGGCGGGCTGATAAGTAATGCCT	362	No
<i>rosa</i> DD41D	Crmr2-like	<i>A. pisum</i>	Apismar4.1	1	1355	32	AAGGGTGTCTCAAAAAGAACGCCGATTTTAA	361	Yes
			Apismar4.2	1	1299	32	GGTTTTTCAATARRAGCGCTCGAWSTTTSAT	361	No
			Apismar4.3	1	1316	27	GGTGCGGCAGAGCCRACTGACGAGTTT	362	Yes
LTIR	DD41D	<i>A. pisum</i>	Apismar5.1	1	2307	466	TCACCAATTTAGGGAACACTGAATTTCTCGGCT	370	Yes
	DD40D	<i>A. pisum</i>	Apismar5.2	1	2423	460	AATGTGTCAAACCTTCTAGAGGTGTTTCTACACC	351	No

Additional file 3). *rosa* DD41D is represented by 44 copies restricted to *A. pisum* (*Apismar4*) and *D. noxia* genomes. They are clustered with *Crmr2* found in the Diptera Mediterranean fruit fly *Ceratitis rosa*. The second group, characterized by a long TIR, named *LTIR*-like elements, mainly comprises sequences from the pea aphid (*Apismar5.1*, *Apismar5.2*) and may correspond to a new subfamily.

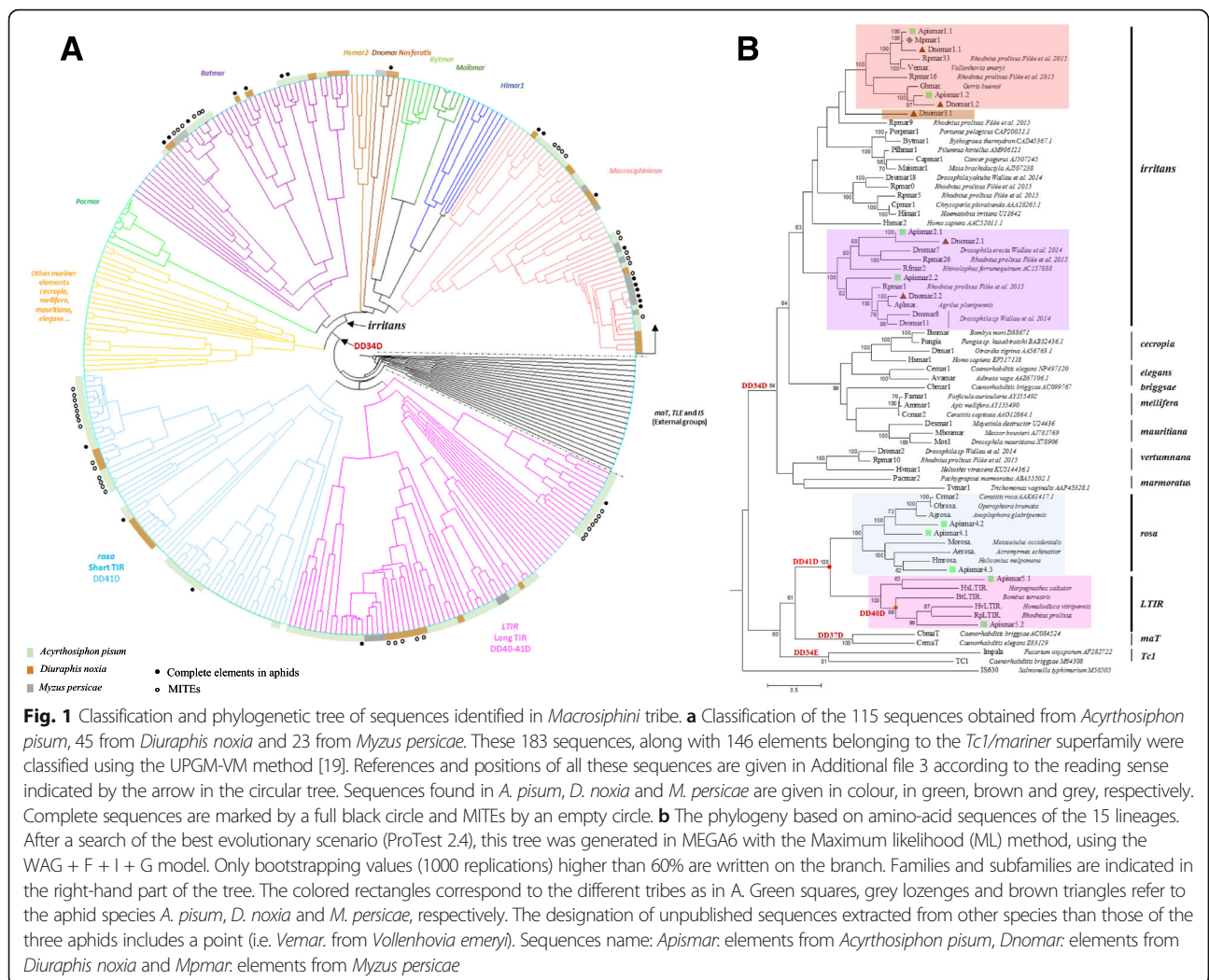
In the same genome, at least four lineages can coexist. However, large differences are observed among species (Fig. 1). Indeed, in *M. persicae*, a potential autonomous element (*Mpmar1*) from *Macrosiphinimar*, related to short sequences, is identified. No *rosa* elements are detected and only deleted/truncated copies belonging to *LTIR*-like, *Dnomar*-like and *Batmar*-like elements are detected. In *D. noxia*, five *irritans* lineages are found. They include potential autonomous elements (*Dnomar*) and a few deleted/truncated copies of the same lineage. Two lineages are composed by short sequences belonging to *rosa* and *LTIR* clades. Furthermore, the genome of *A. pisum* is free of *Dnomar*-like elements. The other lineages are mainly represented by deleted/truncated copies and only a few complete sequences (*Apismar1–5*) can be detected. Hence, the large diversity of these elements among species may reflect the independent evolutionary history of these lineages or specific properties of the genome.

TIRs show a higher degree of identity in the *irritans* subfamily, suggesting a possible recent common ancestor, while they seem to be less conserved in *rosa* and *LTIR* elements (Additional file 4). In addition, TIRs do not present palindromic motifs, but only mirror repeats can be detected in *Apismar2.1* and *Dnomar2.1* belonging to *Batmar*-like elements (Table 1).

Otherwise, the screening of NCBI-nr and WGS databases (Eukaryotes) with the complete elements identified in aphid's genomes reveals only one sequence having a level of similarity above 90%, with cover queries up to 90%. In fact, it concerns a complete element belonging to the *irritans* subfamily found in the genome of the Coleoptera *Agrilus planipennis*, which is closely related to *Dnomar2.2* from *D. noxia* with 92% of similarity (Fig. 1, Additional file 3).

Protein and phylogenetic analyses

The protein sequences of the 15 full clusters are characterized by an ORF encoding about 339 to 370 aa (Fig. 2, Table 1 and Additional file 2). They are aligned with 56 other copies of the *Tc1-mariner* superfamily belonging to non-aphids species. The topology of the ML phylogenetic tree is roughly similar to the classification based on nucleotide sequences (Fig. 1, Additional file 3). Indeed, the five tribes, previously described, are supported by high bootstrap values (98–100%). The percentage of



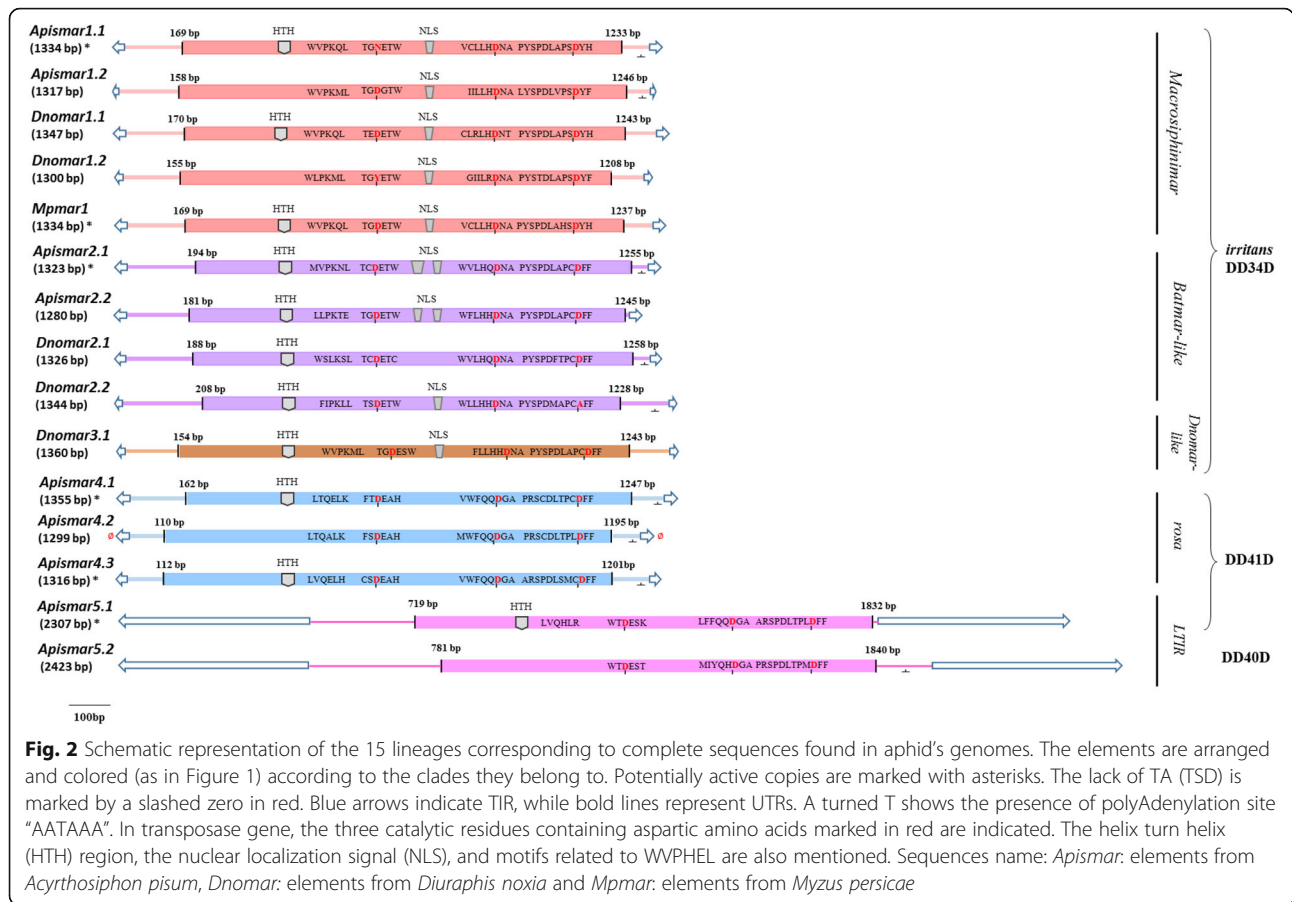
identity between these clades varies from 28 to 59% (Additional file 5).

Only six complete sequences (*Apismar1.1*, *Mpmar1*, *Apismar2.1*, *Apismar4.1*, *Apismar4.3* and *Apismar5.1*) present an intact ORF with no frameshift or codon stop, suggesting that they might be active (Table 1, Additional file 2). An analysis of the transcriptomes of the two species (*A. pisum* and *M. persicae*) was performed using these 6 sequences with a complete ORF. Five sequences (*Apismar1.1*, *Mpmar1.1*, *Apismar4.1*, *Apismar4.3* and *Apismar5.1*) present a full-length transcript, while the last one *Apismar2.1* presents an internal deletion leading to the loss of 140 aa. The sequences related to the conserved motifs, especially WVPHEL and YSPDLA, as well as the catalytic site DD34D considered as the *mariner* signature [47, 48], are detected in most of the sequences belonging to the *irritans* subfamily: *Macrosiphinimar*, *Batmar*-like elements and *Dnomar*-like elements (Fig. 2, Additional file 2). The less conserved motif is WVPHEL,

localized between the HTH motif and the first D. The catalytic site is relatively well conserved (7 out of 10) with a length polymorphism between the three residues. Two sequences are deprived of HTH and one of NLS. These three copies are probably inactive.

In the *rosa* clade, close to *Crmar2*-like elements, the catalytic domain is DD41D rather than the canonical DD34D (Fig. 2, Additional file 2). While the NLS motif is lacking, the HTH is located from position 88 to 110 in *Apismar4.1* and from 90 to 112 in *Apismar4.3*.

The classification and phylogenetic tree showed the presence of a monophyletic clade related to *rosa* DD41D (43% ± 0.016 of similarity), designated *LTIR*. This monophyletic group, characterized by long sequences (> 2.3 kb) with a long TIR (> 460 bp), can be divided into two tribes based on the transposase similarities. The NLS motif is absent and in the catalytic domain the distance between the second and the third D is of 40 aa for *Apismar5.2* and 41 aa for *Apismar5.1*. Otherwise, HTH



motif is only present in *LTIR* DD41D (Fig. 2, Additional file 2). The phylogenetic tree also indicated that *rosa* DD41D and *LTIR* DD40-41D elements are closer to *maT* and *Tc1* than to *mariner* subfamilies (Fig. 1). The comparison of the sequences surrounding the catalytic site is summarized in Fig. 3. The flanking sequences of the second D is clearly distinct between the different groups (*rosa/LTIR/maT* vs the *mariner* subfamilies).

MITEs occurrence: Structure and evolution

MITEs are defined as short non-autonomous copies which are known to derive from autonomous ones. They do not encode functional transposase but can be *trans*-mobilized thank to the transposase of complete copies.

MITEs, detected in the present work, represent 43 copies i.e. 23.5% of all extracted sequences. Only the *Dnomar*-like tribe is free of MITEs (Table 2). For the others, there is a large-size polymorphism, and MITEs are clustered into 11 sublineages based on the breaking points of the main internal deletion and the TIR sequences. All of these sequences, except one (*MITE1.1 sub2*), can be related to a full-length copy (Figs. 1 and 4 and Additional file 3). Microhomologies have been found at the breaking points of the internal deletions for

most of the MITEs. According to the nomenclature proposed by Negoua et al. [49], they are of the BPEE type for seven sublineages of MITE, and of the BPNN type for two other sublineages (Table 2). For the remaining (*MITE1.1*) no microhomology can be detected.

In the *irritans* clade, represented by the *Macrosiphiniar* tribes and *Batmar*-like elements, only *A. pisum* and *M. persicae* contain MITEs, with size varying between 908 and 1165 bp. The first tribe (*MITE1.1*) includes nine copies from the pea aphid clustered in two sublineages (*sub1* and *sub2*) which only share the first 12 nucleotides of the TIRs. An additional lineage (*MITE1.2*), closely related to *MITE1.1sub1*, is found in *M. persicae*. These two sublineages present similar TIRs and an average identity of 81.8%. However, they do not have similar breaking points (Fig. 4). These two types of MITEs are related to putative autonomous copies found in each species (*Apismar1.1* and *Mpmar1.1* respectively) showing 99% of identity.

A similar situation is observed for the *rosa* clade when *MITE4.1sub1* and *MITE4.2* are compared. The *MITE4.1* lineage, includes 12 copies with lengths from 349 to 548 bp, comprised two sublineages. Although clearly related, these sublineages seemed to result from

		D	D	D/E
<i>rosa</i> DD41D	Crmar2	: FSDEA	QQDGA	SCDLTPLDF
	Apismar4.1	: FTDEA	QQDGA	SCDLTPCDF
	Apismar4.2	: FSDEA	QQDGA	SCDLTPLDF
	Aerosa.	: FSDEA	QQDGA	SPDLTPCDF
	Apismar4.3	: CSDEA	QQDGA	SPDLSPCDF
<i>LTIR</i> DD40-41D	BtLTIR.	: WTDER	QQDGA	SPDLTPLDF
	HasLTIR.	: FSDEA	QQDGA	SPDLTSLDF
	Apismar5.1	: WTDES	QQDGA	SPDITPLDF
	Apismar5.2	: WTDES	QHDGA	SPDLTPMDF
<i>maT</i> DD37D	CemaT	: FTDEK	QQDGA	SPDLNPM DY
	CbmaT	: WTDEK	QQDWA	SPDLNPM DF
<i>Tc1</i> DD34E	Impala	: WSDEC	MHDNA	SPDLNPIE-
	CbTc1	: FSDES	QQDND	SPDLNPIE-
<i>mariner</i> subfamilies DD34D	Mos1	: TGDEK	LHDNA	SPDLAPSDY
	Hsmar1	: TCDEK	LHDNA	SPDLSP TDY
	Cemar1	: TGDEK	LHDNA	SPDLAPT DY
	Famar1	: TGDEK	HHDNA	SPDLAPSDY
	Cpmar1	: TMD ET	HQDNA	SPDLAPSDY
	Dnomar3.1	: TGDES	HHDNA	SPDLAPCDF
	Rfmar2	: TGDET	HHDNA	SPDLAPNDF
	Apismar2.1	: TCDET	HQDNA	SPDLAPCDF
	Dnomar2.1	: TCDET	HQDNA	SPDFTPCDF
	Dnomar2.2	: TSDET	HHDNA	SPDMAPCAF
	Apismar2.2	: TGDET	HHDNA	SPDLAPCDF
	Apismar1.1	: TGN ET	LHDNA	SPDLAPSDY
	Dnomar1.1	: TEDET	LHDNT	SPDLAPSDY
	Mpmar1	: TGD ET	LHDNA	SPDLAHS DY
	Apismar1.2	: TGDGT	LHDNA	SPDLVPSDY
Dnomar1.2	: TGYET	LRDNA	STD LAPSDY	

Fig. 3 Multiple alignments of catalytic motifs of *Tc1*, *mariner*, *maT* families with the 15 lineages identified in aphids

independent internal deletions of the *Apismar4.1* complete element. The *D. noxia* genome contains two copies of a MITE of 578 bp (*MITE4.2*) which are also closely related to the autonomous element *Apismar4.1* (Fig. 4).

For the *LTIR* DD41D tribe, *MITE5.1*, only found in *D. noxia*, comprises five copies (790–822 bp) with the same breakpoints, and are related to the autonomous element *Apismar5.1*. No *MITE5.1* was retrieved in the *A. pisum* genome. Furthermore, *MITE5.2* of *LTIR* DD40D tribe identified in the pea aphid is composed of seven short copies (411 and 441 bp). They are divided into two sublineages depending on the breakpoint positions, probably resulting from independent internal deletions (Fig. 4).

Globally, these results show that (i) MITEs in aphid species are less frequent than in *Drosophila ananassae* (about 240 copies) [41] and in *Rhodnius prolixus* (about 400 copies) [20]; (ii) *irritans* clades do not generate MITEs smaller than 900 bp, in contrast to *rosa* and *LTIR*-like elements clades; (iii) three MITE sublineages (*MITE2.2*, *MITE4.2* and *MITE5.1*) are closely related to autonomous copies found in other species; (iv) orphan MITE sublineages can be detected with no full-length

partner (*MITE1.1 sub2*). In the later case, it cannot be excluded that active copies still exist in other populations or closely related species.

The distribution of MITEs and their relationship with full-length elements show that their phylogeny is inconsistent with that of the species. Several scenarios involving the existence of ancestral polymorphism, current population polymorphism (presence/absence of autonomous copies and/or MITEs), stochastic loss of autonomous copies and/or horizontal transfers can be proposed.

To infer the dynamics of MITEs identified in the aphid genomes, we generated consensus sequences for each sublineage in order to estimate their period of amplification from their percentage of divergence, as proposed by Le Rouzic et al. [50] and Wallau et al. [41]. Except for two sequences of the *MITE4.1 sub2* showing 69 and 72% of identity with the consensus of this lineage, all others exhibit a level of identity higher than 85% (Fig. 5). While the transposition rate (*trans*-mobilization) of these copies is unknown, we observed that some of them are almost identical (97–99% of identity) suggesting that these copies are still *trans*-mobilizable or were recently inactivated. The remaining sequences (identity level from 85% to 95%) are less conserved and probably correspond to ancient *trans*-mobilization, and are no longer mobilizable.

Discussion

The three species of aphids, *A. pisum*, *D. noxia* and *M. persicae*, present different genome sizes (541 Mb, 393 Mb and 398 Mb respectively), which correspond to different TE equipment [35, 37], i.e. 38% and 11.5% for the first two species (no information being available for *M. persicae*), suggesting as previously proposed that the contribution of TEs to genome size variation is greater relative to other sources of variation [41, 51, 52].

In the present work, we focused on a survey of *MLE*-related elements in aphid genomes. Our data are in agreement to the previous observation since a total of 115, 45 and 23 sequences, extracted from *A. pisum*, *D. noxia* and *M. persicae*, respectively, are clustered into 22 lineages. The relative abundance of *MLE*-related elements in these three aphids' genomes is low compared to other insect genomes. For instance, *mariner* subfamilies are represented by 10,836 copies in the 700 Mb genome of the Hemiptera *Rhodnius prolixus* [20] and 642 copies in the 156 Mb genome of the *Drosophila eugracilis* [41]. Otherwise, the *Tc1-mariner* superfamily is poorly represented in each aphid genome compared to other superfamilies of DNA transposons, such as *piggyBac* or *hAT* (personal data). This observation might be an illustration of the competition that may occur between

Table 2 List of MITEs detected in the aphid's genomes. NR = no related autonomous copy identified. Presence of short direct repeat (microhomologies) in the region of deletion breakpoints are indicated: BPEE for Breaking Point Exact and BPNN for Breaking Point Near (according to the nomenclature proposed by Negoua et al. [49])

Clade	Tribe	Species	ID MITE	Length (bp)	Sublineage	Copy number	TIR sequences	Autonomous element related to MITE		Breakpoints
								sequences	Average identity (%)	
<i>irritans</i>	Macrosiphinimar	<i>A. pisum</i>	MITE1.1	923–1165	sub1	5	CGAGGRTGTCCAGAAAAGTAAGTGACT	Apismar1.1	90.8	-
			MITE1.2	959, 1007	sub2	4	CGAGCGTGTCCCAAAAARTAAAGTCTCCAT	NR	-	
	Batmar-like	<i>M. persicae</i>	MITE2.1	908–931	sub1	2	CGAGCGTGTCCWGAAGWAAGTGACT	Mpmar1.1	92	BPEE
			MITE2.2	908–912	sub1	3	CGAGGTRTGACAATAAAATAACGAGACTTT	Apismar2.1	98.6	
<i>rosa</i>	Crmr2-like	<i>A. pisum</i>	MITE4.1	349–548	sub1	3	AAGGGTGTCAAAAAGAACGCCGGATTTRAA	Apismar4.1	94.5	BPEE
			MITE4.2	578	sub2	4	RGRTRYCWCAAAAARAAAGSGYGGATTTKRAA		74.6	
	DD41D	<i>D. noxia</i>	MITE5.1	790–822	sub1	5	TCACCAATTTAGGGATCAGTGAATTTCTCGGC...	Apismar5.1	86.2	BPEE
			MITE5.2	411–441	sub1	4	AATGTGCAAACTTCTAGAGGTGTTTCTACAC...	Apismar5.2	90.25	BPEE
BPNN					sub2	3			90	

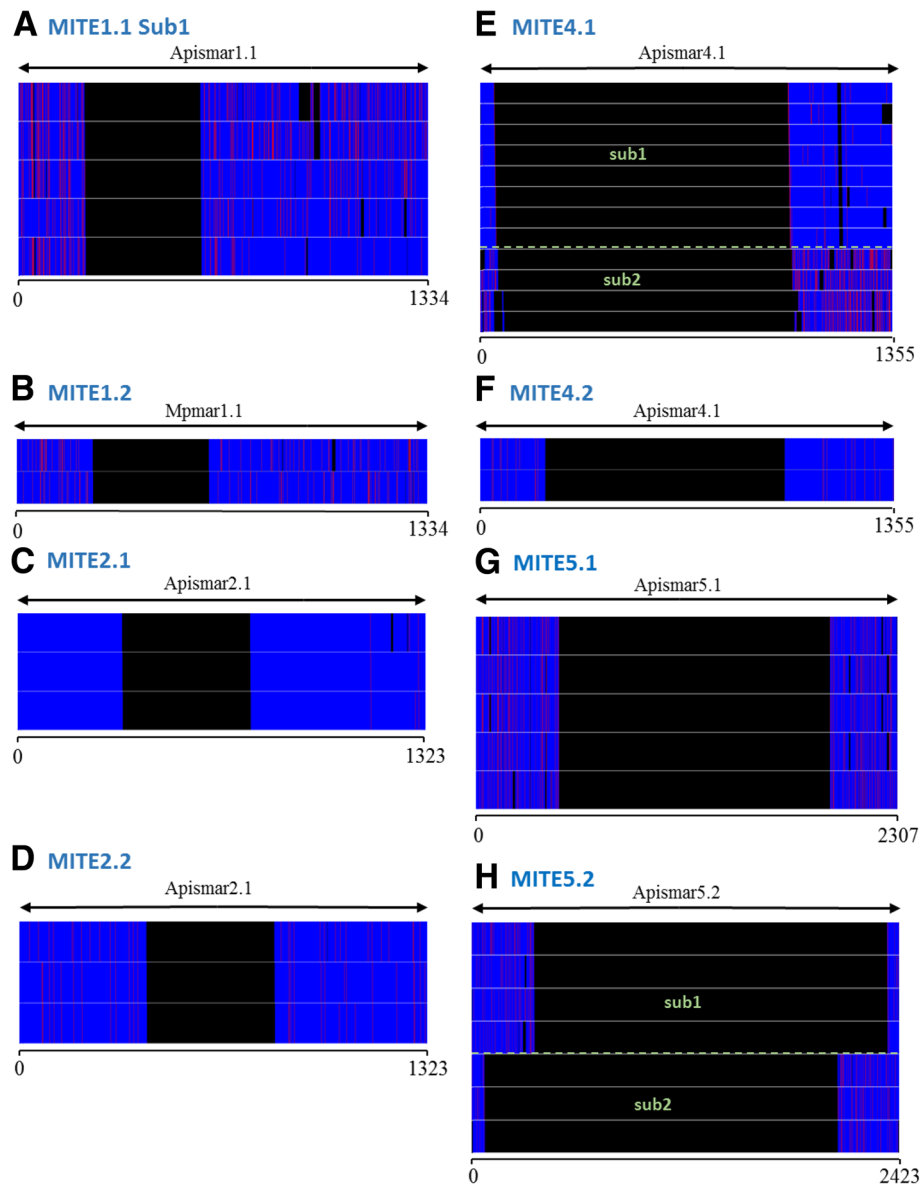


Fig. 4 Sequence alignments of MITE lineages with a longer autonomous partner. For each alignment (a-h), sequences are in blue, showing substitutions in red and gaps in black. The autonomous copies related to MITE and the global structure of the copies are shown on top, with arrowheads corresponding to TIR. Similar copies in length and sequence-defined sublineages (numbered in green). Given the lack of homology with the full potential element, *MITE1.1 sub2* is not represented. **a, c, e** and **h** are found in *A. pisum*, **b** and **d** in *M. persicae*, **f** and **g** in *D. noxia*

superfamilies as described by Abrusán and Krambeck [53]. However, today without a complete and detailed overview of TE equipment of these genomes, we do not have strong arguments to conclude that such a result is due to competition.

In the *mariner* family, only members of the *irritans* subfamily are identified in the aphid's genomes. They belong to the *Macrosiphinimar*, *Batmar*-like and *Dnomar*-like tribes, and are characterized by the DD34D catalytic site. Moreover, only three lineages might still be active (*Apismar1.1*, *Mpmar1.1* and *Apismar2.1*). No sequence related

to other *mariner* subfamilies (i.e. *mauritaniana*, *mellifera*, *cecropia*, *elegans*) is found in these genomes, although they have been identified in vitro in other species belonging to a closely related aphid species such as *Aphis glycines* [33] and seven tree aphids [34].

However, sequences belonging to the *rosa* family (initially closely related to the *mariner* family [9]) have been detected in *A. pisum* and *D. noxia*; and a novel clade (*LTIR*-like) has been identified. Such *LTIR* elements including the DD41D motif, designated as *Lsra* transposons, were described by Zhang et al. [54].

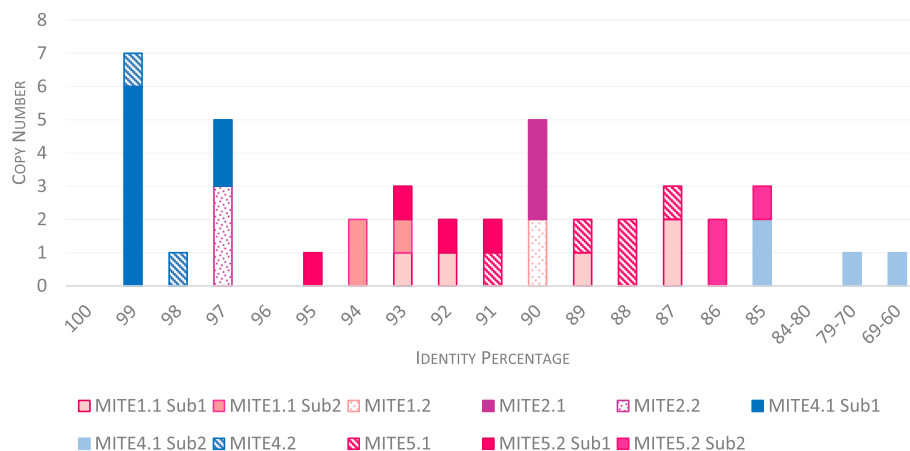


Fig. 5 Evolution analysis of different MITEs sublineages. Based on the comparison of consensus with copies, the similarity rates are identified. While copy sublineages with a high level of similarity present recent invasion, the decrease of this percentage refers to an ancient element. Filled, hatched and dotted patterns correspond to *A. pisum*, *D. noxia* and *M. persicae*, respectively. Colors match to the different tribes as in Fig. 1

This clade is closely related to the *rosa* subfamily but is characterized by long TIRs (about 460 bp vs 28–32 bp). Moreover, conservation of some specific amino acid residues in their catalytic region, especially the final aspartic acid (D) rather than glutamic acid (E), and phylogenetic analysis revealed that *rosa* and *LTIR*-like elements are more closely related to *maT* elements than to *Tc1* and *mariner* ones. Therefore, we suggest that *rosa* DD41D and *LTIR*-like elements constitute a large new family belonging to *Tc1/mariner*.

Distribution, diversity and phylogeny of these elements in the three aphids' genomes are probably the result of vertical transmissions associated to an ancestral polymorphism. In such a situation, closely related sequences derived from the same ancestral copy can be found in several species, while copies derived from different ancestral copies and found in the same genome, can be more distantly related (see for instance [55–57]). Host genomes are also able to repress TE activity [58, 59], leading to their elimination by stochastic loss or vertical extinction. Therefore, the absence of members of the *rosa* family may be due to a stochastic loss during the evolutionary trajectory of *M. persicae*. A similar observation was illustrated in some *Drosophila* species for *mariner* subfamilies [41, 60].

The high level of similarity between MITEs and autonomous partner indicates that short sequences are internally deleted elements, deriving from complete copies. Most of them exhibit direct repeat microhomologies exactly (BPE) or nearly (BPN) to the deletion breakpoints, suggesting that these internal deletions are probably due to abortive gap repair [49, 61, 62]. However, MITEs and related complete copies can be found in two different species, as

described in the *R. prolixus* and *Drosophila* genus [20, 41]. This is the case for *MITE2.2*, *MITE4.2* and *MITE5.1*. To explain such observations, two scenarios can be proposed. First, the ancestral autonomous element at the origin of MITEs may have been lost after the MITE amplification, but was maintained in another species. Another hypothesis consists in the emergence of MITEs after internal deletion(s) of a complete copy, these MITEs being then mobilized by the transposase of another copy closely related to the first one.

Finally, horizontal transfer may also occur for all these sequences between distantly related species. For instance, the *mariner* autonomous transposon *Dnomar2.2* from *D. noxia* is closely related to the sequence of *Agrilus planipennis*. Despite a divergence time of about 361 Mya between these two species (<http://www.timetree.org/home>), the phylogenetic tree of these elements is inconsistent with that of the species. Moreover, HT could also explain the patchy distribution of MITE elements in aphids. However, in all these cases, the transfer mechanism(s) remain unknown and only propositions are suggested, like those proposed in Silva et al. [63] and Loreto et al. [64].

Conclusion

Our results represent the first in silico evidence of diversity and possible evolutionary scenarios of elements belonging to the three clades: *irritans*, *rosa* and a new one named *LTIR*-like elements in aphid genomes. This latter clade is characterized by long TIRs and subdivided into two distinct subgroups based on the catalytic domain signature DD40D or DD41D. Moreover, based on protein and phylogenetic analyses, the *rosa* and *LTIR* transposons are related to

maT DD37D elements, indicating a recent common ancestor. We also demonstrated the presence of several MITE lineages deriving from internal deletion of autonomous elements. Finally, this study proposes an update of the classification of the *Tc1/mariner* superfamily. Data analyses will offer a basis for future research aiming to understand the role of transposable elements during evolution and to develop biotechnological applications for the genetic control of aphid species.

Additional files

Additional file 1. *mariner* and *rosa* transposases sequences used as queries in the tBLASTN search (Species, Clades, Accession number). (PDF 408 kb)

Additional file 2. Amino acid sequences of the transposase of the 15 complete elements. The three aspartic residues of the catalytic domain are marked in red, the sequences related to the WPHL-specific motif of the DNA binding domain are indicated in black bold as well as the helix turn helix (HTH) region (underlined) and the NLS (in blue). Stop codons are represented by asterisks (*). (PDF 425 kb)

Additional file 3. Sequences classified by UPGM-VM method according to the reading sense indicated by the arrow in the circular tree. Deleted or truncated sequences are indicated by an asterisk (*). (PDF 276 kb)

Additional file 4. TIR sequences for each clade. TIR consensus per clade was generated using the Web-Logo server (<http://weblogo.berkeley.edu/logo.cgi>). At each position the nucleotides are stacked one on top of another with the most frequent one on the top. It displays the frequency of bases at each position, with height indicating the proportion of occurrence. The vertical scale is in bits with maximum of two bits possible at each position, indicating that there can be possibility of four different bases at each position. For LTIR-like elements, only the first 57 nucleotides are presented. (PDF 111 kb)

Additional file 5. Pairwise divergence matrix between amino acid lineages. Fifteen complete sequences have been aligned using Aliview. The alignment was then transferred in GENEDOC software [65] to obtain the identity percentage. Sequences name: *Apismar*: elements from *Acyrtosiphon pisum*, *Dnomar*: elements from *Diuraphis noxia* and *Mpmar*: elements from *Myzus persicae*. (PDF 354 kb)

Abbreviations

BPEE: Breaking point exact exact; BPNN: Breaking point near near; HTH: Helix turn helix; LTIR-like elements: Long terminal inverted repeats like elements; MITEs: Miniature inverted-repeat transposable elements; MLE: *mariner*-like element; NCBI-nr: Non-redundant nucleotide; NLS: Nuclear localization signal; TE: Transposable element; TIR: Terminal inverted repeats; TLE: *Tc1*-like element; TSD: Target site duplication; UPGM-VM: Unweighted pair group method with variation of metric; WGS: Whole genome sequence

Acknowledgements

Authors thank Aurélie Hua-Van for its helpful comments and Malcolm Eden for the English review of the manuscript.

Funding

This work was financially supported by the Tunisian Ministry of Higher Education and Scientific Research, the University of Tunis El Manar, the Centre National de la Recherche Scientifique and the Paris-Sud University.

Availability of data and materials

All the data supporting these findings is contained within the manuscript.

Authors' contributions

MB, MM and PC conceived and designed research. MB performed research. MB, JF,JK, MMK and JDR contributed to the analysis. MB, MM and PC drafted the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable

Ethics approval and consent to participate

Not applicable

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Laboratoire Evolution, Génomes, Comportement, Ecologie CNRS, Université Paris-Sud, IRD, Université Paris-Saclay, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France. ²Faculté des Sciences de Tunis, UR11ES10 Génomique des Insectes Ravageurs de Cultures, Université de Tunis El Manar, 1002 Tunis, Tunisie.

Received: 4 January 2017 Accepted: 9 June 2017

Published online: 29 June 2017

References

1. Biémont C, Vieira C. Genetics: Junk DNA as an evolutionary force. *Nature*. 2006;443(7111):521–4.
2. Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet*. 2007;41:331–68.
3. Chénéais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene*. 2012;509(1):7–15.
4. Feschotte C, Zhang X, Wessler SR. Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. *Mobile DNA II*. Washington DC: ASM Press; 2002. p. 1147–58.
5. Brillat 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):105–20.
6. Plasterk RH, Izsvák Z, Ivics Z. Resident aliens: The *Tc1/mariner* superfamily of transposable elements. *Trends Genet*. 1999;15(8):326–32.
7. Lohe AR, Hartl DL. Autoregulation of *mariner* transposase activity by overproduction and dominant-negative complementation. *Mol Biol Evol*. 1996;13(4):549–55.
8. 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.
9. Gomulski LM, Torti C, Bonizzoni M, Moralli D, Raimondi E, Capy P, et al. A new basal subfamily of *mariner* elements in *Ceratitiss rosa* and other tephritid flies. *J Mol Evol*. 2001;53(6):597–06.
10. Claudianos C, Brownlie J, Russell R, Oakeshott J, Whyard S. *maT* a clade of transposons intermediate between *mariner* and *Tc1*. *Mol Biol Evol*. 2002;19(12):2101–9.
11. Medhora MM, MacPeck AH, Hartl DL. Excision of the *Drosophila* transposable element *mariner*: Identification and characterization of the *Mos* factor. *EMBO J*. 1988;7(7):2185.
12. Robertson HM, MacLeod EG. Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. *Insect Mol Biol*. 1993;2(3):125–39.
13. Auge-Gouillou C, Bigot Y, Pollet N, Hamelin MH, Meunier-Rotival M, Periquet G. Human and other mammalian genomes contain transposons of the *mariner* family. *FEBS Lett*. 1995;368(3):541–6.
14. Sinzelle L, Chesneau A, Bigot Y, Mazabraud A, Pollet N. The *mariner* transposons belonging to the *irritans* subfamily were maintained in chordate genomes by vertical transmission. *J Mol Evol*. 2006;62(1):53–65.

15. Laha T, Loukas A, Wattanasatitarpas S, Somprakhon J, Kewgrai N, Sithithaworn P, et al. The *bandit*, a new DNA transposon from a hookworm-possible horizontal genetic transfer between host and parasite. *PLoS Negl Trop Dis*. 2007;1(1):e35.
16. Dupeyron M, Leclercq S, Cerveau N, Bouchon D, Gilbert C. Horizontal transfer of transposons between and within crustaceans and insects. *Mob DNA*. 2014;5:4.
17. 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*. *Mol Biol Evol*. 2016;33(4):1094–09.
18. Robertson HM, Soto-Adames FN, Walden K, Avancini RM, Lampe D. The *mariner* transposons of animals: Horizontally jumping genes. In: Kado CI, editor. *Horizontal gene transfer*. San Diego CA: Academic Press; 2002. p. 173–85.
19. 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.
20. Filée J, Rouault JD, Hary M, Hua-Van A. *Mariner* transposons are sailing in the genome of the blood-sucking bug *Rhodnius prolixus*. *BMC Genomics*. 2015;16(1):1.
21. Bigot Y, Brillet B, Auge-Gouillou C. Conservation of palindromic and mirror motifs within inverted terminal repeats of *mariner*-like elements. *J Mol Biol*. 2005;351(1):108–16.
22. Kidwell MG, Lisch DR. Perspective: Transposable elements, parasitic DNA, and genome evolution. *Evolution*. 2001;55(1):1–24.
23. Jacobson JW, Medhora MM, Hartl DL. Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proc Natl Acad Sci U S A*. 1986;83(22):8684–8.
24. Maruyama K, Hartl DL. Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. *J Mol Biol*. 1991;33(6):514–24.
25. Medhora M, Maruyama K, Hartl DL. Molecular and functional analysis of the *mariner* mutator element *mos1* in *Drosophila*. *Genetics*. 1991;128(2):311–8.
26. Barry EG, Witherspoon DJ, Lampe DJ. 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*. 2004;166(2):823–33.
27. Muñoz-López M, Siddique A, Bischerour J, Lorite P, Chalmers R, Palomeque T. Transposition of Mbouar-9: Identification of a new naturally active *mariner*-family transposon. *J Mol Biol*. 2008;382(3):567–72.
28. Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM. Hyperactive transposase mutants of the *Himar1 mariner* transposon. *Proc Natl Acad Sci U S A*. 1999;96(20):11428–33.
29. Rholl DA, Trunck LA, Schweizer HP. In vivo *Himar1* transposon mutagenesis of *Burkholderia pseudomallei*. *Appl Environ Microb*. 2008;74(24):7529–35.
30. O'brochta DA, Atkinson PW. Transposable elements and gene transformation in non-drosophilid insects. *Insect Biochem Mol Biol*. 1996;26(8):739–53.
31. Wang W, Swevers L, Iatrou K. *mariner* (*mos1*) transposase and genomic integration of foreign gene sequences in *Bombyx mori* cells. *Insect Mol Biol*. 2000;9(2):145–55.
32. Delaurière L, Chénais B, Hardivillier Y, Gauvry L, Casse N. *mariner* transposons as genetic tools in vertebrate cells. *Genetica*. 2009;137(1):9–17.
33. Mittapalli O, Rivera-Vega L, Bhandary B, Bautista MA, Mamidala P, Michel AP, et al. Cloning and characterization of *mariner*-like elements in the soybean aphid, *Aphis glycines* Matsumura. *Bull Entomol Res*. 2011;101(6):697–704.
34. Kharrat I, Mezghani M, Casse N, Denis F, Caruso A, Makni H, et al. Characterization of *mariner*-like transposons of the mauritiana subfamily in seven tree aphid species. *Genetica*. 2015;143(1):63–72.
35. Richards S, Gibbs RA, Gerardo NM, Moran N, Nakabachi A, Stern D, et al. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol*. 2010;8(2):e1000313.
36. Jurka J. *mariner* families from *Acyrtosiphon pisum*. *Rebase Reports*. 2008;8(3):340.
37. Nicholson SJ, Nickerson ML, Dean M, Song Y, Hoyt PR, Rhee H, et al. The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC Genomics*. 2015;16(1):1.
38. 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.
39. Larsson A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014;30(22):3276–8.
40. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460–1.
41. 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):1.
42. Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics*. 2002;18(2):298–05.
43. Gao Y, Mathee K, Narasimhan G, Wang X. Motif detection in protein sequences. In *Proceedings of the Sixth Intl. Symp. on String Processing and Information Retrieval*; 1999. p. 63–72.
44. Narasimhan G, Bu C, Gao Y, Wang X, Xu N, Mathee K. Mining protein sequences for motifs. *J Comput Biol*. 2002;9(5):707–20.
45. Abascal F, Zardoya R, Posada D. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics*. 2005;21(9):2104–5.
46. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725–9.
47. Robertson HM. The *mariner* transposable element is widespread in insects. *Nature*. 1993;362(6417):241–5.
48. Hartl DL, Lozovskaya ER, Nurminsky DI, Lohe AR. What restricts the activity of *mariner*-like transposable elements? *Trends Genet*. 1997;13(5):197–01.
49. Negoua A, Rouault JD, Chakir M, Capy P. Internal deletions of transposable elements: The case of Lemi elements. *Genetica*. 2013;141(7–9):369–79.
50. Le Rouzic A, Boutin TS, Capy P. Long-term evolution of transposable elements. *Proc Natl Acad Sci U S A*. 2007;104(49):19375–80.
51. Kidwell MG. Transposable elements and the evolution of genome size in eukaryotes. *Genetica*. 2002;115:49–63.
52. Vieira C, Nardon C, Arpin C, Lepetit D, Biémont C. Evolution of genome size in *Drosophila*. Is the invader's genome being invaded by transposable elements? *Mol Biol Evol*. 2002;19(7):1154–61.
53. Abrusán G, Krambeck HJ. Competition may determine the diversity of transposable elements. *Theor Popul Biol*. 2006;70(3):364–75.
54. Zhang HH, Shen YH, Xiong XM, Han MJ, Zhang XG. Identification and evolutionary history of the DD41D transposons in insects. *Genes Genomics*. 2016;38(2):109–17.
55. Capy P, Anxolabéhère D, Langin T. The strange phylogenies of transposable elements: Are horizontal transfers the only explanation? *Trends Genet*. 1994;10(1):7–12.
56. Green CL, 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):371–86.
57. Prasad MD, Nurminsky DL, Nagaraju J. Characterization and molecular phylogenetic analysis of *mariner* elements from wild and domesticated species of silkworms. *Mol Phylogenet Evol*. 2002;25(1):210–7.
58. Rozhkov NV, Aravin AA, Zelentsova ES, Schostak NG, Sachidanandam R, McCombie WR, et al. Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *RNA*. 2010;16(8):1634–45.
59. Tóth KF, Pezic D, Stuwe E, Webster A. The piRNA pathway guards the germline genome against transposable elements. In: *Non-coding RNA and the Reproductive System*. Springer Netherlands; 2016. p. 51–77.
60. Capy P, David JR, Hartl DL. Evolution of the transposable element *mariner* in the *Drosophila melanogaster* species group. *Genetica*. 1992;86:37–46.
61. Rubin E, Levy AA. Abortive gap repair: Underlying mechanism for *Ds* element formation. *Mol Cell Biol*. 1997;17(11):6294–02.
62. Brunet F, Giraud T, Godin F, Capy P. Do deletions of *Mos1*-like elements occur randomly in the *Drosophilidae* Family? *J Mol Evol*. 2002;54(2):227–34.
63. Silva JC, Loreto EL, Clark JB. Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol*. 2004;6(1):57–71.
64. Loreto ELS, Carareto CMA, Capy P. Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity*. 2008;100(6):545–54.
65. Nicholas KB, Nicholas HBJ, Deerfield DW. GeneDoc: Analysis and visualization of genetic variation. *EMBNWNEWS*. 1997;4(1):14.