Molecular characterization and phylogenetic position of a new mariner-like element in the coastal crab, Pachygrapsus marmoratus

Quynh-Trang Bui, Laurence Delaurière, Nathalie Casse, Violaine Nicolas, Marc Laulier, Benoît Chénais

Available online at www.sciencedirect.com
ScienceDirect
Gene 396 (2007) 248–256
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Molecular characterization and phylogenetic position of a new *mariner*-like element in the coastal crab, *Pachygrapsus marmoratus*

Quynh-Trang Bui [a], Laurence Delaurière [a], Nathalie Casse [a], Violaine Nicolas [b], Marc Laulier [a], Benoît Chénais [a,*]

[a] *Laboratoire de Biologie et Génétique Evolutive (EA3265), Avenue Olivier Messiaen, 72085 Le Mans cedex, France*
[b] *Museum National d'Histoire Naturelle, Département de Systématique et Evolution, USM 601, UMR 5202, 55 rue buffon, 75005 Paris, France*

Received 14 September 2006; received in revised form 1 March 2007; accepted 12 March 2007
Available online 23 March 2007
Received by A. Bernardi

Abstract

Mariner-like elements (*MLEs*) are class-II transposable elements that move within the genome of their hosts by means of a DNA-mediated "cut and paste" mechanism. *MLEs* have been identified in several organisms, from most of the phyla. Nevertheless, only a few of the sequences characterized contain an intact open reading frame. Investigation of the genome of a coastal crab, *Pachygrapsus marmoratus*, has identified nine *Pacmmar* elements, two of which have an open reading frame encoding a putatively functional transposase. Nucleic acid analyses and comparison with the previous data showed that the GC contents of *MLEs* derived from coastal organisms such as *P. marmoratus* are significantly higher than those of terrestrial *MLEs* and significantly lower than those of hydrothermal ones. Furthermore, molecular phylogeny analyses have shown that *Pacmmar* elements constitute a new lineage of the *irritans* subfamily within the *mariner* family.

Keywords: Transposable element; *Mariner* transposon; GC content; Transposase; Crustacea; Crab

1. Introduction

Mariner-like elements (*MLEs*) are mobile genetic elements able to move from one locus to another within the genome of their eukaryotic hosts. *MLEs* belong to the *IS630/Tc1/maT/mariner* superfamily of class-II transposons that move using a DNA-mediated "cut and paste" mechanism. *MLEs* are one of the simplest forms of autonomous DNA characterized by a single gene flanked by two short (26 to 38 pb long) imperfect inverted terminal repeats (ITRs) (Jacobson et al., 1986; Hartl, 1989). The total length of *MLEs* is about 1300 base pairs (bp) and the unique intron-less gene, encoding a transposase of about 340–360 amino acid residues, is boarded by 5′ and 3′ untranslated regions which are about 135 bp and 50 bp long, respectively (Plasterk et al., 1999). In the well-characterized *mos1 MLE*, the transposase is organized in two domains: 1) The N-terminal domain of about 150 amino acids accounts for the ITRs and target site binding. This DNA binding region contains a helix-turn-helix motif (HTH), which is required for recognition of the ITRs of the corresponding elements (Wang et al., 1999), one or two nuclear location signals (NLS), accounting for the transposase importation into nucleus, and a relatively conserved motif WVPHEL (Robertson, 1993). The latter motif was shown to be involved in the interaction of transposase molecules in the synaptic complex created during the transposition process (Augé-Gouillou et al., 2005; Richardson et al., 2006). 2) The C-terminal domain of about 130 amino acids contains the catalytic core of the transposase, the (D, D34D) triad, in which 34 indicates the number of amino acids separating the last two aspartic acids (D), and the well-conserved motif YSPDLAPD (Robertson, 1993) in which the last D corresponds to the last D of the triad. The D,D(34)D motif

Abbreviations: bp, base pair; ORF, open reading frame; SDS, sodium dodecyl sulfate.

* Corresponding author. Laboratoire de Biologie et Génétique Evolutive, Université du Maine, UFR Sciences et Techniques, Avenue Olivier Messiaen, F-72085 Le Mans cedex 9, France. Tel.: +33 2 43 83 32 51; fax: +33 2 43 83 39 17.
E-mail address: bchenais@univ-lemans.fr (B. Chénais).

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.03.004

is typical of the *mariner* family (Jacobson et al., 1986; Hartl, 2001).

To date, phylogenetic analyses of the *MLEs* sequence evolution have identified at least five *MLE* subfamilies with typically 40–56% identical amino acids: *mauritiana*, *cecropia*, *mellifera*/*capitata*, *irritans*, *elegans*/*briggsae* (Robertson and Lampe, 1995; Gomulski et al., 2001; Shao and Tu, 2001; Feschotte and Wessler, 2002; Rouleux-Bonnin et al., 2005). The recent discovery of *Tvmar1*, the first representative element to be found in a human parasite (*Trichomonas vaginalis*), has suggested the existence of a hitherto-unknown branch of the *mariner* family (Silva et al., 2005). Species phylogenies based on their *MLEs* are often different from those based on the usual characters. As previously suggested, this may be due to ancestral polymorphism or may be attributable to horizontal transfer of *MLEs* across boundary species (Capy et al., 1994a,b; Lampe et al., 2003; Silva et al., 2004, Hua-Van et al., 2005).

The ubiquity of *MLEs* in the genome of continental organisms encompassing plants, fungi, invertebrates, and mammals has been well documented (Robertson, 2002). In contrast, there has been little attempt to search for *MLEs* in aquatic organisms, and especially in marine invertebrates. The presence of an *MLE* fragment has been demonstrated in genomes of two marine flatworms, *Stylochus zebra* and *Bdelloura candida* (Robertson, 1997). Our laboratory has isolated transposase fragments from the coastal Norway lobster *Nephrops norvegicus*, the hydrothermal bivalve *Bathymodiolus thermophilus*, the gutless worm *Riftia pachyptyla*, and the crab *Bythograea thermydron* (Casse et al., 2000). These short fragments were obtained using the degenerated primers designed by Robertson (1993) that correspond to the WVPHEL and YSPDLAP motifs of the transposase. In addition, an *MLE* has been found that is closely linked to trypsin genes in the genome of the parasitic marine crustacean, *Lepeophtheirus salmonis* (Kvamme et al., 2005). Finally, complete *MLE* sequences were obtained by our group in the genome of several marine coastal and hydrothermal crustacean (Casse et al., 2002; Halaimia-Toumi et al., 2004; Casse et al., 2006), and all these *MLEs* belong to the *irritans* subfamily.

Here, we describe *MLEs* found in the genome of a coastal crab, *Pachygrapsus marmoratus*. Nine *Pacmmar* elements (*Pacmmar1.1* to *1.9* for *P. marmoratus mariner* and clone number, according to Robertson and Asplund, 1996) were obtained from two individuals. *Pacmmar1.1* and *1.2* presented an ORF encoding a putatively functional transposase. Moreover, these *MLEs* belong to a new lineage of the *irritans* subfamily.

2. Materials and methods

2.1. Sample collection and DNA extraction

P. marmoratus individuals were collected on the French Atlantic coast. Genomic DNA (gDNA) from two individuals was purified separately according to Laulier et al. (1995). Following the treatment with RNase and proteinase K, DNA was extracted using phenol/chloroform, precipitated with ethanol, and resuspended in Tris–EDTA buffer (10 mM Tris–HCl, pH 8, 1 mM EDTA).

2.2. PCR and inverse PCR amplifications

The sequences of full-length *Pacmmar* were obtained by a polymerase chain reaction (PCR) using the 5′ ITR (5′-TACGAGGGGCGGTCAGAAAGTTATG-3′) designed from the *Bytmar1* element (Halaimia-Toumi et al., 2004) as a primer, in an Eppendorf Mastercycler. Reactions were done in a 25-µl reaction volume using 100 ng gDNA, 20 pmol of primer, 2 mM deoxynucleotide triphosphate mixture, 1.5 mM $MgCl_2$, 2.5 µl of 10X reaction buffer (10 mM Tris, 50 mM KCl, pH 8) and 1 unit of Taq DNA polymerase (Promega). Amplification was as follows: an initial denaturing step at 94 °C for 2 min, then 30 cycles of denaturing at 94 °C for 1 min, annealing at 55 °C for 30 s, extension at 72 °C for 1 min, followed by a final extension at 72 °C for 10 min were performed. PCR products were run on a 1% agarose gel containing 1 µg ml^{-1} of ethidium bromide. The primer used to amplify *MLEs* in the genome of *P. marmoratus* does not correspond to the real sequence of *Pacmmar* ITRs but are the 25 first bases of the 5′ ITR of *Bytmar1*. In order to amplify the specific *Pacmmar* ITRs and sequences flanking the *Pacmmar* elements, an inverse PCR strategy was undertaken. Since *Pacmmar* elements are numerous in *P. marmoratus* genome (see Fig. 2), our hypothesis was that some *Pacmmar* elements are located side by side. Then, using primers designed in the internal region of the element and oriented out of the element, it should be possible to amplify the ITRs and the region located between two successive copies of *Pacmmar*. Therefore, outgoing primers were designed, the first one in the 5′ UTR region, and the second overlapping the 3′ UTR and the 3′ end of the transposase ORF. The primer sequences were: 5′-TATATGTAACCCA-GCCAA-GCGGCGGCAAGTTCAA-3′ and 5′-AGTGGCGAAAGAG-GATGGAACGCT-GCATAGAG-3′. Amplifications were performed using 1 unit of Expand Long Template PCR System Kit (Roche Applied Science), in reaction buffer 2, which can be used to amplify fragments of up to 12–15 kb, and supplemented with 22.5 mM of $MgCl_2$. The primers were supplied at 300 mM, and the deoxynucleotide triphosphate mixture at 500 µM. The PCR cycling conditions were as follows: denaturing at 94 °C for 2 min, then 10 cycles of denaturing at 94 °C for 10 s, annealing at 62 °C for 30 s, and extension at 68 °C for 5 min, followed by 20 cycles consisting of 94 °C for 10 s, 62 °C for 30 s, extension at 68 °C for 5 min in the first cycle, and then for an extension time incremented by 20 s in each of the next 19 cycles. The PCR fragments obtained were eluted, cloned and sequenced.

2.3. Cloning and sequencing

PCR amplification products of approximately 1300 bp were eluted from 0.8% agarose gel using Wizard SV Gel and PCR Clean-up System (Promega). The eluted products were ligated into the plasmid pGEM-T easy vector system I (Promega) overnight at 16 °C and then introduced in DH5α bacteria. The

competent cells were mixed with 5 μl of ligation product 30 min on ice, and then heat shocked 3 min at 42 °C. Transformed bacteria were placed in 1 ml of Luria Broth (LB) medium for 1 h 30 at 37 °C, then centrifuged 4 min at 4000 *g* and plated on LB agar medium containing ampicillin (100 μg/ml), 10 μl of 400 mM isopropyl-β-d-thiogalactopyranoside and 40 μl of 2% X-gal (5-Bromo-4-Chloro-3-indolyl-β-d-galactopyranoside) for blue/white screening for recombinants, and incubated overnight at 37 °C. Plasmids from white colonies were purified using Wizard plus miniprep Kit (Promega). The size of the insert was checked by *Apa*I/*Pvu*I (QBiogen) digestion. Plasmids containing an insert of approximately 1300 pb long were sequenced by Genome Express (http://www.gexbyweb.com/).

2.4. Computational analyses

The sequence alignments were performed by using Infobiogen facilities available at http://www.infobiogen.fr/. Nucleotide sequences were aligned using Clustal W software. Motif signatures of *MLEs* were searched for using software from the Pôle Bioinformatique Lyonnais (http://pbil.univ-lyon1.fr/) and our own *MLE* database (http://www.univ-lemans.fr/sciences/lbge/). The complete protein sequences of transposase genes from several elements of the *Tc1/maT/mariner* superfamily were downloaded from GenBank, and aligned using clustal W with default parameters. Phylogenetic analyses were performed using PAUP 4.0b10 (Swofford, 2002). Parsimony analyses were performed using the tree-bisection-reconnection (TBR) branch swapping option with 10 random addition replicates. The robustness of internal nodes was estimated by 100 bootstrapping replicates. All amino acids were given equal weight, and any gaps were treated as missing data. The calculation of the GC contents of the nucleic acid sequences was carried out using software available on the Infobiogen web site (http://www.infobiogen.fr/). The GC contents calculated correspond to each full-sized *MLE*, each ORF encoding a *MLE* transposase, and to those at positions 1, 2 and 3 (GC1, GC2 and GC3) of the codons used in the *MLE* ORF. GC contents were calculated at the website http://www.infobiogen.fr/services/analyseq/cgi-bin/freqbnk_in.pl. The XpX dinucleotide frequencies were checked by calculating the observed XpX frequency/expected XpX frequency (XpXobs/exp) ratio at the web site http://www.infobiogen.fr/services/analyseq/cgi-bin/freqsq_in.pl.

2.5. Statistics

The statistical analyses were performed with SigmaPlot software, version 9.0. (Systat Software, Inc 2004) using a linear regression test for the GC contents, whereas the XpX frequency and GC content data in the different *MLE* groups were analysed using nonparametric Kruskal–Wallis and Mann–Whitney tests.

2.6. Southern blot

The distribution of *Pacmmar MLEs* in the genome of the two individuals of *P. marmoratus* was checked by Southern blot hybridization (Southern, 1975). The consensus sequence was used to define an enzyme that cut outside the *Pacmmar* sequences in order to estimate the number of *MLEs* present in the genome of this species. Consequently *Apa*I (QBiogen) was used to digest the gDNA. The high level of sequence identity between all *Pacmmar* elements allowed us to use a single radiolabeled probe consisting of the *Pacmar1.1* element. Labeling was performed using the random Nonaprimer Kit (QBiogen) and [α32P] dATP (Perkin-Elmer Life Science). The digested gDNA was migrated in a 0.8% agarose gel, and transferred to a positively charged Nylon membrane (Hybond-N, Amersham). The membrane was hybridized overnight at 65 °C in 0.5 M Na2HPO4–NaH2PO4 pH=7.4, 7% sodium dodecyl sulfate (SDS) with the radiolabeled probe. The membrane was then washed at 65 °C for 30 min in 2× SSC–0.1% SDS (1× SSC=0.15 M NaCl, 0.015 M trisodium citrate), for 1 h in 1× SSC–0.1% SDS, and finally for 30 min in 0.2× SSC–0.1% SDS, after which membranes were exposed overnight to X-ray films with an intensifying screen at −80 °C.

3. Results

3.1. Characterization of Pacmmar element

3.1.1. Analysis of the nucleic acid sequences

Nine full-sized *Pacmmar* elements (AcN AM231069–AM231077) were amplified by PCR using *Bytmar1* 5′ ITR as a primer from the genome of two *P. marmoratus* individuals. All the elements had a total length of between 1329 and 1386 bp, with a duplicated TA dinucleotide at each end. The pairwise comparison of the nine nucleotide sequences indicated identity level between 87 and 99%. A consensus sequence was constructed and called *Pacmmarcons* (Fig. 1) which displayed 99% and 100% identity with *Pacmmar1.1* and *Pacmmar1.2* respectively.

In order to restore a full sequence, the 5′ and 3′ ITRs of *Pacmmar* obtained by inverse PCR were added to the *Pacmmarcons* in place of the ITRs resulting from the amplification using the primer *Bytmar1*. *Pacmmarcons* contained two imperfect ITRs 33-bp in length. *Pacmmar* ITRs are characterized by two kinds of motif: the cleavage signal sequence and the binding site for the transposase. The first motif is located just beside the duplicated TA dinucleotide and corresponds to the consensus motif 5′ YYAGRT 3′ (Langin et al., 1995). *Pacmmar* ITRs have conserved four of the six nucleotides, with G instead of Y and G instead of T at positions 2 and 6 respectively. These differences are also found in *Hsmar2* (Robertson and Martos, 1997) and *Bytmar1* (Halaimia-Toumi et al., 2004). Lampe et al. (2001) have performed a sequence-logo analysis of the ITRs of the various subfamilies of *mariner* element. Two regions (positions 3–8 and 14–18) appear to be conserved, with two nucleotides in particular being most highly conserved (positions 5 and 15). The sequence logos of these regions are 5′ AGGTBK 3′ and 5′ WARRK 3′ respectively. The first region is relatively conserved in *Pacmmar* ITRs, but with T instead of G at position 6. The second region is conserved in the 3′ ITR, whereas there is a G instead of W at position 14 in the 5′ ITR. These differences between the 3′ ITR and 5′ ITR (positions 13, 14, 29, 30, Fig. 1)

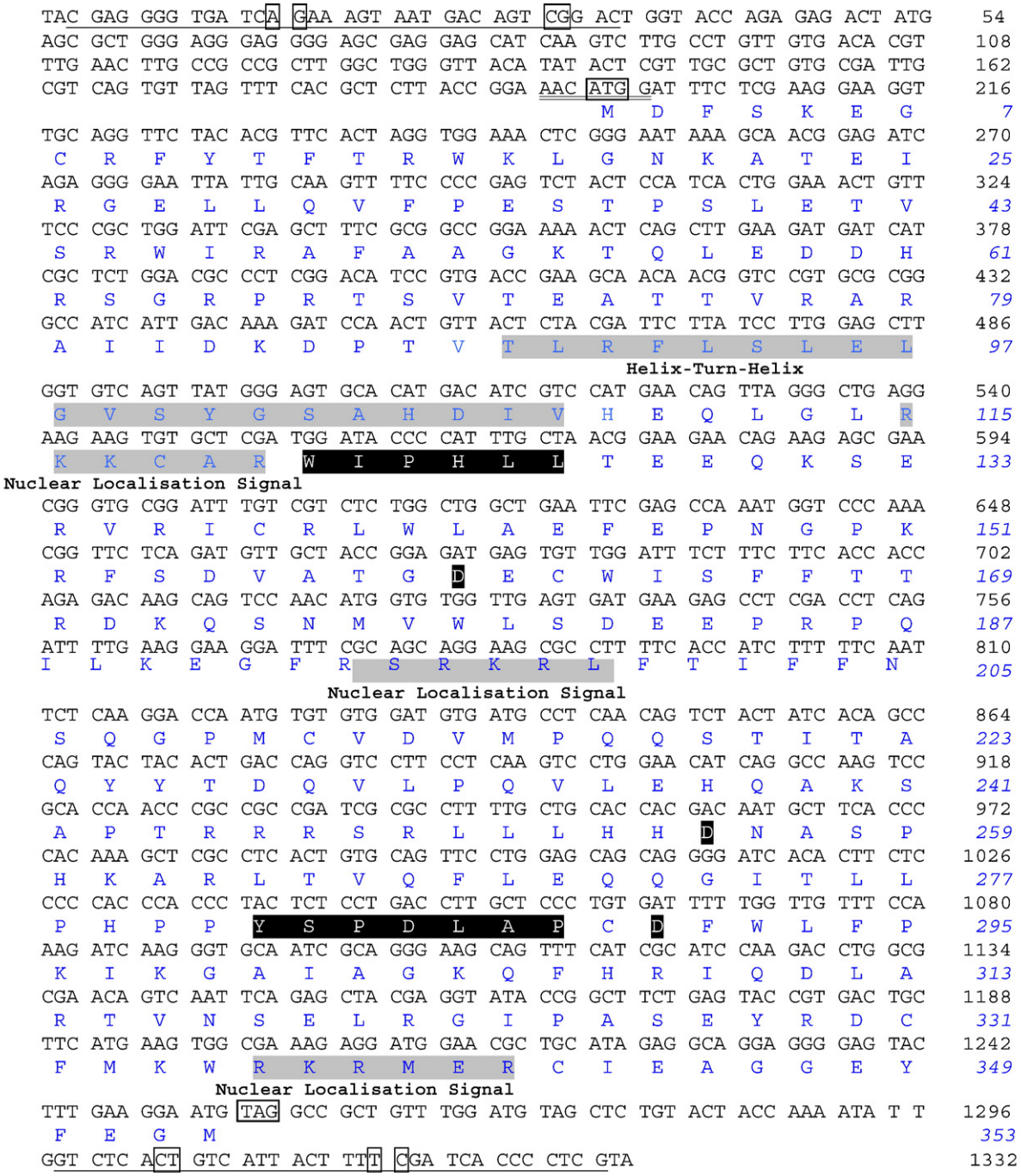

Fig. 1. Nucleotide sequence and amino acid translation of the consensus *Pacmmar MLE*. The ITRs are underlined. Distinct nucleotide between the two ITRs are boxed. The start and the stop codons of the transposase ORF are boxed. The Kozak's sequence is double underlined. The motifs corresponding to *MLE* transposase signatures, amino acid residues of the [D,D34D] triad, and the two conserved motifs, [WV(I)PHE(L)L] and [YSPDLAP], are in black filled boxes with letters typed in white. The helix-turn-helix (HTH) and nuclear location signals (NLS) are in light gray filled boxes. The nucleotide and amino acid positions (in italics) are numbered in the right margin.

could change the binding affinity of the transposase to its ITRs as has already been reported for *Mos*, the transposase of which binds better to *Mos* 3′ ITR than to *Mos* 5′ ITR (Augé-Gouillou et al., 2001).

The 5′ UTR of *Pacmmarcons* was 160-bp long, whereas the 3′ UTR was 40-bp long. The 5′ UTR of *Pacmmarcons* displays a typical Kozak's box (PuXXATGPu) encompassing the putative translational start site at nucleotide 196. However no obvious hallmarks of a eukaryotic promoter (TATA box, CAAT box, GC box) could be detected in the region upstream of the ORF. In addition to *Pacmmarcons*, both *Pacmmar1.1* and *Pacmmar1.2* presented a complete ORF with no stop codon or frameshift, suggesting that these *MLEs* are potentially able to produce a complete transposase. The percentage of identity

Table 1
TpA and CpG calculations for *Pacmmar* elements, compared to average values observed in terrestrial, coastal and hydrothermal *MLEs*

	CpG	TpA
Pacmmar[a]	0.77±0.01	0.55±0.05
Terrestrial *MLEs*[b]	1.11±1.12	0.61±0.13
Other coastal *MLEs*[b]	0.76±0.05	0.55±0.03
Hydrothermal *MLEs*[b]	0.80±0.06	0.45±0.13

[a] Average values of CpG and TpA were calculated from all the *Pacmmar* sequences isolated.
[b] Average values of CpG and TpA were from Halaimia-Toumi et al. (2004) and Casse et al. (2006).

between *Pacmmarcons* or *Pacmmar1.1* on the one hand and *Pacmmar1.2* on the other is 99% and 100% respectively. Furthermore, the seven remaining *Pacmmar* elements all had one to five codons or frameshifts interrupting the ORF.

3.1.2. Nucleotide composition features

The GC content has been calculated for each of the nine *Pacmmar* elements. The average GC content was 50.6%± 1.09, which was very close to the average value previously found in the *MLEs* of two coastal crabs *Cancer pagurus* (50.95±0.53%) and *Maia brachydactyla* (50.92±0.47%; Casse et al., 2006). In contrast, it was significantly higher ($p<0.001$ according to Mann–Whitney test) than that of invertebrate, plant and mammalian *MLEs* (40.5±3.45%), and lower than that of some hydrothermal organisms, the crab *B. thermydron* (54.35±1.05%; Halaimia-Toumi et al., 2004) and the amphipoda *Ventiella sulfuris* (53.64±0.88%; Casse et al., 2006). The GC content of the full-sized elements was highly correlated to the GC content of their respective ORF (approximately 1100 pb), with a linear regression coefficient value of 0.98 ($p<0.001$, $N=9$). This observation shows an equal repartition of the GC content along the element without concentration of GC in *ITRs* and/or *UTRs*. Furthermore, the link between the percentage of GC in the ORF and the codon usage at each of the three positions (GC1, GC2, and GC3) was investigated, and a significant linear correlation was observed between the GC content of ORF and codon position 3 ($R^2=0.79$, $p<0.001$, $N=9$).

Comparison of XpX ratio (Table 1) in *Pacmmar* elements and in other *MLEs* showed that the *Pacmmar* CpG ratio (0.77±0.01) was similar to that observed in coastal and hydrothermal crustaceans, ranging from 0.71 to 0.81 (Halaimia-Toumi et al., 2004; Casse et al., 2006). However, this CpG ratio was significantly lower than that of *MLEs* derived from terrestrial organisms (Table 1) according to the Mann–Whitney test, with $p<0.05$. In contrast, the TpA ratio of *Pacmmar* elements (0.55±0.05) showed significant differences when compared to terrestrial *MLEs* (0.61±0.13) on the one hand and hydrothermal *MLEs* (0.45±0.13) on the other (Table 1).

3.1.3. Amino acid sequence analysis

Following in-silico translation of all *Pacmmar*, nucleic acid sequences showed 51% to 98% identity. As for the nucleic sequences, *Pacmmarcons* and *Pacmmar1.2* displayed 100% identity with each other and 99.4% identity with *Pacmmar1.1*. The three sequences encoded for a potentially functional transposase containing 353 amino acids. The transposases differed by only two amino acids: W instead of G at position 102, and R instead of H at position 109 in *Pacmmar1.1* and *Pacmmarcons* (=*Pacmmar1.2*), respectively. These three *Pacmmar* elements displayed most of the signatures of an *MLE* transposase (Fig. 1); the YSPDLAP motif (Robertson, 1993) was strictly conserved and the WVPHEL motif (Robertson, 1993) was slightly modified in WIPHLL. The catalytic domain included the conserved *MLE* triad [D,D34D] that corresponds to the enzyme core for the DNA cutting and strand transfer reactions required to carry out the transposition process (Lohe and Hartl, 1996). The first D at position 160 was present in a DECW conserved motif, which has been slightly modified from the previously described DEKW or DETW blocks (Shao and Tu, 2001). An FLHDNARPH motif is present in most *MLE* transposases, it was replaced in *Pacmmar* by an LHHDNASPH motif which overlaps the second D in the catalytic triad at position 254. Variations in this motif had already been reported in the nematode *MLE*, *Mcmar1* (Leroy et al., 2003), where it is replaced by the FQQDGARPH motif. The third D was located at position 289, just after the YSPDLAP motif. The N-terminal domain, required for ITR-binding activity, also consists of the HTH domain located from amino acids 88 to 109. Three monopartite putative NLS sequences have been detected; the first, RKKCAR, was located in the DNA-binding domain at positions 115 to 121. The two other NLS sequences, RSRKRL and RKRMER, were found in the carboxyl-terminal domain at positions 192 to 200, and

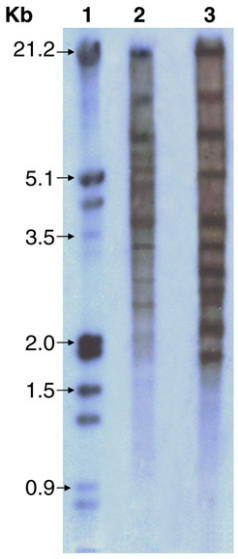

Fig. 2. Southern blot probed with radiolabeled *Pacmmar1.1*. Genomic DNA samples were cut outside the *Pacmmar* sequences using *Apa*I restriction endonuclease, and analysed as described in Section 2.5. Lane 1: molecular weight marker: lambda *Hin*dIII/*Eco*RI; lane 2: gDNA of individual 1; lane 3: gDNA of individual 2. The exposure time was 72 h at room temperature.

336 to 342, respectively. The GRPR-like motif, characteristic of homeodomain proteins and conserved in *Tc1/mariner* transposases (Plasterk et al., 1999), was detected at position 63.

3.2. Distribution in the host genome

The distribution of *Pacmmar MLEs* within the genome of *P. marmoratus* was investigated by Southern blot, using radiolabeled *Pacmar1.1* as the probe. As shown in Fig. 2, numerous bold bands and a feint smear in *Apa*I digests indicated that *MLEs* are probably abundant in the genome of *P. marmoratus*. Analysis of the 3 genomic regions adjacent to the *Pacmmar* insertion sites (Ac N°AM231066–AM231068) did not reveal any similarity with the sequences available in databanks. However, two of these fragments display a repeated sequence approximately 115-bp in length, suggesting that some *Pacmmar* elements may be included in satellite DNA (Fig. 3). *Mboumar*, from the ant *Messor bouveri*, has been recently discovered inserted into satellite DNA (Palomeque et al., 2006). The latter authors suggest that the mobility of the transposable elements in satellite DNA could constitute an important molecular mechanism in the conservation of satellites. The mechanism proposed to explain this hypothesis (Fig. 8 in Palomeque et al., 2006) consisted in an excision process of *MLEs* that brought together satellite DNA monomers that are far from each other. In *Pacmmar*, a TA dinucleotide immediately flanking the ITRs has been observed, which is a typical feature of *mariner* insertion events (Hartl et al., 1997). This finding suggests that *Pacmmar* elements have a transposition mechanism similar to that previously identified for *mariner* elements. As generally reported, it seems that no target DNA sequence specificity is required, with the exception of a dinucleotide TA that is duplicated during the insertion process. Despite the fact that Bigot et al. (1994) reported the existence of a specific degenerated 30-bp palindrome included in a 85-bp conserved region in a hymenopteran species, no specific insertion site for *MLEs* has yet been identified. Vigdal et al. (2002) assumed that target site selection of transposable elements is primarily determined by the structure of the DNA rather than its sequence.

The average GC content of the genomic DNA flanking the *Pacmmar* elements has been calculated to be 55.70±3.13%. This ratio is significantly higher than that observed in the *Pacmmar* elements 50.6±1.09% ($p<0.001$ according to the Mann–Whitney test). Little data is available about the genome of *P. marmoratus* in the databanks, and this makes it impossible to conclude whether this GC content is typical of the gDNA in this particular species, or whether *Pacmmar* elements have an affinity for GC rich sequences.

3.3. Phylogenetic position of Pacmmar

The phylogenetic position of *Pacmmar* in relation to known *mariner* subfamilies and related transposons was determined using the maximum-parsimony method. Four equally parsimonious trees were obtained. They only differed with regard to the resolution of the *Damar1–Himar2–Cpmar1* triad. As shown in Fig. 4, *Pacmmarcons* can obviously be assigned to the *mariner* family (bootstrap support: 85%) and to the *irritans* subfamily (64%). As previously described (Sinzelle et al., 2006), the *irritans* subfamily was found to be divided into four highly supported lineages: *Bytmar1*-like (90%), *Hsmar2*-like (100%), *Himar2*-like (100%), and *Batmar2*-like (87%). The recently-described element *Tvmar1* was included in our analyses, and this decreased the bootstrap support for the *irritans* subfamily to 64% (Fig. 4). The *Tvmar* element was considered to be a new lineage (Silva et al., 2005). It should be noted that eliminating the *Tvmar1* element from the analysis increased the bootstrap support for the *irritans* subfamily to 93%. The *Pacmmarcons* element was introduced into this tree, and formed a monophyletic group with the *Batmar2*-like lineage, but there is very little bootstrap support for this node (<50%). This suggests that *Pacmmarcons* is probably the first representative of a new lineage within the *irritans* subfamily to be detected. In addition, transposase amino acid sequence alignment revealed a DETW motif usually found in the transposases of the *irritans* sub-family, including other "marine" elements such as *Bytmar1* (Halaimia-Toumi et al.,

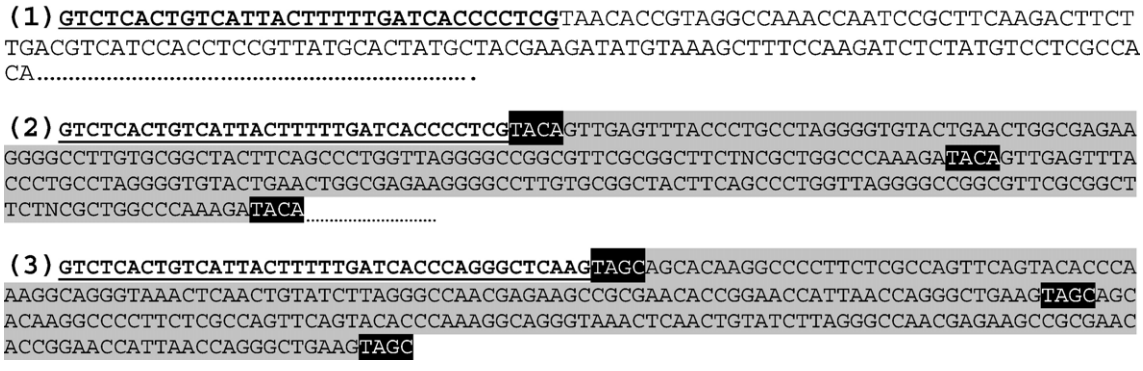

Fig. 3. Sequence of gDNA flanking the 3′ ITR (underlined) of the three *Pacmmar* elements. In both 3′ ITRs (2) and (3), the beginning of three out of five identical repeats is indicated (black boxes), and two full-length sequences out of the five repeats (115 or 116 bp) are shown (gray boxes).

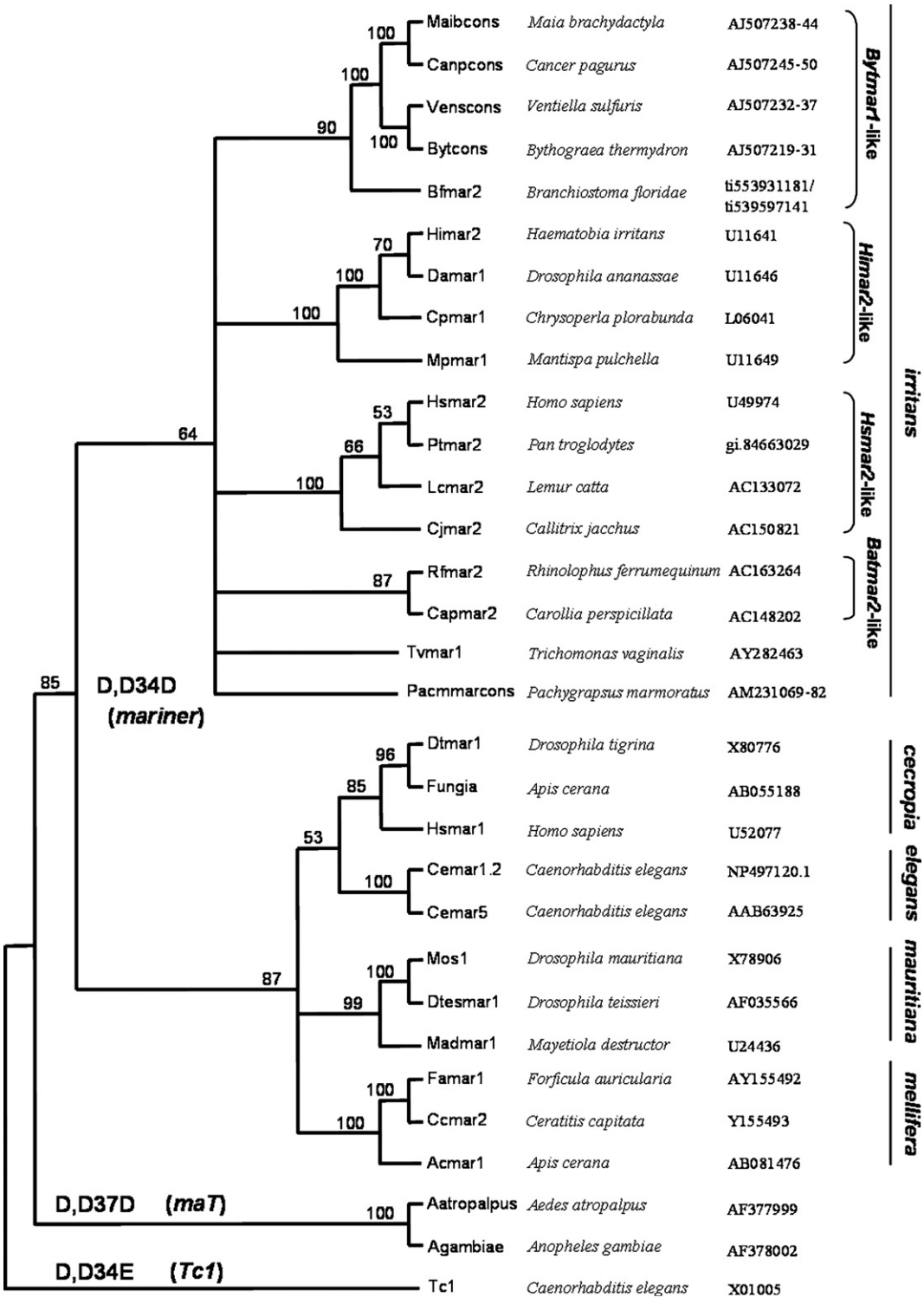

Fig. 4. Phylogenetic position of *Pacmmarcons* in the *Tc1/maT/mariner* superfamily. The consensus tree of the four parsimonious trees is shown (PAUP*, version 4b.10, Swofford, 2002). The tree is rooted using *maT* and *Tc1* sequences. The numbers above branches indicate bootstrap support (1000 repetition). Nodes with bootstrap values of less than 50% are not shown. The species name and *MLE* accession number, and the *mariner* subfamilies are indicated in the right margin.

2004), *Vensmar1*, *Canpmar1* and *Maibmar1* (Casse et al., 2006). However, *Tvmar1* which displays a DESW motif (Claudianos et al., 2002) and *Pacmmar1* which displays a DECW motif, are exceptions to this rule, supporting the hypothesis that each of these new elements constitutes a new lineage in the *irritans* subfamily of *MLEs*.

4. Discussion

Numerous *MLEs* have been characterized in almost all eukaryotic organisms, but only three of them have been shown to be active, including *Mos1*, *Himar1* and *Famar1* (Medhora et al., 1991; Lampe et al., 1999; Barry et al., 2004). Moreover only

8 other *MLEs* with an uninterrupted ORF have been reported, i.e. *Dtmar*, *Ccmar1*, *Crmar2*, *Mcmar1*, *Bytmar1*, *Tvmar1*, *Mboumar* and *Vensmar1* elements (Garcia-Fernandez et al., 1995; Gomulski et al., 1997; Leroy et al., 2003; Gomulski et al., 2001; Halaimia-Toumi et al., 2004; Silva et al., 2005, Palomeque et al., 2006; Casse et al., 2006). This means that the *Pacmmar* element is the twelfth *MLE* with an intact ORF to be reported so far, and the first one to be identified in a marine coastal organism.

From two *P. marmoratus* individuals, we have randomly cloned 9 *Pacmmar* elements two of which contained an uninterrupted ORF, corresponding to an average of ORF proportion of 22%. Eight sequences are less than 10% divergent from the consensus. Such similar sequences are indicative of young elements that have only recently been integrated into the host genome. A great number of *MLEs* was detected in the genome of *P. marmoratus* using Southern blot hybridization, indicating that they are widespread in the host. In order to be integrated into a genome and disperse within it, a transposable element has to be active (Kidwell and Lisch, 2001). All the previous observations support the hypothesis of a recent *MLE* invasion of the *P. marmoratus* genome and suggest the recent activity of *Pacmmar* elements.

The *MLE* sequence of *Pacmmar* is approximately 1300 bp in length, including ITRs of 33 bp, and UTRs of 160 and 40 bp for 5′ and 3′ UTR, respectively. These lengths are typical of *MLEs* (Plasterk et al., 1999), with the exception of those of *Himar*, which have a total length of up to 1500–1600 bp (Lampe et al., 1999), and for *Mcmar1*, which has unexpectedly long perfect ITRs of 355 bp (Leroy et al., 2003). Despite the fact that no obvious RNA-polymerase II promoter element could be found in the 5′ UTR, *Pacmmar* elements do display a typical Kozak's box around the translation start site, suggesting that the downstream ORF could be efficiently translated. These ORFs displayed most of the typical features of a functional *mariner* transposase including the DD34D triad and conserved stretches of amino acids (see Section 3.1.3 for details).

An unusual property of *Pacmmar* results from its nucleotide composition. In fact, we found that the GC content of the *MLEs* of coastal organisms, including *Pacmmar*, is higher than those of the *MLEs* of terrestrial organisms, and slightly lower than those of the *MLEs* of hydrothermal deep-sea species. We did, however, find a significant correlation between the high GC content and the codon usage at position 3 of the ORF encoding for the transposase. As suggested by Halaimia-Toumi et al. (2004), the high GC level could be a result of codon bias adaptation in marine invertebrates. The GC level is linked to DNA stability and bending, and a high GC level may modulate the ability of a transposase to form a synaptic complex.

According to Karlin and Burge (1995), the XpX obs/exp ratio was considered to be under-represented if ≤ 0.78, and over-represented if >1.23. This means that the CpG and TpA ratios found in *Pacmmar* (0.77 ± 0.01 and 0.55 ± 0.06 respectively) should be classified as being under-represented. We have previously shown that the CpG and TpA frequencies in *MLEs* from hydrothermal marine organisms were significantly lower than those of terrestrial *MLEs* (Halaimia-Toumi et al., 2004). Here we report that the CpG and TpA ratios in *Pacmmar* elements are intermediate between terrestrial and hydrothermal *MLEs*. According to the study of Regev et al. (1998), a low CpG ratio could be a consequence of the methylation of cytosine in some crustaceans. TpA shortages have already been reported as an usual feature of transposons (Karlin and Burge, 1995), but little is known about the particular case of the *MLEs*. Lerat et al. (2002) suggested that this TA under-representation in *MLEs* avoids the presence of too many transcription signals and stop codons (TAA or TAG) and then allows to select mRNA. The low TpA ratio of *Pacmmar* could also be the result of an evolutionary process intended to limit the insertion of TA targeting transposons, including *MLEs*, into the *Pacmmar* element itself. If this is true, then the shortage of TpA in *MLEs* could be driven by a self-protection mechanism.

Phylogenetic analysis clearly shows that *Pacmmar* is a *mariner* transposon belonging to the *irritans* subfamily, but it does not cluster with other marine *MLEs* isolated from hydrothermal and coastal crabs (Halaimia-Toumi et al., 2004 and Casse et al., 2006). Based on tree topology and bootstrap support, *Pacmmar* is probably the first representative of a new lineage within the *irritans* subfamily: the *Pacmmar1*-like elements.

The large number of copies of *Pacmmar* in the genome of *P. marmoratus*, the low degree of polymorphism among them, indicating a recent amplification of the element in the host genome, and the presence of elements containing an ORF in the two individuals analysed all suggest that *Pacmmar* could be an active element. To check this hypothesis, transposition assays would be required. The characterization of the *Pacmmar ITR* sequences by inverse PCR, which is reported here, has provided the set of data required before efficient transposition assays can be undertaken. Indeed, Lampe et al. (2001) have demonstrated that a relatively small amount of divergence in *ITR* sequences is sufficient to substantially reduce the transposase/*ITR* interaction.

In conclusion, the *Pacmmar* elements described here, and characterized by their high GC content, constitute a new lineage of *MLEs* in the *irritans* subfamily of the *mariner* family, and are potentially active in the genome of *P. marmoratus*.

Acknowledgements

This work was supported by a PhD grant from Agence Universitaire de la Francophonie to Q-T. B., and L.D. was the recipient of a PhD grant from Région Pays de la Loire. The authors are grateful to Monika Ghosh for revising the English text.

References

Augé-Gouillou, C., Hamelin, M.H., Demattei, M.V., Periquet, M., Bigot, Y., 2001. The wild-type conformation of the *Mos-1* inverted terminal repeats is suboptimal for transposition in bacteria. Mol. Genet. Genomics 265, 51–57.

Augé-Gouillou, C., Brillet, B., Germon, S., Hamelin, M.-H., Bigot, Y., 2005. *Mariner Mos1* transposase dimerizes prior to ITR binding. J. Mol. Biol. 351, 117–130.

Barry, E.G., Witherspoon, D.J., Lampe, D.J., 2004. A bacterial genetic screen identifies functional coding sequences of the insect *mariner* transposable

element *Famar1* amplified from the genome of the Earwig, *Forficula auricularia*. Genetics 166, 823–833.

Bigot, Y., Hamelin, M.H., Capy, P., Periquet, G., 1994. *Mariner*-like elements in hymenopteran species: insertion site and distribution. Proc. Natl. Acad. Sci. U. S. A. 91, 3408–3412.

Capy, P., Anxolabéhère, D., Langin, T., 1994a. The strange phylogenies of transposable elements: are horizontal transfer the only explanation? Trends Genet. 10, 7–12.

Capy, P., et al., 1994b. Horizontal transmission versus ancient origin: *mariner* in the witness box. Genetica 93, 161–170.

Casse, N., Pradier, E., Loiseau, C., Bigot, Y., Laulier, M., 2000. *Mariner*, a mobile DNA transposon in the genomes of several hydrothermal invertebrates. InterRidge News 9, 15–17.

Casse, N., Pradier, E., Demattei, M.V., Bigot, Y., Laulier, M., 2002. *Mariner* transposons are genetic parasites occurring in the genome of hydrothermal invertebrates. Cah. Biol. Mar. 43, 325–328.

Casse, N., Bui, Q.T., Nicolas, V., Renault, S., Bigot, Y., Laulier, M., 2006. Species sympatry and horizontal transfers of *Mariner* transposons in marine crustacean genomes. Mol. Phylogenet. Evol. 40, 609–619.

Claudianos, C., Brownlie, J., Russell, R., Oakeshott, J., Whyard, S., 2002. *maT*-A clade of transposons intermediate between *mariner* and *Tc1*. Mol. Biol. Evol. 19, 2101–2019.

Feschotte, C., Wessler, S.R., 2002. *Mariner*-like elements are widespread and diverse in flowering plants. Proc. Nat. Acad. Sci. U. S. A. 99, 280–285.

Garcia-Fernandez, J., Bayascas-Ramirez, J.R., Marfany, G., Munoz-Marmol, A.M., Casali, A., Baguna, J., Salo, E., 1995. High copy number of highly similar *mariner*-like transposons in planarian (Platyhelminthe): evidence for trans-phyla horizontal transfer. Mol. Biol. Evol. 12, 421–431.

Gomulski, L.M., Torti, C., Malacrida, A.R., Gasperi, G., 1997. *Ccmar1*, a full-length *mariner* element from the Mediterranean fruit fly, *Ceratitis capitata*. Insect Mol. Biol. 6, 241–253.

Gomulski, L.M., et al., 2001. A new basal subfamily of *mariner* element in *Ceratitis rosa* and other tephrirtid flies. J. Mol. Evol. 53, 597–606.

Halaimia-Toumi, N., et al., 2004. The GC-rich transposon *Bytmar1* from the deep-sea hydrothermal crab, *Bythograea thermydron*, may encode three transposase isoforms from a single ORF. J. Mol. Evol. 59, 747–760.

Hartl, D.L., 1989. Transposable element *mariner* in *Drosophila* species. In: Berg, D.E., Howe, M.M. (Eds.), Mobile DNA. ASM Press, Washington, DC, pp. 531–536.

Hartl, D., 2001. Discovery of the transposable element *mariner*. Genetics 157, 471–476.

Hartl, D.L., Lohe, A.R., Lozovskaya, E.R., 1997. Modern thoughts on an ancient mariner: function, evolution, regulation. Ann. Rev. Genet. 31, 337–358.

Hua-Van, A., Le Rouzic, A., Maisonhaute, C., Capy, P., 2005. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. Cytogenet. Genome Res. 110, 426–440.

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

Karlin, S., Burge, C., 1995. Dinucleotide relative abundance extremes: a signature. Trends Genet. 11, 283–290.

Kidwell, M.G., Lisch, D.R., 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. Evol. Int. J. Organ. Evol. 55, 1–24.

Kvamme, B.O., Kongshaug, H., Nilsen, F., 2005. Organisation of trypsin genes in the salmon louse (*Lepeophtheirus salmonis*, Crustacea, copepoda) genome. Gene 352, 63–74.

Lampe, D.J., Akerley, B.J., Runbin, E.J., Mekalanos, J.J., Robertson, H.M., 1999. Hyperactive transpsase mutants of the *Himar1 mariner* transposon. Proc. Natl. Acad. Sci. U. S. A. 96, 11428–11443.

Lampe, D.J., Walden, K.K., Robertson, H.M., 2001. Loss of transposase–DNA interaction may underlie the divergence of mariner family transposable elements and the ability of more than one mariner to occupy the same genome. Mol. Biol. Evol. 18, 954–961.

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

Langin, T., Capy, P., Daboussi, M.J., 1995. The transposable element Impala, a fungal member of the Tc1-*mariner* superfamily. Mol. Gen. Genet. 246, 19–28.

Laulier, M., Pradier, E., Bigot, Y., Periquet, G., 1995. An easy method for preserving nucleic acids in field samples for later molecular and genetic studies without refrigerating. J. Evol. Biol. 8, 657–663.

Lerat, E., Capy, P., Biémont, C., 2002. The relative abundance of dinucleotides in transposable elements in five species. Mol. Biol. Evol. 19 (6), 964–967.

Leroy, H., Castagnone-Sereno, P., Renault, S., Augé-Gouillou, C., Bigot, Y., Abad, P., 2003. Characterization of *Mcmar1*, a *mariner*-like element with large ITR from the phytoparasitic nematode *Meloidogyne chitwoodi*. Gene 304, 35–41.

Lohe, A.R., Hartl, D.L., 1996. Autoregultation of *mariner* transposase activity by overproduction and dominant-negative complementation. Mol. Biol. Evol. 13, 549–555.

Medhora, M., Maruyama, K., Hartl, D.L., 1991. Molecular and functional analysis of the *mariner* mutator element *Mos1* in *Drosophila*. Genetics 128, 311–318.

Palomeque, T., Carillo, J.A., Munoz-Lopez, M., Lorite, P., 2006. Detection of a mariner-like element and miniature inverted-repeat transposable element (MITE) associated with the heterochromatin from ants of the genus Messor and their possible involvement for satellite DNA evolution. Gene 371, 194–205.

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

Regev, A., Lamb, M.J., Jablonka, E., 1998. The role of DNA methylation in invertebrate: developmental or genome defense? Mol. Biol. Evol. 15, 880–891.

Richardson, J.M., Dawson, A., O'Hagan, N., Taylor, P., Finnegan, D.J., Walkinshaw, D., 2006. Mechanism of *Mos-1* transposition: insights from structural analysis. EMBO 1–11.

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

Robertson, H.M., 1997. Multiple *mariner* transposons in flatworms and hydras are related to those of insects. J. Heredity 88, 195–201.

Robertson, H.M., Lampe, D.J., 1995. Distribution of transposable elements in arthropods. Annu. Rev. Entomol. 40, 333–357.

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

Robertson, H.M., Martos, R., 1997. Molecular evolution of the second ancient human transposon, Hsmar2, illustrates pattern of neutral evolution in the human genome lineage. Gene 205 (1–2), 219–228.

Robertson, H.M., 2002. Evolution of DNA transposons in eukaryotes. Mobile DNA II. ASM Press, Washington DC, pp. 1093–1110.

Rouleux-Bonnin, F., Petit, A., Demattei, M.V., Bigot, Y., 2005. Evolution of full-length and deleted forms of the *mariner*-like element, *Botmar1*, in the Genome of the Bumble Bee, *Bombus terrestris* (Hymenoptera: Apidae). J. Mol. Evol. 59, 736–747.

Shao, H., Tu, Z., 2001. Expanding the diversity of *IS630-Tc1-mariner* superfamily: discovery of a unique DD37E transposon and reclassification of DD37D and DD39D transposons. Genetics 159, 1103–1115.

Silva, J.C., Loreto, E.L., Clark, J.B., 2004. Factors that affect the horizontal transfer of transposable elements. Curr. Issues Mol. Biol. 6, 57–72.

Silva, J.C., Bastida, F., Bidwell, S.L., Johnson, P.J., Carlton, J.M., 2005. A potentially functional mariner transposable element in the protist *Trichomonas vaginalis*. Mol. Biol. Evol. 22, 126–134.

Sinzelle, L., Chesneau, A., Bigot, Y., Mazabraud, A., Pollet, N., 2006. The *mariner* transposons belonging to the irritans subfamily were maintained in chordate genomes by vertical transmission. J. Mol. Evol. 62 (1), 53–65.

Southern, E.M., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503–517.

Swofford, D., 2002. PAUP* (Phylogenetic Analysis Using Parsimony). Sinauer Associates, Sunderland.

Vigdal, T.J., Kaufman, C.D., Izsvak, Z., Voytas, D.F., Ivics, Z., 2002. Common physical properties of DNA affecting target site selection of sleeping beauty and other *Tc1/mariner* transposable elements. J. Mol. Biol. 323, 441–452.

Wang, H., Hartswood, E., Finnegan, D.J., 1999. *Pogo* transposase contains a putative helix-turn-helix DNA binding domain that recognises a 12 bp sequence within the terminal inverted repeats. Nucleic Acids Res. 27, 455–461.