ORIGINAL ARTICLE

CrossMark

An Analysis of *IS630/Tc1/mariner* Transposons in the Genome of a Pacific Oyster, *Crassostrea gigas*

M. V. Puzakov[1] · L. V. Puzakova[1] · S. V. Cheresiz[2,3]

Abstract

Transposable elements represent the DNA fragments capable of increasing their copy number and moving within the genome. Class II mobile elements represents the DNA transposons, which transpose via excision and the subsequent reinsertion at random genomic loci. The increase of their copy number occurs only when the transposition event is coupled with the replication. *IS630/Tc1/mariner* DNA transposon superfamily is one of the largest and widely distributed among the Class II elements. In this work, we provide a detailed analysis of *IS630/Tc1/mariner* DNA transposons from the Pacific oyster, *Crassostrea gigas*. *IS630/Tc1/mariner* transposons represented in the genome of the Pacific oyster belong to four families, *Tc1* (DD34E), *mariner* (DD34D), *pogo* (DDxD), and *rosa* (DD41D). More than a half of *IS630/Tc1/mariner* elements from *C. gigas* belong to *Tc1* family. Furthermore, *Mariner-31_CGi* element was shown to represent a new and previously unknown family with DD37E signature. We also discovered the full-size transcripts of eight elements from *Tc1, mariner*, and *pogo* families, three of which can, presumably, retain their transposition activity.

Introduction

Transposable elements (TE) represent the DNA fragments capable of increasing their copy number and moving within the genome. They constitute a considerable part of prokaryotic or eukaryotic genomes, accounting for up to 35–69% of genomic DNA in different mammalian species (Lander et al. 2001; Waterston et al. 2002; de Koning et al. 2011) and up to 90% of the genome in some plants (Sergeeva and Salina 2011).

✉ M. V. Puzakov
puzakov@ngs.ru

[1] The A.O. Kovalevsky Institute of Marine Biology Research of RAS, Nakhimov av., 2, Sevastopol, Russia 299011

[2] Department of Medicine, Novosibirsk State University, Pirogova st., 1, Novosibirsk, Russia 630090

[3] State Scientific Research Institute of Physiology and Basic Medicine, P.O. Box 237, Novosibirsk, Russia 630117

Transposable elements exert a considerable influence on genome function by altering structure and activity of the genes due to the genetic rearrangements (Strand and McDonald 1989; Zakharenko et al. 2000). The TE-induced genomic instability has a dual significance, as far as either a particular organism or the common population is concerned. For the individuals, the increased mutability most often results in lethality, while for the population, especially for the one under stress, an increased mutational background offers an ability to gain new beneficial features (Chow and Tung 2000; Del Re et al. 2003; Zakharenko et al. 2006), which can result in the adaptation and further evolution (Finnegan 1992; Piacentini et al. 2014).

Depending on the mechanism of transposition, the TEs are divided into two classes (Kapitonov and Jurka 2008). Class I includes the retrotransposons, which first construct their RNA intermediate copy and then the DNA copy to be inserted into the host genome. Transposing by this mechanism, the retrotransposons can easily increase their copy number and, thus, they account for the majority of the entire transposable genome.

Class II represents the DNA transposons, which transpose via excision and the subsequent reinsertion at random genomic loci. The increase of their copy number occurs

only when the transposition event is coupled with the replication. The proportion of DNA transposon-derived sequences in the genome of a species rarely prevails over the proportion of retrotransposon-derived sequences (Arensburger et al. 2010; Mesquita et al. 2015; Fernández-Medina et al. 2016; Hernandez-Hernandez et al. 2017).

The class of DNA transposons is further divided into the three subgroups: TIR-transposons, helitrons, and polintons. TIR-transposons constitute the most representative and diverse subclass, which includes 21 superfamilies (Kapitonov and Jurka 2008). The presence of terminal inverted repeats (TIR) is a feature of TIR-transposons. The internal portion of a transposon flanked by the TIRs includes the open reading frame for the transposase enzyme responsible for the TE mobility. During the excision, the transposase recognizes the inverted terminal repeats, which are specific to a particular transposon. The TEs, which have a changed TIR unrecognizable by transposase or have lost their TIRs, become immobilized (Wicker et al. 2007).

One of the most diverse and widely distributed TIR-transposon superfamilies is *IS630/Tc1/mariner*. The first representatives of *mariner* family (*Dmmar1* и *Mos1*) were isolated from *Drosophila mauritiana* (Haymer and Marsh 1986; Jacobson et al. 1986). Similar elements were later found in almost every species studied for their presence (Feschotte and Pritham 2007; Liu and Yang 2014).

The genome size of *IS630/Tc1/mariner* transposons varies between 1 and 3000 bp. These elements usually have a single ORF coding for the transposase (~ 350 amino acids (a.a.) in length). The transposase recognizes the TIRs and mediates the TE insertion into the duplicated target site TA. The length of terminal inverted repeats varies between 20 and 460 bp (Shao and Tu 2001; Claudianos et al. 2002).

Despite the fact that the number of species with fully sequenced genomes is increasing annually, the TE representation and activity across the marine species, including the invertebrates, remains rather poorly studied (Puzakov et al. 2017). According to the data obtained through the automatic annotation services, the content of TE-derived sequences in the genomes of mollusks is relatively lower, than that in the vertebrate genomes (35–52%). In the former, this figure varies between 2 and 8% (Yoshida et al. 2011; Takeuchi et al. 2012; Zhang et al. 2012; Simakov et al. 2013; Albertin et al. 2015). The Pacific oyster, *Crassostrea gigas*, belongs to a large, while scarcely studied genetically, phylum of the mollusks (Mollusca). The sequencing revealed that its genome is rather polymorphic and it harbors the repetitive sequences, including the TEs (Zhang et al. 2012). In this work, we present the results of a detailed analysis of *IS630/Tc1/mariner* family transposons in *C. gigas*.

Materials and Methods

In this work, the DNA transposons of *IS630/Tc1/mariner* superfamily found in the Pacific oyster, *C. gigas* (Jurka 2012, 2013; Bao and Jurka 2013a, b, c), are studied. Additionally, the sequences of the other 53 *IS630/Tc1/mariner* superfamily DNA transposons representing *Tc1, pogo, mariner, maT, mosquito, plants, IS630*, and *rosa* families were used for the phylogenetic analysis and DDE/D domain analysis (Online Resource 1).

The terminal inverted repeats were determined by *blastn* algorithm (Zhang et al. 2000). The presence of functional amino acid sequences was determined by *blastx* on the basis of their homology to the transposases of known transposable elements (Altschul et al. 1997). The copy number *C. gigas* elements was calculated as the number of homologs exceeding 10% of the transposon size using a *blastn* instrument (Zhang et al. 2000). Both the overall number of homologous loci and the number of full-length elements (those retaining > 95% of sequence length and the intact TIRs) were counted.

Multiple alignment of amino acid sequences was made by MUSCLE software (Edgar 2004) using the standard settings. The appropriate amino acid substitution model was selected using the Akaike information criterion. The search best protein models and phylogenetic analysis were performed using MEGA7.0 software (Kumar et al. 2016). The activity of the DNA transposons was evaluated by using the presence of transcripts homologous to the predicted transposase sequences as a proxy. The search for transcripts in the TSA and EST databases was performed with the use of *tblastn* (Altschul et al. 1997).

In order to estimate the evolution dynamics of the elements from *C. gigas*, we determinate the identity percentage of the consensus sequences of elements with their corresponding full-size copies. Consensus sequences were derived using the relative majority rule. This approach was previously used by Bouallègue et al. (2017).

Results and Discussion

IS630/Tc1/mariner Elements of *C. gigas*

Repbase contains 31 potentially functional (autonomous) and 21 non-autonomous *IS630/Tc1/mariner* superfamily elements from *C. gigas* (Online Resource 2). We also found and uploaded to Repbase six novel elements of this superfamily (*Mariner-32_CGi, Mariner-33_CGi, Mariner-34_CGi, Mariner-36_CGi, Mariner-37_CGi, Mariner-38_CGi*). For two elements from Repbase

(*Mariner-2_CGi, Mariner-N1_CGi*), we found the more complete transposase-encoding sequences and, thus, the non-autonomous *Mariner-N1_CGi* transposon was re-classified and renamed as the autonomous *Mariner-35_CGi*.

The majority of the autonomous *IS630/Tc1/mariner* DNA transposons from the oyster have the sequence length of 1–3 kbp, which is typical for this superfamily. Five elements (*Mariner-38_CGi, Mariner-31_CGi, Mariner-35_ CGi, Mariner-9_CGi* и *Mariner-34_CGi*) have sequence length between 3 and 4.5 kbp, while *Mariner-21_CGi* has a size significantly larger than the typical one (9744 bp). The increased length of TE sequences may be due to the insertions into its sequence or the duplications. Any of these events can result in the loss of the TE autonomous motility, as long as the ORF, the promoter, or the TIRs are damaged. The lengths of ORFs encoded by the majority of *IS630/Tc1/mariner* elements from *C. gigas* lie within the range of 300–400 a.a., which is a standard value for this group of DNA transposons. In ten elements, the length of putative transposase is > 450 a.a., while in three elements— *Mariner-3_CGi, Mariner-4_CGi*, and *Mariner-6_CGi*—the transposase is shorter than 250 a.a. The length of TIRs of *IS630/Tc1/mariner* DNA transposons from the oyster typically varies between 24 and 52 bp. Although the TIRs of *Mariner-31_CGi* and its non-autonomous copy are somewhat longer (173 and 273 bp, respectively), they are not beyond the scope described for this superfamily (20–460 bp) (Claudianos et al. 2002).

Analysis of Catalytic Domains

The classification of DNA transposons is based on the differences in the catalytic domain of the transposase, which is responsible for the cutting and relegation of DNA chains in the process of insertion. This domain has three characteristic amino acid residues—two aspartates (D) and either glutamate (E) or aspartate as the third one, hence the name of this triad, the DDE/D motif or domain. Based on the different number of amino acid residues between the second aspartate and the third glutamate/aspartate in DDE/D domains, *IS630/Tc1/mariner* superfamily is divided into eight groups: *Tc1* (DD34E) (Henikoff 1992; Capy et al. 1996), *mariner* (DD34D) (Robertson 1995; Capy et al. 1996), *pogo* (DDxD) (Smit and Riggs 1996; Capy et al. 1997), *maT* (DD37D) (Robertson and Asplund 1996; Shao and Tu 2001; Claudianos et al. 2002), *mosquito* (DD37E) (Shao and Tu 2001), *plants* (DD39D) (Jarvik and Lark 1998; Shao and Tu 2001), *rosa* (DD41D) (Gomulski et al. 2001; Zhang et al. 2016b) и *IS630* (DDxE) (Doak et al. 1994; Capy et al. 1996).

DDE/D domains were identified, virtually, in every *IS630/Tc1/mariner* transposon from *C. gigas* (Online Resource 2). The protein sequence of *Mariner-17_CGi* was not represented in Repbase, probably due to the multiple

stop codons in the putative ORF; however, the catalytic triad (DD34E) still remained clearly recognizable. For *Mariner-9_CGi*, a fragmented ORF with its fragments lying between the nucleotide positions 1065–1319, 1323–1832, and 1660–2853 according to the nucleotide sequence of the element is represented in Repbase. This ORF is coding for a protein sequence of 653 a.a. in length. However, our further analysis suggests that another ORF (with its fragments localized between the positions 1065–1658 and 1660–2856) coding for a protein 596 a.a. in length is more likely to represent the gene sequence of a putative transposase. *Mariner-31_CGi* contains two ORFs, and two independent amino acid sequences—Mariner-31-CGi_p1 and Mariner-31-CGi_ p2—were analyzed.

The fragments of identified DDE/D domains including the second and third amino acid residues of DDE/D triad and 10 a.a. both upstream and downstream of this region were aligned with the similar fragments of the transposases of the known *IS630/Tc1/mariner* superfamily transposons (Fig. 1). The group of DNA transposons (18 elements) with the ORFs coding for the amino acid sequences with DD34E catalytic domain was the largest one. As a result of multiple alignment, those DNA transposons were shown to have a high degree of homology with the known elements of *Tc1* family. *Mariner-35_CGi* (DD36E) element has a DDE/D domain highly homologous to those of DD34E elements and, most possibly, it also belongs to *Tc1* family.

The second large group of DNA transposons (ten elements) codes for the proteins with catalytic domains typical for *pogo* family—DD30D (*pogo*-like) and DD35D (*Fot1*-like).

The three groups of DNA transposons (with DD34D, DD41D, and DD37E domains) included only two elements each. The elements with DD34D (*Mariner-2_CGi, Mariner-11_CGi*) and DD41D (*Mariner-4_CGi, Mariner-12_ CGi*) domains show similarities with the *mariner* and *rosa* groups of transposons, respectively. The situation with the elements with DD37E domain (*Mariner-18_CGi, Mariner-31_CGi*) turned out to be more intriguing. For both elements, the pattern (disposition) of amino acids in the alignment is different from the representatives of *mosquito* family (*Ae-atropalpus1* и *An-gambiae1*). In addition, three amino acid residues (SEV) of *Mariner-31_CGi* are localized to a completely different locus in the alignment than in all other elements of the *IS630/Tc1/mariner* superfamily (Fig. 1). *Mariner-18_CGi* and *Mariner-31_CGi* elements can belong to the new, unique clades of elements with DD37E signature or represent the extensively modified versions of DD34E or other elements.

No elements with DD37D, DD39D, or DDxE domains have been found among the *IS630/Tc1/mariner* transposons of *C. gigas*. To date, the DDxE elements (*IS630* family) have been found only in prokaryotes (Doak et al. 1994; Capy et al.

```
          Mariner-1_CGi    RHQNHITLQQDNARPHVAR---VVRDLFVQQ------N-VDVL-P---WPAVSPDLSPIE-HVWDEMERRL-
          Mariner-5_CGi    FPKEQYIFQDDNAPVHRAR---STQDFIHRN------G-IRTM-S---WPAQSPDLNIIE-NLWLLLKRKL-
          Mariner-7_CGi    FPQNGYFFQDDNAPVHRAR---STQEYVARN------R-INCL-S---WPAQSPDLNIIE-NVWLFIKRKL-
          Mariner-8_CGi    FPHGNYVFQDDNAPVHRSR---LLSAYMEEN------N-INTT-S---WPAQSPDINIIE-NIWLKLKRHL-
          Mariner-14_CGi   FPNNQYIFQDDNAPVHRAR---VVKDFVTRE------G-ITTL-E---WPAQSPDLNIIE-NCWKKMKHEI-
          Mariner-15_CGi   TNDLPYVLQQDNARPHTAA---YTRDWFTQN------H-VTVM-E---WPAASPDLNPIE-NVWQIMKDRV-
          Mariner-16_CGi   FSDSPWMLQEDNCPVHQSR---QTVAWKDEN------N-IDTL-T---WPSQSPDINIIE-NVWRTIKIRL-
          Mariner-17_CGi   LPXGSYLFQDDNAPVHRSR---LTQEFVARN------R-IKNM-S---WPAQSPDLNIIE-NVWLYIKRKL-
          Mariner-19_CGi   FSDKPYLFQDDNAPVHRAR---ITCEYKVNN------N-IPSI-T---WPAQSPDINIIE-NIWLKIKRSL-
          Mariner-20_CGi   FGNNTWYFQDDNAPCHRSN---ESENWKHQN------R-IPQI-S---WPPQSPDPSPIE-NVWLLMKNKI-
DD34E     Mariner-22_CGi   FGNLPWHFMDDNCPVHRSR---ETENWKMRN------D-IPPF-F---WPPQSPDLNPIE-NVWLVLKNHV-
          Mariner-23_CGi   FSNKPYMFQDDNAPPHASR---MTKAWKEXN------N-IPTL-T---XPAQSPDVNIIE-NIWLLLLKNQV-
          Mariner-25_CGi   RQGNARILQQDNARPHTAR---HTQTTLQQN------N-IVTL-D---WPARSPDLSPIE-HVWDILGRRL-
          Mariner-26_CGi   TNGPGLTFQQDNATAHTAR---ATREFLNQN------N-VDML-D---WPSKSPDLSPIE-HVWDELGRRV-
          Mariner-27_CGi   INTQREVFQHDNARPHTAR---ATVDFLADH------N-VTVL-P---WPSRSPDLNPIE-HLWDQLDKRV-
          Mariner-29_CGi   FPNRPWIFQEDNAPCHVSA---RANAWKITN------N-ISTL-P---WPAQSPDLNIIE-NVWKTLKTRI-
          Mariner-32_CGi   IGEDGHRFQQDNDPKHTSK---LTKKFMADN------N-INWWNE---WPA-SPDLNPIE-MVWSELKRCV-
          Mariner-33_CGi   GYGDNNFYMDDNAPCHRSK---VVKDWMANN------N-LRSL-I---WPPQSPDLNPIE-NLWRDLGIAK-
          Impala           FLQSGDIFMHDNASVHTAR---IVKALLEEL------G-VDLM-T---WPPYSPDLNPIE-NLWALMKAEI-
          Tc1              NVGRGFVFQQDNDPKHTSL---HVRSWFQRR------H-VHLL-D---WPSQSPDLNPIE-HLWEELERRL-
          Bari-1           FPTTEWILQQDNAPCHKGR---IPTKFLNDL------N-LAVL-P---WPPQSPDLNIIE-NVWAFIKNQR-

DD36E     Mariner-35_CGi   KFPNGFRIYQDNDSKHVSK---STKEWMQQK------NLLDKIMK---TPASSPDINPIE-NVWAAMKLYL-

          Mariner-2_CGi    VDLGNIILHLDNAPAHTAA---STCLEIELL------G-FNLL-E---HPPYSPDLAPMDFAIFPFIKSQ--
          Mariner-11_CGi   ANKENVIYHQDNAPSHTAE---QTNLELDLL------G-FSRI-S---HPPYSPDLAPMDFALFPYLKAK--
          Dmmar1           KRQHRVIFLHDNAPSHTAR---AVRDTLETL------N-WEVL-P---HAAYSPDLAPSDYHLFASMGHA--
DD34D     Hsmar1           VNRKGPILLHDNARPHVAQ---PTLQKLNEL------G-YEVL-P---HPPYSPDLSPTDYHFFKHLDNF--
          Famar1           TNRKGVVFHHDNARPHTSL---VTRQKLLEL------G-WDVL-P---HPPYSPDLAPSDYFLFRSLQNS--
          Ccmar1           ANRKGVLFHYDNARPHVAK---PTLAKLKEM------N-WEIM-P---HSPYSPDIAPSDYHLFRSLQNN--
          Cpmar1           MKKKKVLFHQDNAPCHKSL---RTMAKIHEL------G-FELL-P---HPPYSPDLAPSDFFLFSDLKRM--

          Mariner-9_CGi    PSHRPVLLIMDGHSSHISL---EILQVAKEN------E-IEILCL---PPHTTSELQPLDKCVFKPLKTE--
          Mariner-24_CGi   GDARPQLLILDGHSSHESL---AILESAIQN------D-IHIISL---PPHTTHALQPLDRSVFGPMNTA--
          Mariner-28_CGi   GPKRPQLLILDSHSSHEVL---GLLQEAVRE------N-IIILAM---PPHTSHHLQPLDKCGFGPFSHA--
          Mariner-30_CGi   RINQKAIAIFDIFAAHRTN---SLIEKLKSH------N-IQPLYV---PGACTDKLQPLDVAVNYDYKEK--
          Mariner-34_CGi   GPHRPQLLILDSHRSHEVL---GLLQMAKAN------G-IHILAL---PPHTTAWLQPLDRTVFGPLSPS--
          Mariner-38_CGi   GPERPQLLILDGHGSHETL---GLLELAAEE------N-IHILAL---PPHTTHFLQPLDRTVFGPFNKA--
          Fot1             DPADARLLIVDGHGSHATE---QFMAKCYLN------N-VYLLFL---PAHCSHVLQPLDLGCFSSLKAA--
DDxD      Tan1             SVGKYSLLVLDGHGSHLTP---EFDQSCAEN------E-VIPICM---PAHSSHLLQPLDVGCFSVLKRT--
          Flipper          RPKEARLLVLDGHGSHETT---QFMLECFKN------N-IHLLFL---PPHTSHVLQPPDLSIFSPLKKE--
          Mariner-10_CGi   KKKRHILMFLDNAASHSHE---LRLSHVQ----------LKFL-----PANTTSVLQPLDQGIIRAFKAR--
          Mariner-13_CGi   RKKRKIALIVDNCPAHP-----KIAGLKAVE--------LVFL-----PPNTTSKTQPMDQGIIQNLKVH--
          Mariner-36_CGi   SQNRSILLFVDNCAAHP-----TVDKLRAIK--------LAFL-----PPNTTSKLQPMDQGIIYNMKVK--
          Mariner-37_CGi   RRRRKVAMIVDNCPAHP-----KVKDLKAIT--------LFFL-----PPNTTSKTQPMDQGIIQNLKIH--
          pogoR11          KQNRKILLFIDNATSHT-----TVKDFENIK--------LCFM-----PPNATALLQPLDQGIIHSFKLE--
          Tigger1          KISFKILLLIDNAPGHPRA---LMEMYKEIN--------VVFM-----PANTTSILQPMDQGVISTFKSY--

          Mariner-18_CGi   QYPGHATFVQDGATPHTAL---VTQNWCKDN----LPN-FISKEE---WPGNSPDLNCIE-NLWSILDSEA-
DD37E     Mariner-31_CGi   VLQRGDSLIVDNAPTHRNMSEVVLRNWLPTI------G-VQYI-F---LPTYSPDLNPAE-LCFRKVKILL-
          Ae-atropalpus1   AHDRPVMLWPDLASCHYSK---TVIEWYATN------G-VSVI-PKDLNPPNCPQFRPIE-KYWAITKRRL-
          An-gambiae1      SHDHPVMFWPDLASCHYSK---VVREWYAEK------G-VLFV-PKNLNPPNCPQFRPIE-KYWAIMKRRL-

          Mariner-4_CGi    REMNNVWFQQDGATPHTAK---IVLEWLAEKFGEHFIS-LRSNIE---WPPHSPDLNPLDFFLWGYLKDR--
DD41D     Mariner-12_CGi   VNFDDVWFQQDGATPHTAG---CVLQWLDNTFGDQYIS-YRTANV---WPPHSPDLNPLDFFLWGYLKDR--
          Crmar2.5         MDVDDMWFQQDGATCHTAN---ETMALLRNKFNGRVIS-RNGDVN---WPPRSCDLTPLDFFLWGYLKEK--

          Bmmar1           FNNQEWSFQQDSAPGHKAR---STQSWLETN----VSD-FIRAED---WPSSSPDLNPLDYDLWSVLEST--
DD37D     Bmmar6           FNNRHWVFQQDSAPAHRAK---STQDWLAAR------E-IDFI-RHEDWPSSSPDLNPLDYKIWQHLEEK--
          Cemar6           FKKTKWTFQQDGAPAHKHK---NVQAMCESN------F-FDFI-AFNQWPSSSPDLNPMDYSVWSVLEAK--

          Soymar1          ELGSTIFIQQDNARTHINP---DDPEFVQAA-TQDGFD-IRLM-C---QPPNSPDFNVLDLGFFSAIQSL--
DD39D     Br-oleracea      DFGKTIFIQQDNAKTHVTV---NDEEFQVAA-LQHGLD-IQLM-C---QPPNSPDLNIFDLGYFRAIQAL--
          Ca-sativa        DFEKTIFIQQDNARTHVDP---RDEDFRAAS-SHHGFD-IRLM-C---QPPNSPDLNILDLGFFNAIQTL--

          IS630Ss          RRAKTITLIVDNYIIHKSR---ETQSWLKEN------PKFRVI-Y---QPVYSPWVNHVE-RLWQALHDTI-
DDxE      IS630Se          RRAKTITLVADNYIIHKSR---KVERWLEEN------PKFRLL-F---LPMYSPWLNPIE-RLWLSLHETI-
          RSA              FPNRKLHVILDNLNTHKKN-----EDWLKAH------PNVQFH-F---TPTSAPWLNQVE--VWFSILQGQS
```

1996), while DD39D transposons (*plants* family) include only the elements from the plants (Shao and Tu 2001).

Several oyster transposons were not included into the alignment, since they could not have been reliably aligned with the other members of the superfamily or were lacking the clearly recognizable catalytic triad. Those included *Mariner-21_CGi* with DD83D catalytic domain. Apparently, one or two large insertions occurred into the catalytic triad of this element (sequence not shown). Of note, this element has a huge size of 9744 bp while possessing an average size transposase (376 a.a.).

DDE/D domain of *Mariner-6_CGi* transposase could not have been precisely identified, since the region of the third amino acid residue is deleted or significantly modified. The catalytic domain of *Mariner-3_CGi* element also could not have been identified due to the missing region of the second aspartate of DDE/D domain.

The amino acid sequence of Mariner-31_CGi_p1 is lacking the domains relevant to its transposition activity, but instead it contains Myosin_tail_1 (pfam01576) domain. This domain mediates the assembly of myosin subunits into the macromolecule thread and its stability. This sequence was probably excised by the transposon in the course of a transposition event.

Phylogenetic Analysis

The complete transposase sequences of *IS630/Tc1/mariner* DNA transposons from *C. gigas*, as well as the transposases of 52 known DNA transposons representing the families from *IS630/Tc1/mariner* superfamily, were used for the phylogenetic analysis (Fig. 2, Online Resources 3). The amino acid sequences of the *IS630* elements were used as outgroup.

Elements from *C. gigas* possessing the catalytic triads DD35D (*Mariner-9_CGi, Mariner-24_CGi, Mariner-28_CGi, Mariner-30_CGi, Mariner-34_CGi*, and *Mariner-38_CGi*) and DD30D (*Mariner-10_CGi, Mariner-13_CGi, Mariner-36_CGi*, and *Mariner-37_CGi*) formed a single phylogenetic group with *pogo* family transposons. *Mariner-21_CGi*, which possesses the DD83D signature, probably due to the insertions in the catalytic domain region, was classified into the same clade on the phylogenetic tree. Thus, these 11 elements can all be classified as the members of *pogo* family.

The DD34D clade encompassed all the analyzed elements of *mariner* family, as well as four elements from *C. gigas*. Interestingly, for two of them (*Mariner-3_CGi,*

Mariner-6_CGi), the DDE/D domains could not have been identified; however, according to the phylogenetic analysis both elements belong to *mariner* family (Fig. 2). All elements with DD41D signature (including *Mariner-4_CGi* and *Mariner-12_CGi*) formed the entire *rosa* clade.

The DD34E elements from oyster, as well as from the other organisms, formed a single phylogenetic group; however, the individual branches showed low bootstrap values (<60).

In order to accurately establish the evolutionary relationship of DD34E elements from *C. gigas*, a phylogenetic analysis was performed, where we used the ML method (Fig. 3, Online Resources 4). For this analysis we used 52 known TE representing different families of *IS630/Tc1/mariner* transposons and a subset of the sequences from *C. gigas*, keeping those having DD34E signature. In addition, the sequence of *Mariner-35_CGi* transposase with DD36E signature was included in this analysis, since its similarity to DD34E-transposons has been established as a result of multiple alignment of DDE/D domain fragments (Fig. 1). The *IS630* elements were used as an outgroup in this analysis. Thus, we show that all analyzed elements of *C. gigas* form an entire clade with a bootstrap value of 51. This group also included all *Tc1* family elements and the members of *TRT* subfamily, which belong to *Tc1*-like elements, despite possessing DD37E domains (Zhang et al. 2016a). However, several groups can be identified within this clade, four of which are made of the elements of the oyster. Eleven of these elements from *C. gigas* were more closely related to *Minos* (Fig. 3), while five elements showed similarity to *Tc1*, *Sleeping Beauty*, and others (Fig. 3). *Mariner-32_CGi* and *Mariner-35_CGi* did not show close similarity to any of *Tc1*-like transposons analyzed. This indicates that those elements belong to separate subfamilies of *Tc1*-like transposons.

To date, two groups of *IS630/Tc1/mariner* DNA transposons with DD37E signature have been described, namely, the *mosquito* family elements (Shao and Tu 2001) and the members of *TRT* subfamily (Zhang et al. 2016a). The phylogenetic tree shows that two oyster elements possessing the DD37E catalytic triads do not belong to neither of them (Online Resources 2, Fig. 2). Instead, *Mariner-18_CGi* formed the clade with *Tc1* elements (DD34E), while *Mariner-31_CGi* formed the branch between the *mariner* (DD34D) and *plants* (DD39D) elements.

In order to more accurately classify the oyster elements with DD37E signature, we performed the *tblastn* search of WGS databases of metazoans using the amino acid sequences of *Mariner-18_CGi* and *Mariner-31_CGi* transposases as a query. Only DD37E transposases were selected from the search results. When *Mariner-18_CGi* transposase was used as a query, the nucleotide sequences coding for *Mariner-18_CGi*-like transposases (*Tnp-DD37E(L18)*) were found in only two mollusk species from Crassostrea genus

Fig. 2 Maximum likelihood (ML) tree for *IS630/Tc1/mariner* transposons of *C. gigas*. This tree was generated in MEGA7 using the rtREV + G + I + F model. Bootstrap percentages (100 replications) of > 60% are shown. The analysis involved 90 amino acid sequences (Online Resource 3). Families and subfamilies are indicated in the right-hand part of the tree
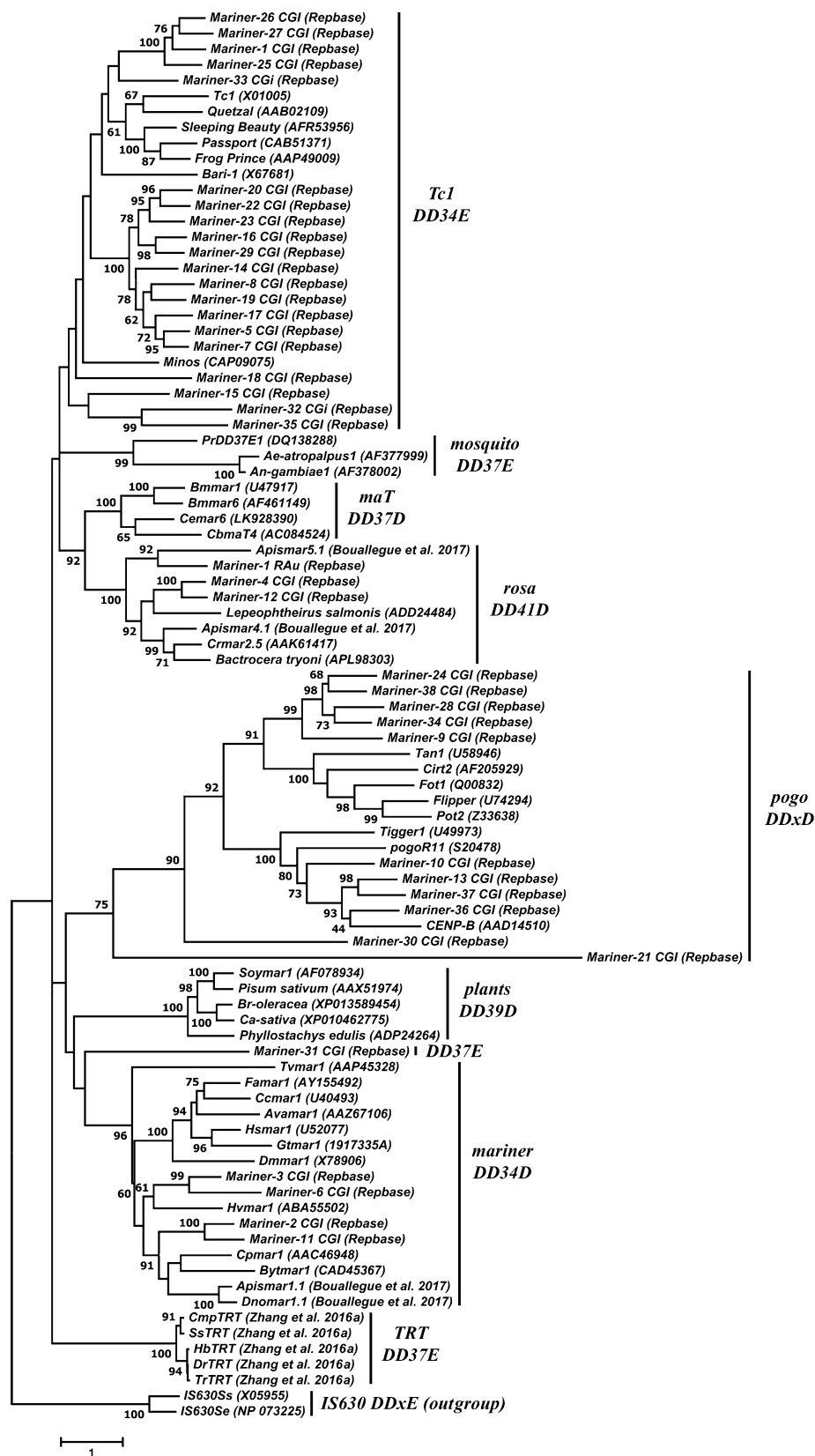
Fig. 3 Phylogenetic relationships of Pacific oyster DD34E elements with other members of the *IS630/Tc1/mariner* superfamily based on their transposases. The best-suited amino acid substitution model for these data was the LG+G+I model. ML phylogenetic tree was then built using the MEGA7 software. Bootstrap values (100 replications) of <60% are not shown. The analysis involved 71 amino acid sequences (Online Resource 4). Families and subfamilies are indicated in the right-hand part of the tree
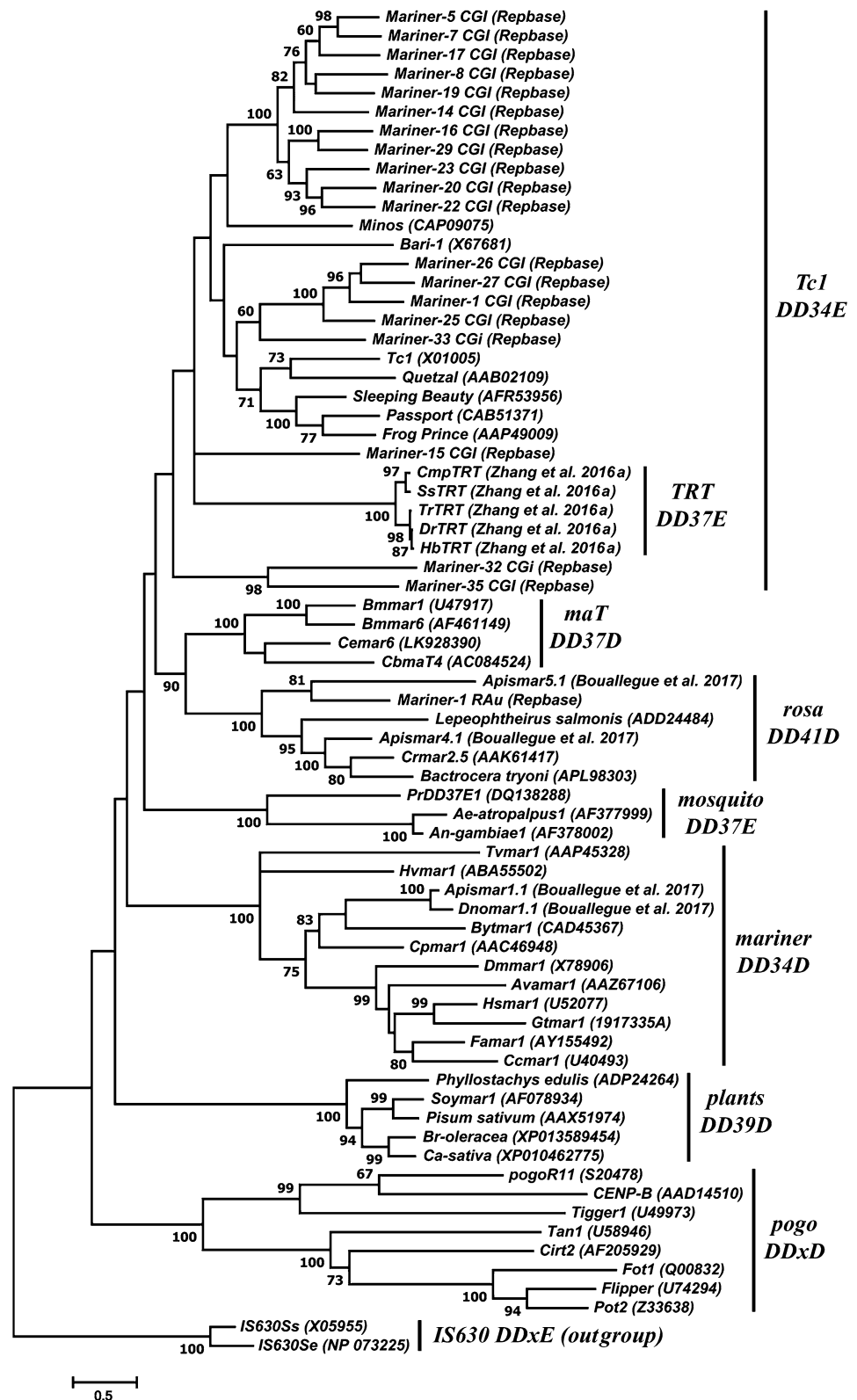
(*C. gigas* and *C. virginica*). Also, *Tnp-DD37E(L18)* were identified in some members of cnidarians (subclass Hexacorallia and class Hydrozoa) and echinoderms (superclass Echinozoa). Among the Arthropoda, *Tnp-DD37E(L18)* were found in the representatives of two crustacean classes (Malacostraca, Maxillopoda), subclass Acari (mites and ticks) and

class Collembola. In the Chordata, the similar proteins were identified only in some fish species (infraclass Teleostei). When *Mariner-31_CGi* transposase was used as a query, the transposases with DD37E signature (*Tnp-DD37E(L31)*) were found in the subclass Pteriomorphia (bivalves) and in some members of subclass Hexacorallia (cnidarians). A mosaic representation of *Mariner-18_CGi*-like and *Mariner-31_CGi*-like elements on the evolutionary tree of metazoans may indicate at the cases of horizontal transfer in the history of these TE. The horizontal transfer is a common feature of *IS630/Tc1/mariner* superfamily elements (Dupeyron et al. 2014; Wallau et al. 2016).

The phylogenetic analysis, which included *Mariner-18_CGi*-like and *Mariner-31_CGi*-like transposases, as well as 52 known DNA transposons from *IS630/Tc1/mariner* superfamily, shows that *Mariner-18_CGi* and most of *Tnp-DD37E(L18)* form an entire individual branch closely related to *Tc1*-like transposons (Fig. 4, Online Resources 5). *Tnp-DD37E(L18)* from copepoda *Eurytemora affinis* forms a separate branch also close to *Tc1*-like transposons. *Tnp-DD37E(L18)* from amphipod crustacean *Parhyale hawaiensis* formed the same clade with *TRT* subfamily elements. We suggest that *Mariner-18_CGi* and *Tnp-DD37E(L18)* elements are the group of DNA transposons evolutionary related to DD34E elements. *Mariner-31_CGi*-like elements formed a separate clade distinct from both *mosquito* family and *TRT* subfamily. These data, together with the fact that *Mariner-31_CGi* possesses differences from DD37E elements in its catalytic domain (Fig. 1), suggest that these DNA transposons represent a new family with DD37E signature.

Thus, according to the analysis of DDE/D domains and the phylogenetic analysis, *Tc1* family, which accounts for more than a half of elements in the genome of *C. gigas*, is the most representative of *IS630/Tc1/mariner* superfamily DNA transposons (Table 1). More than a quarter of *IS630/Tc1/mariner* elements from *C. gigas* belong to *pogo* family. The rest are represented by DNA transposons of *mariner* and *rosa* families, and also by a new family of *Mariner-31_CGi*-like elements.

Activity of *IS630/Tc1/mariner* DNA Transposons from *C. gigas*

Although the elements of *IS630/Tc1/mariner* superfamily have been discovered in virtually every eukaryotic genome studied, only few of them possess the functional transposase and their activity has been confirmed experimentally. To date, the following active *IS630/Tc1/mariner* DNA transposons have been identified: *Impala* and *Fot1* from the fungus, *Fusarium oxysporum* (Daboussi et al. 1992; Langin et al. 1995), *Tc1* and *Tc3* from the nematode *Caenorhabditis elegans* (Emmons et al. 1983; Collins et al.

1989), *Minos, Mos1*, and *Himar1* from drosophila (Bryan et al. 1990; Franz and Savakis 1991; Robertson and Lampe 1995), *Mboumar-9* from the ant *Messor bouvieri* (Munoz-Lopez et al. 2008), and *Passport* from the flounder (Clark et al. 2009). This phenomenon is associated with the fact that the transposase-encoding nucleotide sequences of the majority of known elements are damaged by deletions, frameshifts, and amino acid substitutions. It is suggested that most of *IS630/Tc1/mariner* transposons are now at the terminal stage of their "lifecycle," and the peak of their activity has been far in the past. According to the concept of the "lifecycle" (Schaack et al. 2010), a new active element enters the host genome and starts the colonization of germ line. Then occurs the stage characterized by the copy number increase and the spread of element across the population followed by the increase in the diversity of active elements, the subsequent vertical inactivation of the copies due to the mutation accumulation or negative selection and, finally, the elimination of inactive copies of the element from the genome.

In order to analyze the activity of *IS630/Tc1/mariner* DNA transposons from the Pacific oyster, we performed the search for transcripts homologous to the predicted transposases of the known elements in NCBI Transcriptome Shotgun Assembly (TSA) and Expressed Sequence Tags (EST) databases. Although the presence of transcripts homologous to the transposase is not a direct proof of their functionality or, moreover, of their transpositional activity, it can still serve as an indirect evidence of those.

For 25 elements we found transcripts with a high degree of homology to the putative transposases. However, most of these transcripts represented the chimeric sequences encompassing the fragments of transposase and some other proteins, and thus their presence cannot serve as evidence of a full mRNA of a potentially functional transposase being transcribed. Moreover, a small fraction of the detected transcripts represented fragments of the transposase from which one can hardly draw a conclusion on the potential activity of the studied DNA transposons. With the exception of the chimeric and truncated transcripts, the ones most closely resembling the sequences of their putative transposases were identified only for eight elements (Online Resource 2).

The transcribed RNA sequence of a single element (*Mariner-35_CGi*) was completely correspondent to the predicted transposase, while those of the other seven elements harbored substitutions and deletions/insertions, both shorter (9 bp) and longer (up to 396 bp) ones.

The differences between the transposase sequences predicted by annotation and the transcripts may be due not only to the annotation errors, but also to the fact that the genome (PRJNA70283) and the transcriptome (PRJNA301543) were sequenced from different individuals in separate experiments. Despite the presence of the transposase transcripts,
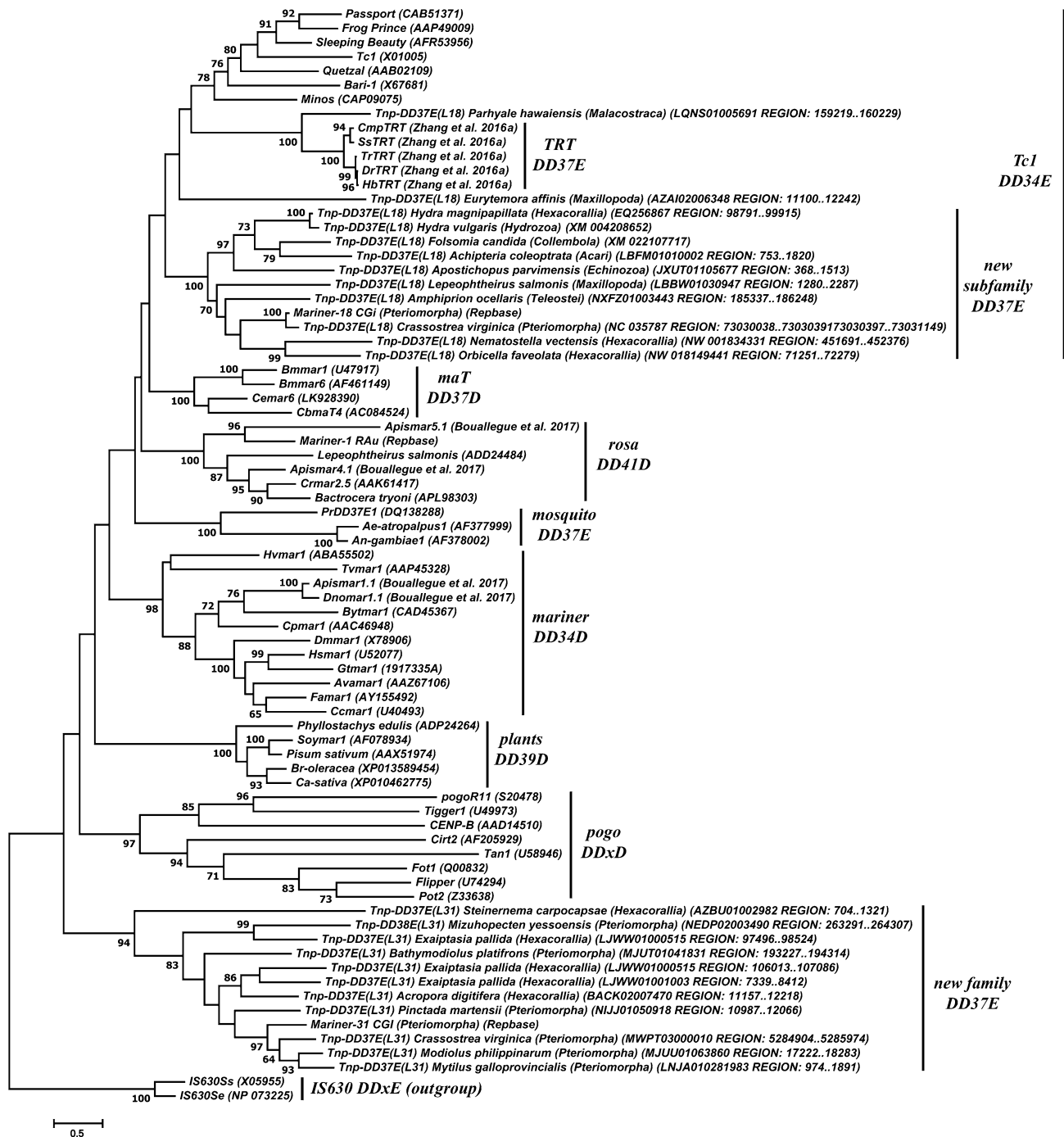
Fig. 4 The phylogeny based on amino acid sequences of the *Mariner-18_CGi*-like (*Tnp-DD37E(L18)*) and the *Mariner-31_CGi*-like (*Tnp-DD37E(L31)*) transposases. This tree was generated in MEGA7 with the ML method, using the LG+G model. Only bootstrapping values (100 replications) higher than 60% are written on the branch. The analysis involved 77 amino acid sequences (Online Resource 5). Families and subfamilies are indicated in the right-hand part of the tree. Species, taxon, and localization in WGS are indicated for the *Tnp-DD37E(L18)* and the *Tnp-DD37E(L31)* transposases

the activity of relevant elements remains in question, especially, since almost all of them have low copy numbers.

The search for the regions homologous to DNA transposons in the genome of *C. gigas* using *blastn* yielded the

copy number of each element per genome (Online Resource 2). This analysis was based only on the sequences of potentially autonomous oyster elements. We considered both the number of homologous loci, and the number of full-size

Table 1 Distribution of *IS630/Tc1/mariner* transposons of Pacific oyster by family

Family	Representatives	Fraction (%)[a]
Tc1 (DD34E)	*Mariner-1_CGi, Mariner-5_CGi, Mariner-7_CGi, Mariner8_CGi, Mariner-14_CGi, Mariner-15_CGi, Mariner-16_CGi, Mariner-17_CGi, Mariner-18_CGi, Mariner-19_CGi, Mariner-20_CGi, Mariner-22_CGi, Mariner-23_CGi, Mariner-25_CGi, Mariner-26_CGi, Mariner-27_CGi, Mariner-29_CGi, Mariner-32_CGi, Mariner-33_CGi, Mariner-35_CGi*	52.7
Mariner (DD34D)	*Mariner-2_CGi, Mariner-3_CGi, Mariner-6_CGi, Mariner11_CGi*	10.5
Pogo (DDxD)	*Mariner-9_CGi, Mariner-10_CGi, Mariner-13_CGi, Mariner-21_CGi, Mariner-24_CGi, Mariner-28_CGi, Mariner-30_CGi, Mariner-34_CGi, Mariner-36_CGi, Mariner-37_CGi, Mariner-38_CGi*	28.9
maT (DD37D)	–	0
Mosquito (DD37E)	–	0
Plants (DD39D)	–	0
Rosa (DD41D)	*Mariner-4_CGi, Mariner-12_CGi*	5.3
IS630 (DDxE)	–	0
New putative family (DD37E)	*Mariner-31_CGi*	2.6

[a]Fraction of family representatives in the total number of unique *IS630/Tc1/mariner* DNA transposons from the Pacific oyster

elements (fragments retaining ≥ 95% of length and TIRs). Most *IS630/Tc1/mariner* elements from the Pacific oyster are represented by the few full-size copies (Online Resource 2, Fig. 5), and only for two of them (*Mariner-2_CGi* and *Mariner-5_CGi*) the number of complete copies is > 25. Such a unimodal distribution with positive asymmetry may serve as an indication that oyster elements have been virtually eliminated from the genome, i.e., that they are at the terminal stage of their "lifecycle".

However, virtually every element has a unique ratio of entire number of copies to the full-size copy number, which results from an individual history of each element. This ratio allows one to judge about how long ago the period of activity of the element occurred, while the entire copy number suggests the scope of transposition activity. For example, *Mariner-2_CGi, Mariner-3_CGi*, and *Mariner-8_CGi*

elements have been, obviously, rather active in the past, as suggested by their high copy numbers (> 100 copies). Such elements as *Mariner-33_CGi* and *Mariner-38_CGi* are represented in the genome by the few sequences (2 copies of each element). An immediate immobilization of such elements by the genome after the initial colonization and, thus, a premature termination of a "lifecycle" is a possibility for such DNA transposons.

In an attempt to evaluate the dynamics of *IS630/Tc1/mariner* DNA transposons in the genome of the Pacific oyster, we used the approach to such problems previously used by Bouallègue et al. (2017). In order to estimate the period of amplification of the elements from *C. gigas*, we determinate the percentage of identity of the consensus sequences of elements with their corresponding full-size copies. As a result, only three elements *Mariner-14_CGi,*

Fig. 5 Number of Pacific oyster *IS630/Tc1/mariner* transposons full-length copies
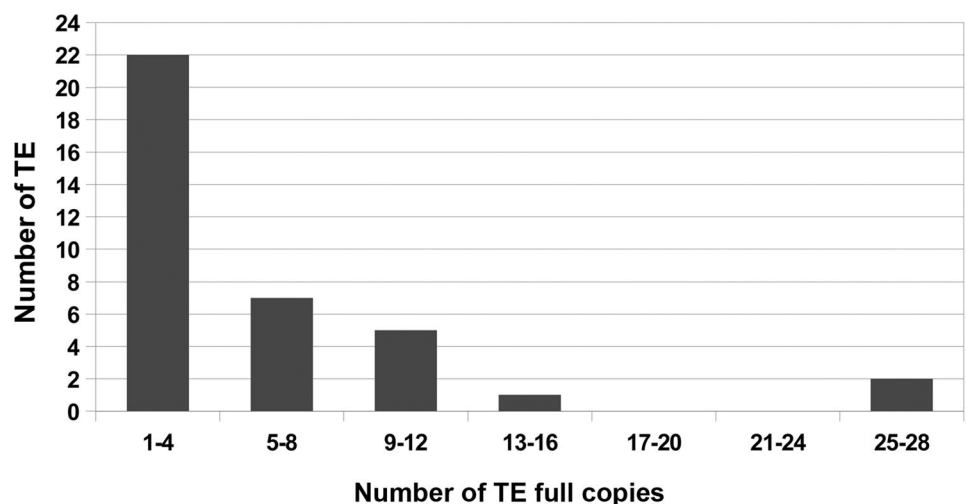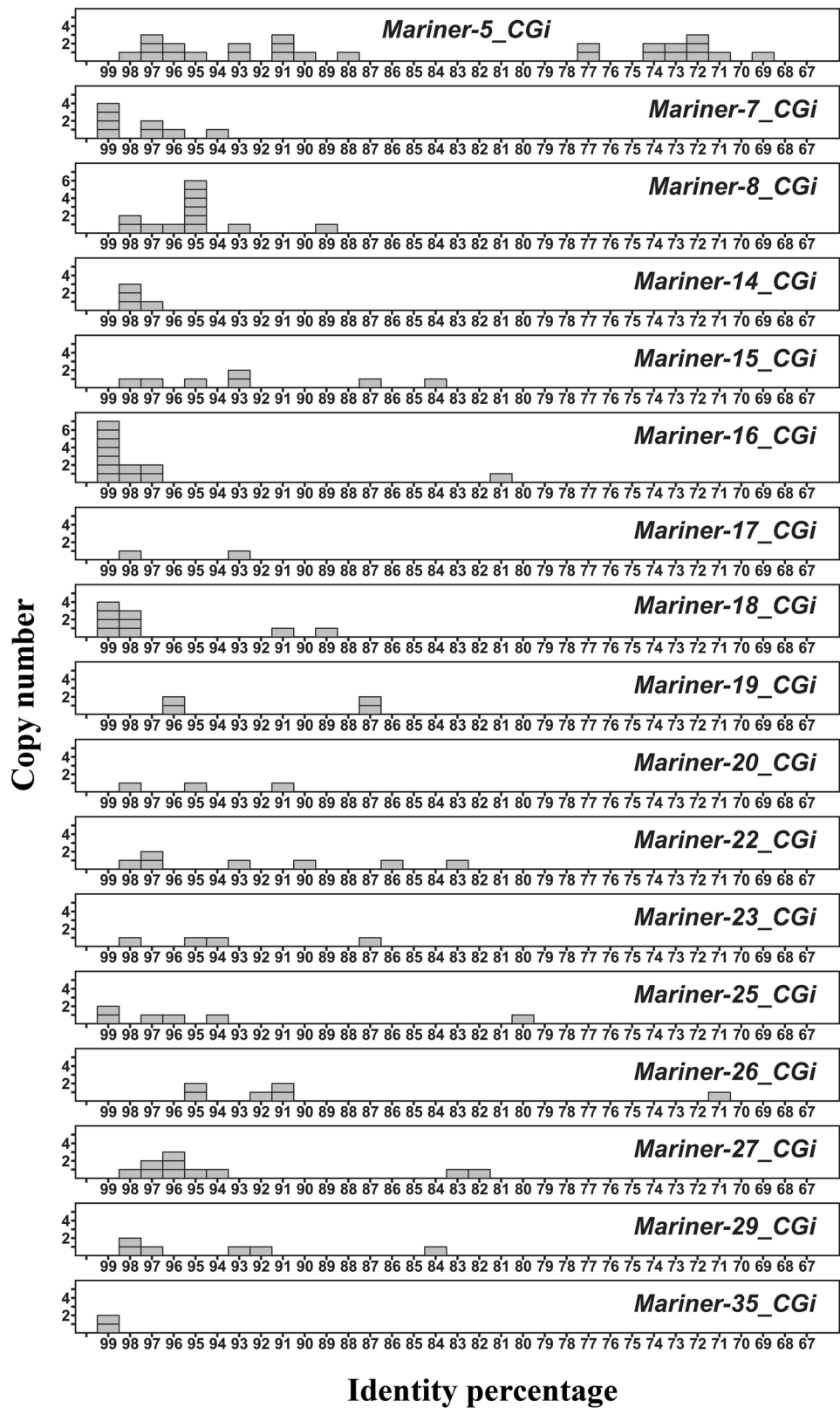
Fig. 6 Evolution dynamics of Pacific oyster *Tc1*-like elements. The identity percentage was estimated by comparing the consensus with the copies. A high level of similarity corresponds to the recent invasion of the element (the beginning of the "lifecycle"), and a decrease in this percentage indicates an earlier event

Mariner-31_CGi, and *Mariner-35_CGi* were shown to have the identity percentage of 97–99% (Figs. 6, 7). Another six elements with the prevalence of copies with high identity (97–99%), also have one or two copies with the identity of 81–96%. For three of nine elements mentioned above (*Mariner-2_CGi, Mariner-30_CGi,* and *Mariner-35_CGi*), the full-size transcribed RNA sequences were found (Online Resource 2). The high degree of conservation of the sequences of those elements may indicate that these copies are still retaining activity or have been inactivated only recently. The other DNA transposons (with the identity of copies ~ 69–98%) are less conserved, which supports the hypothesis that their "lifecycle" has already been completed.

Mariner-5_CGi represents an interesting example of an element, which has undergone a reactivation event and commenced a new "lifecycle" in its evolutionary history in the genome of *C. gigas*. This is suggested by the presence of two "fractions" of its copies with identity percentage of 69–77% and 88–98%, respectively (Fig. 6).

Thus, the predominant majority of *IS630/Tc1/mariner* DNA transposons from the Pacific oyster seem to lack the transposition activity. Among the transcriptionally active elements, only *Mariner-2_CGi* has a higher copy number (197 copies, 27 of which are the full-size copies) and high identity percentage value (94–99%), being probably still functional. *Mariner-30_CGi* and *Mariner-35_CGi* elements
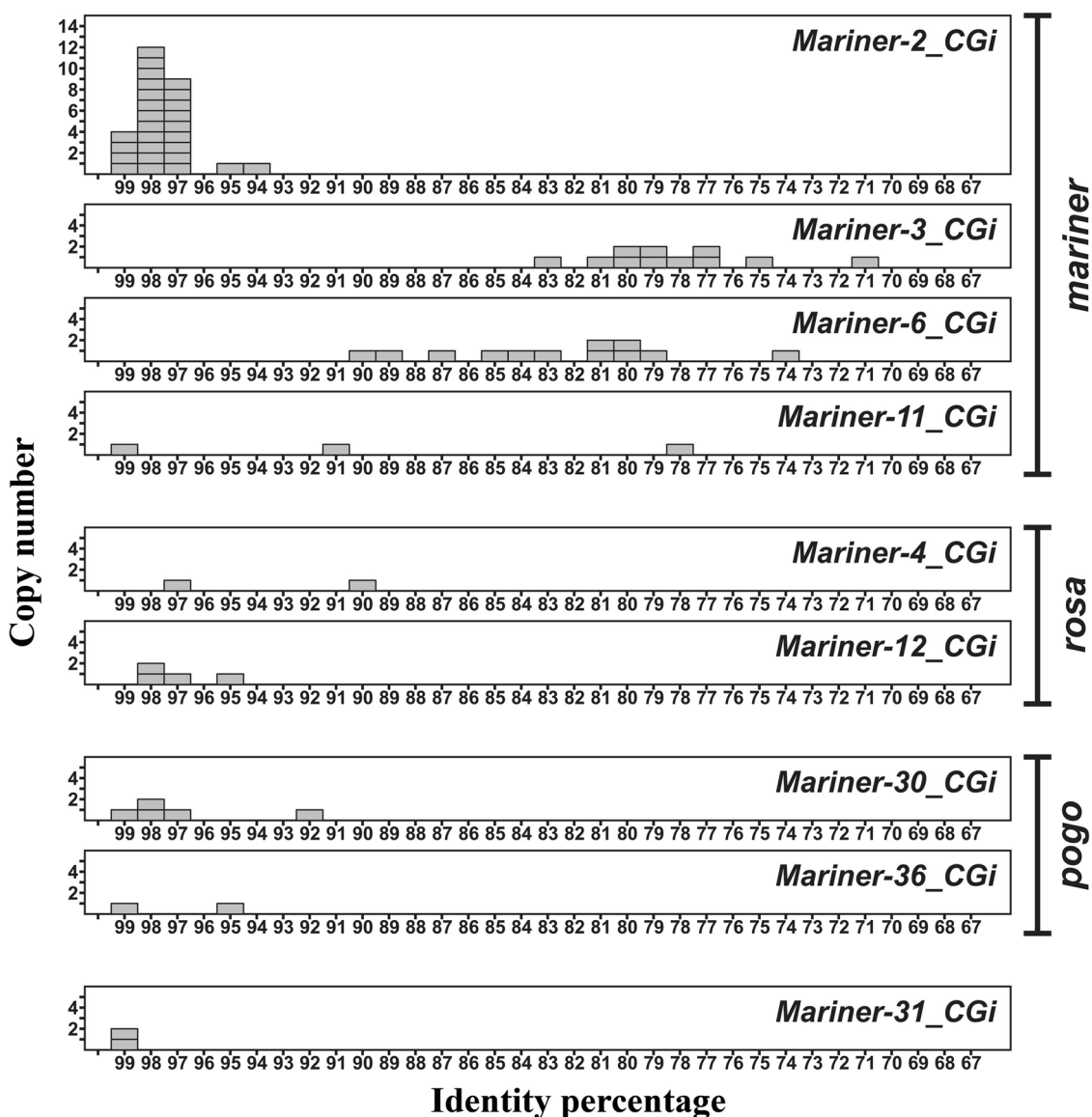

Fig. 7 Evolution dynamics of non-*Tc1*-like elements of *C. gigas*. The identity percentage was estimated by comparing the consensus with the copies. A high level of similarity corresponds to the recent invasion of the element (the beginning of the "lifecycle"), and a decrease in this percentage indicates an earlier event

may also retain their function, as they have the transcribed RNA sequences and the degrees of identity of 92–99% and 99%, respectively. A comparison of transposon copy numbers between the groups of elements with or without the transcriptional activity of transposase did not yield any significant differences.

Conclusion

IS630/Tc1/mariner DNA transposons of the Pacific oyster are represented by four families, *Tc1* (DD34E), *mariner* (DD34D), *pogo* (DDxD), and *rosa* (DD41D), the first of which accounts for more than a half of the discovered *IS630/Tc1/mariner* elements from *C. gigas*. *Mariner-31_CGi* is the representative of a new family with DD37E signature. In addition, we established that eight elements representing three families (*Tc1, mariner,* and *pogo*) demonstrate transcriptional activity. Presumably, only three of those elements may retain the transposition activity.

The studies of TE diversity and dynamics in the species with uninvestigated genomes facilitate a better understanding of both the TE evolutionary history and the molecular evolution of the genomes. Recently, due to growing number of organisms with fully sequenced genomes, several new groups of transposons were found in *IS630/Tc1/mariner* superfamily. For example, the elements with DD41D signature were described in the publications of Zhang et al. (2016b) and Bouallègue et al. (2017), while the *TRT* elements with DD37E signature were presented in the study of Zhang et al. (2016a). The genome of a Pacific oyster harbors two DD37E elements, which represent a new subfamily of *Tc1*-like elements (*Mariner-18_CGi*) and a new and previously unknown family (*Mariner-31_CGi*). Our results not only identify the new TE groups, but also demonstrate that the active elements with DD37E signature have repeatedly emerged in the course of evolution of *IS630/Tc1/mariner* DNA transposons.

The question of molecular domestication of DNA transposons remains a topical issue. Molecular domestication is an evolutionary process by which a TE-derived coding sequence gives rise to a functional host gene. Several criteria have been proposed on how to confirm of the TE-derived gene (Feschotte and Pritham 2007): (i) elements that have undergone domestication exist as single copies in the genomes; (ii) structurally, these genes are lacking the molecular features of DNA transposons, such as flanking TIR and TSD; (iii) orthologs are found in distant species. At least, three independent events of domestication of *pogo*-like transposases are known in the multicellular organisms (Mateo and Gonzalez 2014). The *pogo*-like transposases are suggested to be predisposed to domestication due to some of their specific features (Feschotte and Pritham 2007;

Casola et al. 2008). Eight of eleven elements of *pogo* family in the genome of *C. gigas* possess a single full-size copy, while three of them are lacking TIRs. The study of molecular domestication of these TE presents a promising line of research.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Albertin CB, Simakov O, Mitros T et al (2015) The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524:220–224

Altschul SF, Madden TL, Schäffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Arensburger P, Megy K, Waterhouse RM et al (2010) Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science 330:86–88

Bao W, Jurka J (2013a) DNA transposons from the Pacific oyster genome. Repbase Rep 13(1):599–633

Bao W, Jurka J (2013b) DNA transposons from the Pacific oyster genome. Repbase Rep 13(3):1415–1426

Bao W, Jurka J (2013c) DNA transposons from the Pacific oyster genome. Repbase Rep 13(4):1516–1517

Bouallègue M, Filée J, Kharrat I et al (2017) Diversity and evolution of *mariner*-like elements in aphid genomes. BMC Genom 18:494. https://doi.org/10.1186/s12864-017-3856-6

Bryan G, Garza D, Hartl D (1990) Insertion and excision of the transposable element mariner in Drosophila. Genetics 125:103–114

Capy P, Vitalis R, Langin T et al (1996) Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? J Mol Evol 42:359–368

Capy P, Langin T, Higuet D et al (1997) Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? Genetica 100:63–72

Casola C, Hucks D, Feschotte C (2008) Convergent domestication of *pogo*-like transposases into centromere-binding proteins in fission yeast and mammals. Mol Biol Evol 25:29–41

Chow KC, Tung WL (2000) Magnetic field exposure stimulates transposition through the induction of DnaK/J synthesis. Biochem Biophys Res Commun 270(3):745–748

Clark KJ, Carlson DF, Leaver MJ et al (2009) *Passport*, a native *Tc1* transposon from flatfish, is functionally active in vertebrate cells. Nucleic Acids Res 37:1239–1247

Claudianos C, Brownlie J, Russell R et al (2002) *maT*: a clade of transposons intermediate between *mariner*. and *Tc1*. Mol Biol Evol 19:2101–2109

Collins J, Forbes E, Anderson P (1989) The *Tc3* family of transposable genetic elements in *Caenorhabditis elegans*. Genetics 121:47–55

Daboussi MJ, Langin T, Brygoo Y (1992) *Fot1*, a new family of fungal transposable elements. Mol Gen Genet 232:12–16

de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD (2011) Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 7:e1002384

Del Re B, Garoia F, Mesirca P et al (2003) Extremely low frequency magnetic fields affect transposition activity in *Escherichia coli*. Radiat Environ Biophys 42(2):113–118

Doak TG, Doerder FP, Jahn CL, Herrick G (1994) A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc Natl Acad Sci USA 91(3):942–946

Dupeyron M, Leclercq S, Cerveau N, Bouchon D, Gilbert C (2014) Horizontal transfer of transposons between and within crustaceans and insects. Mob DNA 5:4

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Emmons SW, Yesner L, Ruan K, Katzenberg D (1983) Evidence for a transposon in *Caenorhabditis elegans*. Cell 32:55–65

Fernández-Medina RD, Granzotto A, Ribeiro JM, Carareto CM (2016) Transposition burst of *mariner*-like elements in the sequenced genome of *Rhodnius prolixus*. Insect Biochem Mol Biol 69:14–24

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–368

Finnegan DJ (1992) Transposable elements. Curr Opin Genet Dev 2(6):861–867

Franz G, Savakis C (1991) *Minos*, a new transposable element from *Drosophila hydei*, is a member of the *Tc1*-like family of transposons. Nucleic Acids Res 19:6646

Gomulski LM, Torti C, Bonizzoni M et al (2001) A new basal subfamily of *mariner* elements in *Ceratitis rosa* and other tephritid flies. J Mol Evol 53:597–606

Haymer DS, Marsh JL (1986) Germ line and somatic instability of a white mutation in *Drosophila mauritiana* due to a transposable genetic element. Dev Genet 6:281–291

Henikoff S (1992) Detection of Caenorhabditis transposon homologs in diverse organisms. New Biol 4:382–388

Hernandez-Hernandez EM, Fernández-Medina RD, Navarro-Escalante L et al (2017) Genome-wide analysis of transposable elements in the coffee berry borer *Hypothenemus hampei* (Coleoptera: Curculionidae): description of novel families. Mol Genet Genom 292(3):565–583. https://doi.org/10.1007/s00438-017-1291-7

Jacobson JW, Medhora MM, Hartl DL (1986) Molecular structure of a somatically unstable transposable element in Drosophila. Proc Natl Acad Sci USA 83:8684–8688

Jarvik T, Lark KG (1998) Characterization of *Soymar1*, a mariner element in soybean. Genetics 149:1569–1574

Jurka J (2012) DNA transposons from the Pacific oyster genome. Repbase Rep 12(12):2460–2461

Jurka J (2013) DNA transposons from the Pacific oyster genome. Repbase Rep 13(4):1547

Kapitonov VV, Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet 9(5):411–412

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874

Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Langin T, Capy P, Daboussi MJ (1995) The transposable element *impala*, a fungal member of the *Tc1*-mariner superfamily. Mol Gen Genet 246:19–28

Liu Y, Yang G (2014) *Tc1*-like transposable elements in plant genomes. Mob DNA 5:17

Mateo L, Gonzalez J (2014) *Pogo*-like transposases have been repeatedly domesticated into CENP-B-related proteins. Genome Biol Evol 6:2008–2016. https://doi.org/10.1093/gbe/evu153

Mesquita RD, Vionette-Amaral RJ, Lowenberger C et al (2015) Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. Proc Natl Acad Sci USA 112:14936–14941

Munoz-Lopez M, Siddique A, Bischerour J et al (2008) Transposition of *Mboumar-9*: identification of a new naturally active *mariner*-family transposon. J Mol Evol 382:567–572

Piacentini L, Fanti L, Specchia V et al (2014) Transposons, environmental changes, and heritable induced phenotypic variability. Chromosoma 123:345–354

Puzakov MV, Puzakova LV, Zakharov IK (2017) Diversity and distribution of mobile genetic elements in marine invertebrates genomes. Vavilov J Genet Breed 21(2):269–283. https://doi.org/10.18699/VJ16.16-o

Robertson HM (1995) The *Tc1-mariner* superfamily of transposons in animals. J Insect Physiol 41:99–105

Robertson HM, Asplund ML (1996) *Bmmar1*: a basal lineage of the mariner family of transposable elements in the silkworm moth, *Bombyx mori*. Insect Biochem Mol Biol 26(8–9):945–954

Robertson HM, Lampe DJ (1995) Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. Mol Biol Evol 12:850–862

Schaack S, Gilbert C, Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25:537–546

Sergeeva EM, Salina EA (2011) Transposable elements and plant genome evolution. Russ J Genet: Appl Res 1(6):565–576. https://doi.org/10.1134/S2079059711060086

Shao H, Tu Z (2001) Expanding the diversity of the *IS630-Tc1-mariner* superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. Genetics 159(3):1103–1115

Simakov O, Marletaz F, Cho SJ et al (2013) Insights into bilaterian evolution from three spiralian genomes. Nature 493:526–531

Smit AFA, Riggs AD (1996) Tiggers and other DNA transposon fossils in the human genome. Proc Natl Acad Sci USA 93:1443–1448

Strand DJ, McDonald JF (1989) Insertion of a copia element 5′ to the *Drosophila melanogaster* alcohol dehydrogenase gene (*adh*) is associated with altered developmental and tissue-specific patterns of expression. Genetics 121(4):787–794

Takeuchi T, Kawashima T, Koyanagi R et al (2012) Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. DNA Res 19:117–130

Wallau GL, Capy P, Loreto E, Le Rouzic A, Hua-Van A (2016) VHICA, a new method to discriminate between vertical and horizontal transposon transfer: application to the mariner family within Drosophila. Mol Biol Evol 33(4):1094–1109

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982. https://doi.org/10.1038/nrg2165

Yoshida MA, Ishikura Y, Moritaki T et al (2011) Genome structure analysis of molluscs revealed whole genome duplication and lineage specific repeat variation. Gene 483:63–71

Zakharenko LP, Zakharov IK, Voloshina MA, Gracheva EM, Romanova OA, Kochieva EZ, Simonova OB, Georgiev P, Golubovsky MD (2000) *hobo*-induced rearrangements are responsible mutation bursts at the *yellow* locus in natural population of *Drosophila melanogaster*. Mol Gen Genet MGG 263(2):335–341

Zakharenko LP, Kovalenko LV, Zakharov IK, Perepelkina MP (2006) The effect of γ-radiation on induction of the *hobo* element transposition in *Drosophila melanogaster*. Russ J Genet 42(6):619–622

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7(1–2):203–214

Zhang G, Fang X, Guo X et al (2012) The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490(7418):49–54. https://doi.org/10.1038/nature11413

Zhang HH, Li GY, Xiong XM, Min-Jin Han MJ, Zhang XG, Dai FY (2016a) *TRT*, a vertebrate and protozoan *Tc1*-like transposon: current activity and horizontal transfer. Genome Biol Evol 8(9):2994–3005. https://doi.org/10.1093/gbe/evw213

Zhang HH, Shen YH, Xiong XM et al (2016b) Identification and evolutionary history of the DD41D transposons in insects. Genes Genom 38:109–117. https://doi.org/10.1007/s13258-015-0356-4