Multiple Invasions of *Visitor*, a DD41D Family of *Tc1/mariner* Transposons, throughout the Evolution of Vertebrates

Dan Shen¹⁻² ‡, Bo Gao¹ ‡, Csaba Miskey², Cai Chen¹, Yatong Sang¹, Wencheng Zong¹, Saisai Wang¹, Yali Wang¹, Xiaoyan Wang¹, Zoltán Ivics² and Chengyi Song¹*

1, College of Animal Science & Technology, Yangzhou University, Yangzhou, Jiangsu, 225009, China
2, Division of Medical Biotechnology, Paul Ehrlich Institute, Langen,63225, Germany

‡ These authors contributed equally to this work.

* Corresponding authors: cysong@yzu.edu.cn

© Crown copyright 2020
Abstract

Although the DD41D (named as Visitor, VS) family of Tc1/mariner transposons was discovered in Arthropods, Mollusca and Ctenophora, the evolution profile of this family is still largely unknown. We found that VS is widespread in the animal kingdom, including 140 species of 18 orders in invertebrates and 30 species of 12 orders in vertebrates, and one land plant species. Our data revealed multiple horizontal transfer (HT) events in both invertebrates and vertebrates and invasion into multiple lineages of mammals, including Chiroptera (seven species), Dasyuromorpha/Marsupialia (one species), Didelphimorphia/Marsupialia (one species), Diprotodontia/Marsupialia (two species) and Primates (one species). Phylogenetic analysis revealed a close relationship of VSs to DD37D/maT and DD34D/mariner and confirmed that VSs with the DD40D signature identified previously is not a distinct family, but originated from DD41D/VS. Age analysis revealed that the most recent invasion of VSs was found in ray-finned fishes and a toad, followed by relatively young invasions in bats and marsupials, whereas VSs in mammals, jawless fishes and lizards were mainly represented by ancient copies, suggesting old age. Phylogenetic analyses and comparison of pairwise distances between VSs and recombination-activating gene 1 (RAG1) support HT events of VSs in vertebrates. The intact VSs from bats were non-functional as determined by the transposition activity assay. Some vertebrate lineages and species were identified as the hot hosts of Tc1/mariner transposons. Overall, our study presents the evolution profile of VSs and suggests that VSs play roles in diversifying and shaping the genomes of diverse animal lineages.

Key words: Tc1/mariner, Visitor, DD41D, horizontal transfer, evolution

Introduction

Genomes contain diverse repetitive DNA sequences of transposable elements (TEs), which are divided into two classes: Class I transposons (retrotransposons) and Class II transposons (DNA transposons). TEs account for up to 35–69% of genomic DNA in different mammalian species (Waterston et al. 2002; de Koning et al. 2011) and up to 90% of the genome in some plants (Sergeeva 2011), playing important roles in plasticity, adaptability and evolution of prokaryotic and eukaryotic genomes (Sotero-
Class II transposons can be divided into three subcategories depending on their transposition mechanism: the classical ‘cut-and-paste’ DNA transposons, the rolling circle DNA transposons and the ‘self-synthesizing’ DNA transposons (Feschotte 2007). They are either autonomous transposons encoding their own transposase or non-autonomous transposons that are either truncated or contain internal deletions (thus, with no functional transposases), but with two intact transposon ends supporting transposition. Class II transposons are a major component of the analysed teleost genomes and the most important contributors to genome size variation across teleost lineages (Gao et al. 2016). Based on the molecular structures of transposons, DNA transposons can be classified into 19 superfamilies, namely, Tc1/mariner, hAT, PiggyBac, CACTA, MuDR, Merlin, Transib, P, PIF/Harbinger, Mirage, Zator, Ginger, Kolobok, Chapaev, Novosib, Rehavkus, PHIS, Sola and Academ (Jurka 2000; Tang et al. 2015).

The Tc1/mariner superfamily is ubiquitous and constitutes the largest group of eukaryotic Class II transposons (Bourque et al. 2018; Burns 2017). Its members share several common characteristics and synapomorphies. In particular, the Tc1/mariner superfamily elements contain at least one open reading frame (ORF) that encodes a transposase with approximately 350 amino acid residues (Benjamin et al. 2007). The transposase commonly contains DNA binding domains with two helix–turn–helix (HTH) motifs and a catalytic triad DDE/D motif (Lohe & Hartl 1996). The transposase encoding DNA is flanked by untranslated sequences and terminal inverted repeats (TIRs). On the basis of variation in the DDE/D signature, Tc1/mariner elements were further classified into at least nine main monophyletic families: DD34E/Tc1, DD36E/IC (Sang et al. 2019), DD34D/mariner, DD37E/TrT, DD37D/maT, DD39D (Zhang et al. 2016), DDxD/pogo, DDxE and DD41D (Gomulski et al. 2001). The DD41D family was first identified in Ceratitis rosa and other tephritid flies (Gomulski et al. 2001), and it was characterized by a total length of approximately 1300 bp), a length of TIR ranging 15–45 bp and a DD41D domain transposase, which was once considered to be a variant of the DD34D catalytic domain of the mariner family. Later, Crmar2 was identified as a member of the DD41D family in insects (Zhang et al. 2016). Moreover, a few transposons encoding transposases with the DD40D motif, a variant of the DD41D catalytic domain, were also identified in insects (Xie et al. 2018). Furthermore, another minor group of DD41D transposons (named as Lsra), which have a long total length of approximately 1500–4400 bp and TIRs of up to 1900 bp was found (Bouallègue et al. 2017). However, all previous evolution analyses of the DD41D family were restricted to invertebrates (insects and oyster), and the distribution of DD41D family
members is still unknown outside of invertebrates. Herein, we describe the taxonomic distribution, intrafamily diversity, horizontal transfer (HT) and evolutionary dynamics of the DD41D (named as Visitor, VS) family across different lineages of vertebrates. We also investigated experimentally the transposition capacity of some intact VS copies detected in bat lineages. These findings will enable a better understanding of the evolution profiles of the DD41D/VS transposons and their biological roles in the evolution of animal genomes.

Materials and Methods

Data Mining

A panel of 10 transposase sequences with DD41D and DD40D domains identified in previous studies (Zhang et al. 2016; Bouallègue et al. 2017) were used as queries in TBLASTN to search against Whole Genome Shotgun databases available on the National Center for Biotechnology Information (NCBI) website with a value of 1e-5. The DD41D or DD40D transposon was determined to exist in a species when the unique DD41D or DD40D motif was detected in the catalytic domain of a transposon in that species. Then, the best hits were extracted with 2 kb flanking sequences and were manually investigated for TIRs and determined the full sequence of VSs. The consensus sequences of VSs were reconstructed for VSs using multiple alignment of copies (at least 10 copies) in each genome by using the online emboss explorer http://www.bioinformatics.nl/emboss-explorer/). The identified VS sequences (representative or consensus) were then used to retrieve more elements of VSs by BLAST. A VS transposon was considered to be present in a genome if the BLAST hit was at least 80% identity and 40% coverage of the queries. VS with a low copy number in the genome, which may be false-positive hits resulting from sequence contamination, were verified further by checking genome assembly status and the mapping the flanking sequences of the transposon insertion to the host genome or to the genomes of closely related species; the unmapped transposons and located on very short contig were designated as sequence contamination and were excluded from the analysis. Altogether, we conclude that our final set of VS sequences is highly unlikely to result from contamination artifacts. The copy number VSs in each genome was estimated by using BLAST (40% coverage and 80% identity) and RepeatMasker (4.0.7) (Smit et al. 2015), with the
consensus sequences or representative sequences of VSs (detailed information is provided in supplementary table S1).

Sequence and Phylogenetic Analyses

The potential ORF of the VS used in the present study was analysed using BioEdit (Tippmann 2004). Secondary structures of the VS were predicted using PSIPRED (Mcguffin et al. 2000). Putative nuclear localization signal (NLS) motifs were predicted using PSORT II Prediction as provided on the PSORT WWW server (http://psort.nibb.ac.jp/). Each pairwise identity was also calculated using BioEdit after removing all ambiguous and gapped sites. Base composition of transposons was measured by MEGA software v. 7.0.26 (Model: compute nucleotide composition) (Kumar et al. 2016)

The phylogenetic trees were inferred with an ultrafast bootstrap value of 1000 using the maximum likelihood method within the IQ-TREE program (Nguyen et al. 2015), based on the most conserved domain of the VS transposase (approximately 170 amino acids [aa]) corresponding to the catalytic ‘DDE/D’ domain. The best model was selected using ModelFinder embedded in the IQ-TREE program (Nguyen et al. 2015). Multiple alignments were performed using the MAFFT program (Kazutaka & Standley 2013). Reference transposon sequences of the Tc1/mariner families were downloaded from GenBank, and from the references of DD35E/TR (Zong et al. 2020), DD36E/IC (Sang et al. 2019), and DD37E/TRT (Zhang et al. 2016), and the representative transposase (not consensus) sequences of VSs for all detected species were used for generating more accurate phylogenetic tree.

The pairwise distances between recombination-activating gene 1 (RAG1) and all consensus sequences of VSs in vertebrates were used to test the HT hypothesis. Multiple alignments of RAG1 and VS were created using MUSCLE (Edgar 2004). Then, comparison distances between RAG1 and VS were calculated by using MEGA software v. 7.0.26 based on two aligned files. The accession numbers and host species are listed in supplementary table S2. The genetic distance between RAG1 and VS in each species using a pairwise comparison is listed in supplementary table S3.

The insertion time of each element was estimated using sequence divergences (k) of VSs from the consensus sequences or representative sequences computed by RepeatMasker with the formula \( T = k/2r \), where \( T \) corresponds to the insertion time in millions of years, \( k \) corresponds to the number of nucleotide substitutions per site and \( r \) corresponds to the neutral mutation rate of the species lineage. We used the following neutral mutation rates: \( 2.9590 \times 10^{-9} \) substitutions/site/year for monkey (Sudhir & Sankar 2002);
2.6920 \times 10^{-9} \text{ substitutions/site/year for bat (Poux et al. 2005); 3.2113} \times 10^{-9} \text{ substitutions/site/year for wallaby, which was also used for Tasmanian devil, tammar wallaby and koala; 2} \times 10^{-9} \text{ substitutions/site/year for} \text{Rhinella marina} \text{ (Sequeira et al. 2011); 3.51} \times 10^{-9} \text{ substitutions/site/year for} \text{Oreochromis niloticus and 5.29} \times 10^{-9} \text{ substitutions/site/year for lizard (Eo & DeWoody 2010).}

**Transposition Assay**

**VS Transposase Cloning**

Genomic DNA from six bat species (*Myotis siligorensis, Myotis adversus taiwanensis, Myotis ricketti, Myotis horsfieldii, Myotis parnellii* and *Myotis pusillus*) received as a gift from Dr Liu (Shenyang Agricultural University, China) were used for determination of *VS* existence and *VS* transposase cloning using polymerase chain reaction (PCR) with primers for each transposase. The target products were cloned into the pLB vector (TIANGEN, China) vector and sequenced (TSINGKE, China).

**Plasmid Constructs**

A two-plasmids transposition assay described by Mitra et al. (Rupak et al. 2013) was used. To construct the donor plasmid, an SV4 neo cassette was cloned from pUC19SBneo (Grabundzija et al. 2010) using primers (*VS* TIR/sv40, *VS* TIR/pA) containing *VS* TIRs (30/31 bp). PCR-generated fragments were phosphorylated and inserted into the *SmaI* site of the pUC19 cloning vector, thereby generating pUC19VSneo. The integrity of all coding regions and transposon TIRs was verified by sequencing. To construct the helper plasmid, the cloned transposase ORFs were subcloned into the pCSBNpA vector (Grabundzija et al. 2010) using the XhoI/NotI sites. All primers used for PCR are listed in supplementary table S4.

**Cell Culture and Transfection**

HepG2 cells were maintained in Dulbecco's Modified Eagle Medium (Gibco, US) supplemented with 10% foetal calf serum and 1% Penicillin Streptomycin (Gibco, US). A total of 3 \times 10^5 HepG2 cells were seeded in each well of a six-well plate 1 day prior to transfection. The cells were transfected with 1 µg DNA consisting of 500 ng donor plasmid and 500 ng helper plasmid using 2 µL of TransIT-LT1 Reagent.
(Mirus, US). On 24 hours post transfection, the cells were replanted into 10 cm plates (1% seeding amount for each well) and selected in 1 mg/mL G418 medium. After 2 weeks of selection, the resistant colonies were stained using methylene blue and counted using ImageJ (https://imagej.net/Welcome).

Results

Extensive Distribution of VS Elements in Both Vertebrates and Invertebrates

Previous research has identified VS in one species of oyster (Puzakov et al. 2018), 25 species of insects (including one species of silkworm) (Xie et al. 2018, Bouallège et al. 2017), spanning across eight orders (Osteida, Coleoptera, Diptera, Hemiptera, Homoptera, Hymenoptera, Lepidoptera and Odonata) and two phyla (Mollusca and Arthropoda) of invertebrates. Here, the VS transposons were identified by using the online TBLASTN program with 10 known DD41D and DD40D sequences as queries (Xie et al. 2018, Bouallège et al. 2017) to search against all available prokaryotic and eukaryotic genomes deposited in NCBI database. In total, 194 VS homology elements from 171 species (fig. 1 and supplementary table S1) were obtained and submitted for phylogenetic analysis using the maximum likelihood method in the IQ-Tree program. The bootstrapped phylogenetic tree confirmed that all identified VSs including previous elements were clustered into a single clade within the Tc1/mariner superfamily, with a 99% bootstrap support. This family appears to be much closer to DD37D/maT and DD34D/mariner than to other families (DD34E/Tc1, DD35E/TR, DD36E/IC and DD37E/TRT) in phylogenetic relationship based on the alignment of DDE domains (fig. 2A, see supplementary fig. S1 for uncollapsed tree). Although DD37D/maT and DD34D/mariner are the two close families of DD41D/VS, the sequence identity matrix across families of Tc1/mariner revealed that they shared a very low level of sequence identity (22% and 16%, fig. 2B), indicating that DD41D/VS, DD34D/mariner and DD37D/maT may share an old common ancestor, but excluding the possibility of DD41D emerging from the family of DD34D/mariner or DD37D/maT.

The VS family is widely distributed in both vertebrates and invertebrates. In total, 193 VS elements were identified in 140 species spanning 18 orders of invertebrates and 30 species spanning 10 orders of vertebrates. In addition, a VS element was identified in the olive tree (Olea europaea subsp. Europaea),
which is a land plant species (fig. 1 and supplementary table S1). In detail, the VS transposons were found in 140 species of 18 orders of invertebrates, belonging to one class (Bivalvia) of Mollusca and four classes (Insecta, Arachnida, Maxillopoda and Malacostraca) of Arthropoda, except for those VSs identified previously. Most of these VS elements were found in 134 insect species, which showed the maximum diversity of species across invertebrates. Furthermore, we found the first evidence of VS invasion in vertebrates, including 13 species in three orders (Cichliformes, Characiformes and Synbranchiformes) of ray-finned fishes (Actinopterygii), which represent the highest diversity across vertebrates. Most of the VSs were detected in the Cichliformes order (11 species) of ray-finned fishes. The VS elements were also detected in three species (Eptatretus burgeri, Lethenteron camtschaticum and Petromyzon marinus) of jawless fishes (Agnatha), one species (Rhinella marina) of Anura, one species (Anolis carolinensis) of Squamate and 12 species of mammals (fig. 1 and table 1). In mammals, the VS family had apparently invaded five orders; namely, Chiroptera (seven species), Dasyuromorpha/Marsupialia (one species), Didelphimorpha/Marsupialia (one species), Diprotodontia/Marsupialia (two species) and Primates (one species) (table 2).
The copy number of VS elements varies significantly in these genomes, from 1 to 3151 in each genome (fig. 3 and supplementary table S1). Most of the VSs were short truncated copies in mammals and lizard, although one or two long copies of VSs (>1000 bp) were present; many long copies of VSs were present in a toad, most fishes and even in invertebrates (table 2, supplementary table S1 and supplementary fig. S1). More than half of the species (129/171) contained at least one intact VS element, such as 78% of insect species (104/134), 88% of fishes (14/16) and 58% of mammals (7/12). Intact VS elements were also detected in one species each of Malacostraca, Ostreida, toad and land plants (table 1). The presence of intact copies of VS elements in a land plant and multiple lineages of animals suggests that this family is a young clade that may still be active in these species.

**Evolutionary Dynamics of VSs in Mammals and Non-mammal Vertebrates**

VS transposases were further grouped into five distinct intraclusters (Clusters A–E) based on the phylogenetic tree (fig. 2A), with significant bootstrap support (≥82%). Cluster A included eight species of vertebrates (including five species of mammals, one species of lizard and two species of jawless fish). Cluster B included 31 species of arthropods, while the taxonomic distribution of Cluster C spanned seven species of bats. The fourth cluster, Cluster D, included 25 species of Arthropods, one species of Mollusca and one plant. The fifth cluster, Cluster E, contained 14 species of fishes, one species of toad. Previously reported rosa and Lsra transposases belonged to Clusters B and D (supplementary fig. S1). The phylogenetic tree and Tree of Life demonstrated that the host and VS phylogenies were incongruent (fig. 2A and 3), combining the patchy taxonomic distribution of VSs, suggesting that VS transposons have been exposed to multiple episodes of HT in both invertebrates and vertebrates (fig. 3). The invasion times of VSs in vertebrates estimated by Kimura divergence also supported that the VSs experienced multiple waves of amplification (fig. 4). VS invasion in jawless fish, koala, opossum, tammar wallaby, tasmanian devil and monkey mainly occurred approximately 40 million years ago (Mya) (fig. 4),
indicating a putatively independent invasion event, which is also well supported by the phylogenetic analysis, wherein these VSs form a distinct clade (Cluster A) with 82% bootstrap support. The VSs in bat lineage experienced a relatively recent wave of amplification with a peak activity at 30 Mya and might represent an independent HT event. In addition, very young and weakly active VSs were also observed in three species (Desmodus rotundus, Eptesicus fuscus and Pteronotus parnellii) of bat lineage. While most VSs in ray-finned fishes and cane toad (Rhinella marina) experienced highly similar and young amplifications with a peak at approximately 15 Mya (fig. 4), lizard VS reached its amplification peak at approximately 20 Mya. VSs in some ray-finned fishes and toad species also share very high sequence identities (≥87%, supplementary fig. S2) and form a distinct clade (Cluster E) in the phylogenetic tree (fig. 2A); they belong to species that shared the last common ancestor at approximately 360 Mya (Hedges & Kumar 2004), indicating that these species may have hosted relatively young HT events of the VSs. Furthermore, the HT events in vertebrates were further confirmed using the comparison of pairwise distances between consensus sequences of the VSs and RAG1, which is usually used to infer HT events of transposons in vertebrates (Pace et al. 2008; Zhang et al. 2016). The distances of most pairwise comparisons of VSs (0.133 ± 0.213, range: 0.000–0.708) were much lower than those calculated for RAG1 (0.245 ± 0.278, range: 0.002–0.931), except for those in certain vertebrate species (Otolemur garnettii, Monodelphis domestica, Phascolarctos cinereus, Sarcophilus harrisii and Anolis carolinensis) in the clade of A (supplementary table S3), as indicated in fig. 5. Taken together, these data suggest multiple waves of invasions in vertebrates and that most vertebrate species may obtain VSs by HT, while some by vertical transmission.

The hot vertebrate hosts of VS, TR, IC, and TRT transposons

The taxonomic distribution patterns of VS, TR, IC, and TRT families in vertebrates revealed that some lineages and species of vertebrates are more susceptible to the invasions of Tc1/mariner transposons (fig. 6, supplementary tables S5 and S6). All four families (VS, TR, IC, and TRT) of Tc1/mariner transposons, of which the taxonomic breadths have been well-defined, were
detected in both of Actinopterygii and Anura, while three families (VS, TR, and TRT) were detected in Squamata (fig. 6A). Twelve orders, including Anura, Squamata, and ten orders (Cichliformes, Characiformes, Synbranchiformes, Esociformes, Perciformes, Scombriformes, Tetraodontiformes, Cyprinodontiformes, Salmoniformes, and Cypriniformes) of Actinopterygii, tend to be more common for HTs of Tc1/mariner transposons and have been invaded by at least three families of Tc1/mariner transposons. Particularly, Anura, Cichliformes, and Characiformes were invaded by all four families (VS, TR, IC, and TRT) of Tc1/mariner transposons, and represent the most common orders of HTs (fig. 6B). Further analysis revealed that nineteen species, including two Anura species (Rhinella marina and Xenopus tropicalis) and seventeen Actinopterygii species (Astyanax mexicanus, Haplochromis burtoni, Maylandia zebra, Neolamprologus brichardi, Amphiliophilus citrinellus, Dicentrarchus labrax, Cyprinodon variegatus, Anoplophora fimbria, Salmo salar, Takifugu rubripes, Stegastes partitus, Esox Lucius, Xiphophorus maculatus, Danio rerio, Takifugu flavidus, Larimichthys crocea, Nothobranchius furzeri), have been invaded by at least three families of Tc1/mariner transposons and tend to be more hospitable to Tc1/mariner transposons than others and are the hot hosts of HTs. While Astyanax mexicanus represents the hottest host of HT events and was invaded by all four families (VS, TR, IC, and TRT) of Tc1/mariner transposons (fig. 6C). In addition, we found that the vertebrate and invertebrate Tc1/mariner elements may differ in DNA base composition across families (VS, TR, IC, and TRT), and the vertebrate invaded Tc1/mariner elements tend to be slightly enriched in GC compared with the invertebrate elements (fig. 5D and 5E, and supplementary table S7).

**Structural Variations of VSs**

The structural organization of VSs was illustrated using representative intact VS elements from six classes of animals and one of the land plants (table 1) and is summarized in fig. 7. Most intact VS transposons had a length between 1.3 and 1.7 kb, with an exception of some VSs in insects with a total length of up to 2.4 kb. These VS contained ORFs encoding transposases with 205–429 aa. The transposases were flanked by long or short TIRs and TA target site duplications. Significant variations in TIR lengths of the VSs were observed, such as short TIRs of 27–33 bp.
and long TIRs of approximately 535 bp identified in insects, which were in agreement with the previous study in rosa and Lsra families (Bouallègue et al. 2017). The VS elements in Cluster A were represented by short TIRs (approximately 30 bp long) in mammals and insects, and longer TIRs (>250 bp) in fishes, and several even longer TIRs (225–535 bp) in some species (Anoplophora glabripennis, Rhagoletis zephyria, Myzus persicae, Rhopalosiphum maidis and Tuta absoluta) of insects. In vertebrates, most intact VS transposons in fishes and the toad contained long TIRs of 285 bp and transposases with 304–359 aa. In mammals, the intact copies of VSs represented by a short type of TIRs were only detected in one primate species (Otolemur garnettii) and seven bat species (Desmodus rotundus, Eptesicus fuscus, Miniopterus natalensis, Myotis brandtii, Myotis davidii, Myotis lucifugus and Pteronotus parnellii). Further analysis revealed that although the putatively intact copy of VS had an ORF encoding 345 aa in Otolemur garnettii flanked by 64 and 65 bp TIRs and TA TSDs, the first D residue of the key DDE catalytic domain in the transposase was mutated from D to G (fig. 7), indicating that the VS in this species may have lost transposition activity, a finding that is also supported by the above invasion and extinction age analysis (fig. 4). The intact copies of VSs in bats were >97% identical in nucleotide sequences with sizes of 1276 to 1288 bp and TIRs of 30 or 31 bp, encoding 356, 357 or 359 aa. Considering the insertion age, those intact copies in bats appear to be recent invasions and may be still functional. In the four species of Marsupialia (tasmanian devil/Sarcophilus harrisii, tammar wallaby/Notamacropus eugenii, koala/Phascolarctos cinereus and opossum/Monodelphis domestica), VS homology sequences are also approximately 1.3 kb in size with 345 or 346 aa (at least 92% sequence identity) but TIRs and TSDs were not detectable. By alignment of VS aa, we found that the transposase proteins consist of DNA binding domains (two HTH motifs), DD41D domains and the GRPR-like motifs, which are highly conserved across animal and plant lineages. A slight variation in spacing between the second and third D residues of DDD motif was observed, such as that in some VS transposases in Hymenoptera (one species) and Lepidoptera (six species) represented by the DD40D domain, as noticed previously (Xie et al. 2018; Bouallègue et al. 2017). However, the DD40D elements group in the Cluster B clade of DD41D/Vs in phylogenetic tree.
analysis, indicating that they have originated from DD41D/VS, and thus do not represent a distinct family.

**Transposition Activity Evaluation of Bat VS Elements**

According to our mining data, VSs in bats are compatible with a recent invasion, with intact copies detected in several species, suggesting that functional VSs may exist in this lineage (fig. 4). To test the activity of the intact VSs in bats, we cloned the transposase coding sequences (ORFs) from four bat species (*Myotis siligorensis*, *Myotis adversus taiwanensis*, *Myotis horsfieldii*, and *Myotis pusillus*) using high fidelity PCR with primers designed based on the transposase consensus sequences of seven bat species VSs (*Desmodus rotundus*, *Eptesicus fuscus*, *Miniopterus natalensis*, *Myotis brandii*, *Myotis davidii*, *Myotis lucifugus* and *Pteronotus parnellii*) (supplementary fig. S5A). The cloned transposases were sequenced and showed >90% identity to the identified seven bat species VS transposase sequences (supplementary fig. S3 and S4). The transposases cloned from *Myotis pusillus*, *Myotis adversus taiwanensis* and *Myotis siligorensis* showed 99% sequence identity to the consensus sequences (nucleotides), which were subcloned into a eukaryotic expression vector as helper plasmids and used for the transposition activity assay based on a two-component plasmid system (donor and helper plasmids). The donor plasmid was constructed using a neomycin expression cassette flanked by the consensus TIR sequences of VSs from bats (supplementary fig. S5B). The donor and helper plasmids were co-transfected into human HepG2 cells, and the transfected cells were selected using G418 antibiotic. The results indicated that the VSs from bats did not show any obvious transposition activity when compared with the *Sleeping Beauty* (SB) transposon system that was used as a positive control (supplementary fig. S5C).
Discussion

Invasion of VS family in Mammals

*Tc1/mariner* constitutes an important superfamily of DNA transposons and displays a high intrafamily diversity with approximately 10 families that have been well defined (Sang et al. 2019; Gao et al. 2017). The VS elements, as a family of the *Tc1/mariner* superfamily, have been identified in 26 species of two orders (Mollusca and Arthropoda) of invertebrates (Puzakov et al. 2018). Herein, our data expand the distribution profile and the intrafamily diversity of the VSs. We found that the VS family has a wide taxonomic distribution in the animal kingdom. The data analysis supports a widespread distribution of this family in both invertebrates and vertebrates, particularly in mammals. It was believed that DNA transposons are rare in mammals; however, we observed a patchy phylogenetic distribution of VSs across several lineages of mammals. As indicated by previous genome and mobilome annotations by our lab and other groups, a strong bias of HTs of TEs was observed across the lineages of vertebrates. Some lineages appear to be more susceptible to TE invasion, while some seem to be immune to HTs; for example, bird genomes are represented by very low TE density (approximately 10%) (Gao et al. 2017), while fish display significant variations of TE contents and high diversity of TEs (Vrljicak et al. 2016; Gao et al. 2016) and mammal genomes are represented by high coverages but low diversity of TEs (Hubley et al. 2016; Ongaro et al. 2019). A significant bias of HTs of DNA transposons was also observed for different lineages of mammals (Pace et al. 2008). Herein, we found HTs of VS in seven species of bats, one species of primates (*Otolemur garnettii*) and four species of Marsupialia (*Sarcophilus harrisii, Monodelphis domestica, Notamacropus eugenii* and *Phascolarctos cinereus*). Interestingly, repeated HTs of the hAT superfamily in some species of the same lineages were also confirmed in previous studies, including *Otolemur garnettii* (bush baby) of primates, *Myotis lucifugus* of bats and *Monodelphis domestica* of Marsupialia (Dunemann & Wasmuth 2019; Schaack 2010). In particular, the bat (Chiroptera) lineage has been thought to be more susceptible to HT of transposons than other groups and has experienced HT events of the main DNA
transposons (*hAT*, *piggyBac*, *Tc1/mariner* and *Helitron*). Our data also showed that *VS*s, as a family of the *Tc1/mariner* superfamily, invaded the most common bat species. These data imply that some taxa, such as bats, *Otolemur garnettii* and *Monodelphis domestica*, might be prone to exchanging DNA transposons or more hospitable to DNA transposons than others. However, the invasion mechanism(s) of the DNA transposons in these lineages is still not clear. Bats feed heavily on insects, which might increase bat exposure to HT events. In addition, bats can be infected by several viruses and parasites, which may be another possible reason for the frequent exchange of their genetic materials. These data also suggest that multiple DNA transposon families, including the *VS* family, play a differential role in the evolution of different mammalian genomes. The kissing bug (*Rhodnius prolixus*) has been suggested as the likely intermediate for the spread of DNA transposons *SPIN* and *OCI* (Gilbert et al. 2013). Herein, we also found that the kissing bug, which belongs to Cluster B and has a close relationship with vertebrates, also carried *VS*, suggesting it to be a putative HT vector of the *VS*s in vertebrates.

**Taxonomic distribution and evolutionary dynamics of DD41D/VS, DD35E/TR, and DD36E/IC, and DD37E/TRT families**

Three families of *Tc1/mariner* transposons, including DD35E/TR (Zong et al. 2020), DD36E/IC (Sang et al. 2019), and DD37E/TRT (Zhang et al. 2016), were discovered recently, and their taxonomic distribution and evolutionary dynamics were well-defined. The evolutionary dynamic of DD41D/VS families was very different from these in DD35E/TR (Zong et al. 2020), DD36E/IC (Sang et al. 2019), and DD37E/TRT (Zhang et al. 2016), where most of them dominated by young and recent activities. While *VS*s experienced multiple waves of amplification in animals, particularly in jawless fish and mammals, the amplifications of *VS*s mainly occurred approximately 30 or 40 Mya. DD35E/TR displayed a very restricted taxonomic distribution and was only presented in 91 species of three vertebrate lineages (Actinopterygii, Anura, and Squamata). While both of DD36E/IC (Sang et al. 2019) and DD37E/TRT (Zhang et al. 2016) exhibited more widely distribution in animals, DD36E/IC (Sang et al. 2019) presented in 141
species of four vertebrate lineages (Agnatha, Actinopterygii, Anura, and Mammalia) and 13 species (Arthropoda) of invertebrates, and DD37E/TRT (Zhang et al. 2016) was detected in Actinopterygii, Anura, and Squamata, Protozoans, and Fungi. Gambol was identified very early and the taxonomic distribution was restricted in invertebrates (Coy & Tu 2005). Here, we demonstrated that VS is widespread in the animal kingdom and invaded into two phyla (Mollusca and Arthropoda) of invertebrates and five lineages (Actinopterygii, Agnatha, Anura, Squamate and Eutheria) of vertebrates, which is very similar to that of DD36E/IC (Sang et al. 2019). Furthermore, our data demonstrated again that some orders and species of vertebrates are more susceptible to the invasions of Tcl/mariner transposons and tend to be the hot hosts of HTs, beside the bat lineage. Ecological factors and host–transposon interactions may contribute to the different evolution dynamics of different families and reflect long-term evolutionary trends (Arkhipova 2018; Cosby et al. 2019; Gilbert & Feschotte 2018). Here we found that the differences in base composition of different families may also contribute to the differential evolution dynamics of these families, however, the mechanism of this phenomenon remains largely unknown, which is worth for further study.

Activity of DNA Transposons in Vertebrates and Mammals

Mammals represent a different evolution profile of the TE landscape, with a low activity and diversity of DNA transposons compared with those in fishes, amphibians and reptiles (Klein & O’Neill 2018; Sotero-Caio et al. 2017; Bourgeois & Boissinot 2019). Sequence analysis of TEs in the human, mouse, rat and dog genomes has indicated that mammalian genomes lack recently active DNA transposons and that most DNA transposons in most lineages are fossils and their activity has ceased in the last 50 Mya (Pace & Feschotte 2007; Venter 2001), except for the bat lineage, where multiple waves of recent DNA transposon activity were identified (Mitra et al. 2013). The piggyBac superfamily in its native form from the bat lineage shows transposition activity (Ray et al. 2007; Pagán et al. 2012). Herein, we calculated the invasion time of VSs for different classes of vertebrates and found that most ray-finned fishes and the toad display very
recent insertions, a finding that is in agreement with previous reports, while most mammalian genomes containing VSs undergo a round of old burst, followed by a decrease in abundance, and even extinction. However, very young insertions of VSs in some species of bats were observed, with intact VS elements identified in these species. We also confirmed that the intact VSs from four bat lineages have lost jumping activity, indicating that most VSs in bats may lose their transposition ability due to the accumulation of mutations during evolution. Most key domains, including HTH, GRPR-like and DDE domains, were highly conserved in these bat VS transposases, except the nuclear localization signal (NLS) was not detectable, which plays an important role in transposition (Benjamin et al. 2007; Everitts et al. 2007). The absence of the NLS may be a reason for the loss of transposase function; however, we cannot exclude other reasons, such as mutations in TIRs, which needs further evaluation.

**Phylogenetic Relationship of VSs with Other Tc1/mariner Families**

The evolution profile of the whole Tc1/mariner superfamily is still incomplete, and the mechanisms involved in their invasion and dispersal in different species have been poorly understood. Herein, the phylogenetic tree analysis revealed that the VS family is much closer to DD37D/maT and DD34D/mariner than to other families (DD34E/Tc1, DD37E/TRT, DD35E/IS630, DD35E/TR and DD36E/IC), indicating that they are sister families. However, the low sequence identities across these families (DD37D/maT, DD34D/mariner and DD41D/Vs) exclude the possibility of DD41D/Vs originating from the DD37D/maT or DD34D/mariner family. Based on the phylogenetic position, DDxD/pogo may be the common ancestor family of DD41D/Vs, DD39D, DD37D/maT and DD34D/mariner, which is also well supported by the highly variable spacing of three characteristic residues (D, D, [E or D]) in the catalytic domains of DDxD/pogo compared with the DD34E/Tc1 (Puzakov et al. 2018), DD35E/TR (Zong et al. 2020), DD36E/IC (Sang et al. 2019), and DD37E/TRT (Zhang et al. 2016) families and the extensive distribution of this family in eukaryotic organisms. In addition, our data demonstrated that the DD40D transposons identified in a previous report (Xie et al. 2018) are scattered in Cluster B of
DD41D/VS in the phylogenetic tree and appear to have originated from DD41D/VS, and therefore probably does not constitute a distinct family of the Tc1/mariner superfamily.

**Supplementary Materials**

Supplementary materials (figures S1-S5, dataset S1-S4 and tables S1-S7) are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

**Acknowledgements**

We thank Yang Liu, from Shenyang Agricultural University, for providing the bat genomic DNA. We thank all members of the Chengyi Song Laboratory and the Zoltan Ivics Laboratory for their hard work and thoughtful insights. This work was supported by grants from the Major Projects of National Genetically Modified Organism Breeding [2018ZX08010-08B], Natural Science Foundation of China [31671313], the Priority Academic Program Development of Jiangsu Higher Education Institutions and the High-end Talent Support Program of Yangzhou University to C.S. and the Postgraduate Research & Practical Innovation Program of Jiangsu Province and the Yangzhou University International Academic Exchange Fund to D.S.

**Literature Cited**

Arkhipova IR. 2018. Neutral theory, transposable elements, and eukaryotic genome evolution. Mol Biol Evol. 35(6):1332–1337.

Bouallégue M, et al. 2017. Diversity and evolution of mariner-like elements in aphid genomes. BMC Genomics. 18(1):494.

Bourgeois Y, and Boissinot S. 2019. On the population dynamics of junk: A review on the population genomics of transposable elements. Genes (Basel). 10(6): 419.

Bourque G, et al. 2018. Ten things you should know about transposable elements. Genome Biol. 19(1):199.
Brillet B, Bigot Y, and Augé-Gouillou C. 2007. Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains. Genetica. 130(2):105–120.

Burns KH. 2017. Transposable elements in cancer. Nat Rev Cancer. 17(7):415–424.

Cosby RL, Chang NC, and Feschotte C. 2019. Host–transposon interactions: Conflict, cooperation, and cooption. Genes Dev. 33(17–18):1098–1116.

Coy MR, and Tu Z. 2005. Gambol and Tc1 are two distinct families of DD34E transposons: analysis of the Anopheles gambiae genome expands the diversity of the IS630-Tc1-mariner superfamily. Insect Mol Biol. 14(5):537–546.

de Koning APJ, et al. 2011. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. 7(12):e1002384.

Dunemann SM, and Wasmuth JD. 2019. Horizontal transfer of a retrotransposon between parasitic nematodes and the common shrew. Mob DNA. 10:24.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5:113.

Eo SH, and DeWoody JA. 2010. Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. Proc Biol Sci. 277(1700):3587–3592.

Evertts AG, et al. 2007. The hermes transposon of Musca domestica is an efficient tool for the mutagenesis of Schizosaccharomyces pombe. Genetics. 177(4):2519–2523.

Feschotte C, and Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet. 41:331–368.
Gao B, et al. 2016. The contribution of transposable elements to size variations between four teleost genomes. Mob DNA. 7:4.

Gao B, et al. 2017. Low diversity, activity, and density of transposable elements in five avian genomes. Funct Integr Genomics. 17(4):427–439.

Gilbert C, et al. 2013. Horizontal transfer of OC1 transposons in the Tasmanian devil. BMC Genomics. 14:134.

Gilbert C, and Feschotte C. 2018. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. Curr Opin Genet Dev. 49:15-24.

Gomulski LM, et al. 2001. A new basal subfamily of mariner elements in Ceratitis rosa and other tephritid flies. J Mol Evol. 53(6):597–606.

Grabundzija I, et al. 2010. Comparative analysis of transposable element vector systems in human cells. Mol Ther. 18(6):1200–1209.

Hedges SB, and Kumar S. 2004. Precision of molecular time estimates. Trends Genet. 20(5):242–247.

Hubley R, et al. 2016. The Dfam database of repetitive DNA families. Nucleic Acids Res. 44(D1):D81–89.

Jurka J. 2000. Repbase Update: A database and an electronic journal of repetitive elements. Trends Genet. 16(9):418–420.

Kalendar R, Lee D, and Schulman AH. 2014. FastPCR software for PCR, in silico PCR, and oligonucleotide assembly and analysis. Methods Mol Biol. 1116:271–302.

Katoh K, and Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Klein SJ, and O’Neill RJ. 2018. Transposable elements: genome innovation, chromosome diversity, and centromere conflict. Chromosome Res. 26(1–2):5–23.

Kumar S, and Subramanian S. 2002. Mutation rates in mammalian genomes. Proc Natl Acad Sci USA. 99(2):803–808.

Kumar S, Stecher G, and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Bio Evol. 33(7):1870–1874.

Laha T, et al. 2007. The bandit, a new DNA transposon from a hookworm—possible horizontal genetic transfer between host and parasite. PLoS Negl Trop Dis. 1(1):e35.

Lander ES. 2001. Initial sequencing and analysis of the human genome. Nature. 409(6822):860–921.

Lohe AR, and Hartl DL. 1996. Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol Biol Evol. 13(4):549–555.

McGuffin LJ, Bryson K, and Jones DT. 2000. The PSIPRED protein structure prediction server. Bioinformatics. 16(4):404–405.

Mitra R, et al. 2013. Functional characterization of piggyBat from the bat Myotis lucifugus unveils an active mammalian DNA transposon. Proc Natl Acad Sci USA. 110(1):234–239.

Waterston RH, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature. 420(6915):520–562.

Nguyen LT, et al. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.

Ongaro L, et al. 2019. The genomic impact of European colonization of the Americas. Curr Biol. 29(23):3974–3986.e4.
Pace JK, and Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res. 17(4):422–432.

Pace 2nd JK, et al. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci USA. 105(44):17023–17028.

Pagán HJT, et al. 2012. Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vespertilionid bats. Genome Biol Evol. 4(4):575–585.

Poux C, et al. 2005. Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. Syst Biol. 54(5):719–730.

Puzakov MV, Puzakova LV, and Cheresiz SV. 2018. An analysis of IS630/Tc1/mariner transposons in the genome of a Pacific oyster, Crassostrea gigas. J Mol Evol. 86(8):566–580.

Ray DA, et al. 2007. Bats with hATs: evidence for recent DNA transposon activity in genus Myotis. Mol Biol Evol. 24(3):632–639.

Sang Y, et al. 2019. Incomer, a DD36E family of Tc1/mariner transposons newly discovered in animals. Mob DNA. 10:45.

Schaack S, Gilbert C, and Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol. 25(9):537–546.

Sequeira F, et al. 2011. Hybridization and massive mtDNA unidirectional introgression between the closely related neotropical toads Rhinella marina and R. schneideri inferred from mtDNA and nuclear markers. BMC Evol Biol. 11:264.

Sergeeva EM, and Salina EA. 2011. Transposable elements and plant genome evolution. Russ J Genet Appl Res. 1:565–576.
Smit A, Hubley R, and Grenn P. 2015. RepeatMasker Open-4.0.

Sotero-Caio CG, et al. 2017. Evolution and diversity of transposable elements in vertebrate genomes. Genome Biol Evol. 9(1):161–177.

Subramanian B, et al. 2019. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. Nucleic Acids Res. 47(W1):W270–W275.

Tang Z, et al. 2015. Repeated horizontal transfers of four DNA transposons in invertebrates and bats. Mob DNA. 6(1):3.

Tippmann HF. 2004. Analysis for free: comparing programs for sequence analysis. Brief Bioinform. 5(1):82–87.

Vrljicak P, et al. 2016. Genome-wide analysis of transposon and retroviral insertions reveals preferential integrations in regions of DNA flexibility. G3 (Bethesda). 6(4):805–817.

Waterston RH, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature. 420(6915):520–562

Xie LQ, et al. 2018. Genome-wide identification and evolution of TC1/Mariner in the silkworm (Bombyx mori) genome. Genes Genomics. 40(5):485–495.

Zhang HH, et al. 2016. TRT, a vertebrate and protozoan Tc1-like transposon: current activity and horizontal transfer. Genome Biol Evol. 8(9):2994–3005.

Zhang HH, et al. 2016. Identification and evolutionary history of the DD41D transposons in insects. Genes Genom. 38:109–117.

Zong W, et al. 2020. Traveler, a New DD35E Family of Tc1/Mariner Transposons, Invaded Vertebrates Very Recently. Genome Biol Evol. 12(3): 66–76.
**Figure legends**

**Fig. 1** Species distribution of *VS* transposons identified in eukaryotes. The number of species (s) that contain *VS* elements is presented for each order in 171 diverse species ranging from the plant to animal kingdoms. There are two different types of *VSs* in terms of catalytic domains, including DD40D and DD41D. The presence of *VS* elements coding these two different domains is indicated by ‘+’.

**Fig. 2** Phylogenetic relationships of the *VS* family. (A) Phylogenetic tree of *VS* elements identified in this study with reference families of the *Tc1/mariner* superfamily based on the DDE domains. Bootstrapped (1000 replicates) phylogenetic trees were inferred using the maximum likelihood method in IQ-TREE (Nguyen et al. 2014). Each sequence (except the DD39D and DD37E subclasses) contains the name of the transposon, the gene sequence number corresponding to the transposon and the Latin abbreviation of the species in which the transposon is located. Five distinct intraclusters (Clusters A–E) of *VS* were grouped based on the phylogenetic tree. (B) Sequence identities between the *VS* family and nine other families. The sequence identities were measured using pairwise comparisons of full-length transposases.

**Fig. 3** Presence and coverage of *VS* elements across eukaryotes. The Tree of Life was used to infer a tree of the 100 species used in this study. EvolView (Balakrishnan et al. 2019) was used to generate the bar graph. Leaves are coloured to indicate which species have both *VS_Cluster B* and *VS_Cluster D* (yellow), only *VS_Cluster A* (violet), only *VS_Cluster B* (blue), only *VS_Cluster C* (grey), only *VS_Cluster D* (light sky blue) and only *VS_Cluster E* (orange). Black bars correspond to the copies of *VS* elements in each species. Connections indicate possible HT events involving invertebrates (orange) or vertebrates (blue).

**Fig. 4** The invasion profiles of *VS* elements in vertebrates. The insertion time of each element was estimated using the formula $T = \frac{k}{2r}$, where $T$ corresponds to the insertion time in millions of years, $k$ corresponds to the number of nucleotide substitutions per site, and $r$ corresponds to the neutral mutation rate of the species lineage.

**Fig. 5** Pairwise distances of *VS* elements and *RAG1*. The distances are obtained from all possible pairwise comparisons (n=58; labelled on the x axis) between six species (Cluster A), six species (Cluster C) and eight species (Cluster E) in which *VS* was identified. The genetic distance between *RAG1* and *VS* in each vertebrate species using a pairwise comparison is listed in supplementary table S3. The coding sequence (CDS) regions of *RAG1* in the NCBI database are available (dataset S3).
Fig. 6 Taxonomic distribution patterns of VS, TR, IC, and TRT families in vertebrate. (A) VS, TR, IC, and TRT transposons distribution in vertebrate lineages. The numbers of species/orders detected for each family indicated for each lineage. (B and C) Venn of distribution patterns across vertebrate orders and species (detailed information of species invaded was listed in supplementary tables S5 and S6). (D and E) Base composition and GC content of VS, TR, IC, and TRT elements in vertebrate and invertebrate. Full elements from 10 vertebrate hosts for each family (VS, TR, IC, and TRT), full elements from 10 invertebrate hosts for VS and Gambol families were used to compare the differences of base composition (detailed information of transposons used was listed in dataset S4). Base composition account and GC contents were summarized in supplementary table S7.

Fig. 7 Structure of intact VS transposons. The structures of eleven transposons from seven classes that contain VS transposons are shown. The blue arrows represent TIRs, white boxes represent transposase domains, black bars represent helices, red boxes represent catalytic domains, green ellipses represent GRPR-like motifs, yellow box represents NLS, and the numbers represent the length of TIRs and transposase. Otga, Otolemur garnettii; Modo, Monodelphis domestica; Mina, Miniopterus natalensis; Anca, Anolis carolinensis; Rhma, Rhinella marina; Orni, Oreochromis niloticus; Epbu, Eptatretus burgeri; Agnl, Anoplophora glabripennis; Duno, Dufourea novaeangliae; Crgi, Crassostrea gigas; Oleu, Olea europaea subsp. Europaea.
fig. 1
fig. 2
fig. 6
Table 1

Summary of V$\alpha$ Elements in Diverse Genomes Ranging from Plant to Animal Kingdoms

| Phylum   | Taxa Distribution (Class) | Number of Species Containing a V$\alpha$ | Number of Species Containing an Intact V$\alpha$ | Length of V$\alpha$ (kb) | Transposase Length of V$\alpha$ (aa) | TIR Length of V$\alpha$ | TSD |
|----------|---------------------------|------------------------------------------|-----------------------------------------------|------------------------|--------------------------------------|-----------------------|-----|
| Chordata | Mammalia Primates         | 1                                        | 1                                             | 1.6                    | 345                                  | 65/64                 | TA  |
|          | Mammalia Marsupialia      | 4                                        | 0                                             | 1.3                    | 345/346                              | -                     | -   |
|          | Mammalia Chiroptera       | 7                                        | 6                                             | 1.3                    | 345–359                              | 30/31                 | TA  |
|          | Reptilia                  | 1                                        | 0                                             | -                      | 348                                  | -                     | -   |
|          | Amphibia                  | 1                                        | 1                                             | 1.7                    | 358                                  | 285                   | TA  |
|          | Actinopterygii            | 13                                       | 13                                            | 1.6                    | 304–359                              | 256–285               | TA  |
|          | Agnatha                   | 3                                        | 1                                             | 1.2                    | 343–374                              | 32                    | -   |
| Arthropoda| Insecta                  | 134                                      | 104                                           | 1.2–2.4                | 205–429                              | 14–535                | TA  |
|          | Arachnida                 | 1                                        | 0                                             | -                      | 375                                  | -                     | -   |
|          | Maxillopoda               | 1                                        | 0                                             | -                      | 355                                  | -                     | -   |
|          | Malacostraca              | 3                                        | 1                                             | 1.3                    | 357–361                              | 29–33                 | TA  |
| Mollusca | Bivalvia                  | 1                                        | 1                                             | 1.5                    | 234/362                              | 31–33                 | TA  |
| Plants   | Angiosperms               | 1                                        | 1                                             | 1.3                    | 348                                  | 27                    | TA  |
| Class      | Species Name                          | Length of VS | Transposase Length of VS | Copy Number* | Copy Number# | TIR Length of VS | Identity of TIRs% | TSD |
|------------|---------------------------------------|--------------|--------------------------|--------------|--------------|------------------|-------------------|-----|
| Mammal     | Otolemur garnettii/Primates           | 1636         | 345                      | 35           | 0            | 64/65            | 78                | TA  |
|            | Sarcophilus harrisii/Marsupialia      | -            | 346                      | 1            | 0            | -                | -                 | -   |
|            | Monodelphis domestica/Marsupialia     | -            | 345                      | 1            | 0            | -                | -                 | -   |
|            | Notamacropus eugenii/Marsupialia      | -            | 346                      | 5            | 0            | -                | -                 | -   |
|            | Phascolarctos cinereus/Marsupialia    | -            | 346                      | 2            | 0            | -                | -                 | -   |
|            | Desmodus rotundus/Bats                | 1220         | 357                      | 8            | 0            | -                | -                 | -   |
|            | Eptesicus fuscus/Bats                 | 1285         | 357                      | 3            | 1            | 31/30            | 90                | TA  |
|            | Miniopterus natalensis/Bats           | 1278         | 357                      | 1            | 1            | 31/30            | 87                | TA  |
|            | Myotis branditi/Bats                  | 1288         | 357                      | 2            | 1            | 31/30            | 81                | TA  |
|            | Myotis davidii/Bats                   | 1288         | 356                      | 2            | 1            | 31/30            | 81                | TA  |
|            | Myotis lucifugus/Bats                 | 1288         | 357                      | 2            | 1            | 30/31            | 84                | TA  |
|            | Pteronotus parnellii/Bats             | 1276         | 359                      | 1            | 1            | 30/31            | 87                | TA  |
| Lizard     | Anolis carolinensis/Reptilia          | -            | 348                      | 2            | 0            | -                | -                 | -   |
| Cane toad  | Rhinella marina/Amphibia              | 1655         | 358                      | 2645         | 12           | 285              | 95                | TA  |
| Fish       | Amphipholus citrinellus/Ray-finned fishes | 1660   | 304                      | 15           | 1            | 285              | 96                | TA  |
|            | Astatotilapia calliptera/Ray-finned fishes | 1654   | 352                      | 36           | 43           | 284/285          | 99                | TA  |
|            | Haplochromis burtoni/Ray-finned fishes | 1656   | 352                      | 200          | 4            | 283/284          | 97                | TA  |
|            | Labeotropheus fuelleborni/Ray-finned fishes | 1656   | 330                      | 39           | 1            | 285/287          | 94                | TA  |
| Species                                      | Ray-finned fishes | TA |
|---------------------------------------------|-------------------|----|
| *Maylandia zebra*                           |                   | TA |
| *Mchenga conophoros*                        |                   | TA |
| *Neolamprologus brichardi*                  |                   | TA |
| *Oreochromis niloticus*                      |                   | TA |
| *Pundamilia nyererei*                       |                   | TA |
| *Rhamphochromis esox*                       |                   | TA |
| *Simochromis diagramma*                     |                   | TA |
| *Astyanax mexicanus*                        |                   | TA |
| *Mastacembelus armatus*                     |                   | TA |
| *Eptatretus burger*                         |                   | TA |
| *Lethenteron camtschaticum*                 |                   | TA |
| *Petromyzon marinus*                        |                   | TA |

NOTE. — *, copy number of sequences with > 40% query coverage and > 80% identity; #, copy number of intact transposons; - represents not detectable.