Genes Genom (2015) 37:153–159
DOI 10.1007/s13258-014-0241-6

RESEARCH ARTICLE

Evolutionary fate of SVA2 elements in primate genomes

Yonsik Kwak · Yun-ji Kim · Jinchuan Xing · Kyudong Han

Received: 12 October 2014 / Accepted: 30 October 2014 / Published online: 7 November 2014

Abstract SVA (SINE/VNTR/*Alu*) transposable element, one of non-LTR retrotransposons, emerged in the primate genome about 25 million years ago. Currently, ~2,800 SVA copies exist in the human genome. Recently, a group of transposable elements named SVA2 is discovered. SVA2 elements share the VNTR region with the SVA element but do not contain the SINE-R region of the SVA elements. In this study, we studied the SVA2 evolution and the impact of the SVA2 elements on primate genomes. We first identified 144, 139, 136, 139, and 116 SVA2 elements in the human, chimpanzee, gorilla, orangutan, and rhesus macaque genomes, respectively. To examine the evolutionary state and structure of the elements, we performed comparative genomics, comparing human SVA2 with its orthologous counterpart from non-human primates. The result suggests that SVA2 subfamily is not, at present, retrotranspositionally active in the primate genomes because none of the SVA2 elements identified in this study are species-specific. In addition, we found that four human SVA2 elements locate in human genes and two of them have miRNA target sites, indicating that they might regulate gene expression and involve in the gene-related human diseases.

Keywords SVA element · SVA2 element · Primate · Evolutionary history

Electronic supplementary material The online version of this article (doi:10.1007/s13258-014-0241-6) contains supplementary material, which is available to authorized users.

Y. Kwak · Y. Kim · K. Han (✉)
Department of Nanobiomedical Science & BK21 PLUS NBM Global Research Center for Regenerative Medicine, Dankook University, Cheonan 330-714, Republic of Korea
e-mail: kyudong.han@gmail.com

Y. Kwak · Y. Kim · K. Han
DKU-Theragen Institute for NGS Analysis (DTiNa), Cheonan 330-714, Republic of Korea

J. Xing
Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

J. Xing
Human Genetics Institute of New Jersey, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Introduction

Mobile genetic elements including DNA transposon and retrotransposon account for approximately 44 % of the human genome (Batzer and Deininger 2002). DNA transposons are transposed into a new genomic region using a "cut and paste" mechanism while retrotransposons are mobilized using a "copy and paste" mechanism. Retrotansposons are subdivided into two groups: autonomous versus non-autonomous. Autonomous elements can produce proteins which are required for their transposition (Ostertag et al. 2001). Long interspersed element and ERV (endogenous retrovirus) families are well-known autonomous elements. On the other hand, non-autonomous elements are not able to encode the proteins but use the enzymatic machinery produced by autonomous elements for their mobilization (Boeke 1997; Callinan and Batzer 2006). short interspersed element (SINE) and SVA families belong to non-autonomous elements.

SVA elements are distributed across all human chromosomes and about 2,800 copies have been identified in the human genome (Wang et al. 2005). SVA element was integrated into the primate genome after the divergence of

Old World monkey and other ape lineages (<25 million years ago) (Wang et al. 2005). SVA element is considered one of the youngest retrotransposons in primate genomes. Previous studies about SVA elements suggested that the elements are still retrotranspositionally active and thus could alter genomic architecture in the hominid primate genomes (Wang et al. 2005; Goodier and Kazazian 2008). SVA element is capable to cause genetic diseases by integrating into genes randomly (Ostertag et al. 2003). SVA elements mobilize using a mechanism called "TPRT (target primed reverse transcription)". In this process, SVA element is transcribed into RNA and the RNA intermediate is reversely transcribed into a new genomic region, leading to 7–20-bp-long target site duplications (TSDs) neighboring each side of the new SVA element (Luan et al. 1993; Cost et al. 2002; Christensen and Eickbush 2005).

SVA2 was considered the precursor of SVA elements in previous studies (Jurka et al. 2005; Damert et al. 2009; Hancks et al. 2012). Approximately 100 copies of SVA precursors, SVA2, were identified in the rhesus macaque genome (Han et al. 2007). Since then, four SVA2 elements have been found in the human chromosome 19 (Damert et al. 2009). SVA subfamilies are classified into six different subfamilies named SVA_A to SVA_F, based on SINE-R sequence (Wang et al. 2005). SVA elements of the six subfamilies have SINE-R, *Alu*-like sequences and VNTR while SVA2 elements are lack of SINE-R and *Alu*-like sequences, but contain VNTR and heterologous non-SVA sequence (Damert et al. 2009). Thus, the previously identified SVA2 elements did not belong to any of the six SVA subfamilies. It limited further studies about SVA2 and SVA2 elements in other primates have not been identified. In this study, we investigated SVA2 elements in primate genomes. First, we computationally collected SVA2 elements from the human and four non-human primate genomes and experimentally verified them using wet-bench techniques. Through the analyses, we identified more than 100 SVA2 elements in each of the primate genomes and traced the evolutionary history of SVA2 elements. In addition, we detected four human SVA2 elements inserted at 3′ UTR of exonic regions of genes (*ABHD2*, *TMEM69*, *CCDC137*, and *C1QTNF7*) in the human genome.

Materials and methods

Identification of SVA2 elements

To identify SVA2 elements, we constructed a domestic SVA2 consensus sequence, based on a SVA2 consensus sequence from GIRI (http://www.girinst.org) and four human SVA2 sequences previously identified in the human genome. Using the domestic SVA2 sequence as a query to BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat), we extracted SVA2 candidates from five different primate genomes (hg19; February 2009 freeze, PanTro4; February 2011 freeze, gorGor3; May 2011 freeze, ponAbe2; July 2007 freeze, rheMac3; October 2010 freeze). Then, we computationally validated the candidates using GIRI and RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) and also manually inspected them through a close examination focusing on the sequence structure of the SVA2 candidates.

PCR validation

For the verification of the candidates, we amplified them using PCR assay using five different DNA samples of human (*Homo sapiens*; Coriell Cell Repository, Camden, NJ, USA), chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*les) and rhesus macaque (*Macaca mulatta*) as a DNA template. Genomic DNA for four apes was kindly provided by Dr. Takenaka Primate Research Institute, Kyoto University). We designed a pair of specific primers for each SVA2 candidate and computationally simulated the PCR with the primers using UCSC In-silico PCR. In case where SVA2 candidate locates on unsequenced genomic region (N sequences) of the primate genome in interest, we designed primers based on the nucleotide sequence of its orthologous locus from other primate genomes. Primers which worked with the In-silico PCR were used to amplify SVA2 candidates. Each PCR reaction contained 10 μl of Taq DNA polymerase, 200 nM of each oligonucleotide primer, and 10 ng of target template DNA. PCR amplification was initiated with denaturation step of 5 min at 95 °C, followed by 35 cycles of denaturation step of 30 s at 95 °C, annealing step of 40 s at the optimal annealing temperature, and extension step of 1–4 min at 68 °C. 5., and ended with a final extension step of 10 min at 68 °C. PCR products were confirmed by 1 % agarose gel electrophoresis, stained with Eco Dye (BioFact, Korea), and visualized using UV fluorescence.

Genomic environment analysis

To investigate the genomic environment of human SVA2 element, we analyzed the gene density and GC content of the genomic regions flanking human SVA2 elements. Gene density was calculated in 2 Mb flanking sequence centered on each human SVA2 element by using UCSC genome browser (http://genome.ucsc.edu). GC content was estimated in 20 kb flanking sequence centered on each human SVA2 element by using the EMBOSS GeeCee server (http://emboss.bioinformatics.nl/cgi-bin/emboss/geecee).

Results and discussion

SVA2 element in primate genomes

Using the domestic SVA2 consensus sequence as a query to BLAT, we identified SVA2 candidates from five different primate genomes; 196 SVA2 candidates in the human genome (hg19; February 2009), 196 SVA2 candidates in the chimpanzee genome (PanTro4; February 2011), 195 SVA2 candidates in the gorilla genome (gorGor3; May. 2011), 204 SVA2 candidates in the orangutan genome (ponAbe2; July 2007), and 150 SVA2 candidates in the rhesus macaque genome (rheMac3; October. 2010). The 941 SVA2 candidates were subjected to a manual inspection and 275 false positive were eliminated from the initial data. The remaining SVA2 candidates were further verified by PCR analysis and DNA sequencing of the PCR products. Seven of the verified SVA2 elements contained a partially unsequenced region but we were able to verify them through the combined method of PCR and DNA sequencing of the PCR products, as described in detail in the materials and method section. In addition, we manually inspected the SVA2 candidates using structural analyses based on their TSD and SVA2-specific sequences. In final, we confirmed that 144, 139, 136, 139, and 116 SVA2 elements are present in the human, chimpanzee, gorilla, orangutan, and rhesus macaque genomes, respectively (Table 1).

Structure of human SVA2 element

SVA2 element consists of VNTR and heterologous non-SVA sequence (Shen et al. 1994; Strichman-Almashanu et al. 2002). Nonetheless, both ends of the elements are flanked by TSDs which are a hallmark of TPRT mechanism (Fig. 1). We examined the size of SVA2 elements present in the human genome (Fig. 2). The size was variable, ranging from 68 to 2,906 bp. The average size of 144 human SVA2 elements was approximately 455 bp, shorter than that of human SVA elements. As we expected, the size difference of the elements mostly resulted from VNTR variation (Fig. 2). The size of the human SVA2 VNTR ranged from 45 to 2,803 bp. Interestingly, we found nine truncated SVA2 elements (Fig. S1). They had non-heterologous SVA region but were deficient of VNTR. The VNTRs have a high sequence similarity between them, which is likely to cause the recombination between them. We believe the recombination between VNTRs led to the truncated form of SVA2 elements but could not rule out other mechanisms.

In addition, we analyzed TSDs of the human SVA2 elements. TSDs were variable in terms of its nucleotide sequence and length, which means that there is no preferential genomic target site for the insertion of human SVA2 element. The size of the TSDs was in range of 4–21 bp, with 9 bp average (Table S1).

Classification	Number of loci				
	Human	Chimpanzee	Gorilla	Orangutan	Rhesus macaque
Blat results	196	196	195	204	150
Discarded after manual inspection	52	58	65	66	34
Confirmed SVA2 elements	144	139	130	138	116
Confirmed by PCR analysis	–	–	6	1	–
Total	144	139	136	139	116

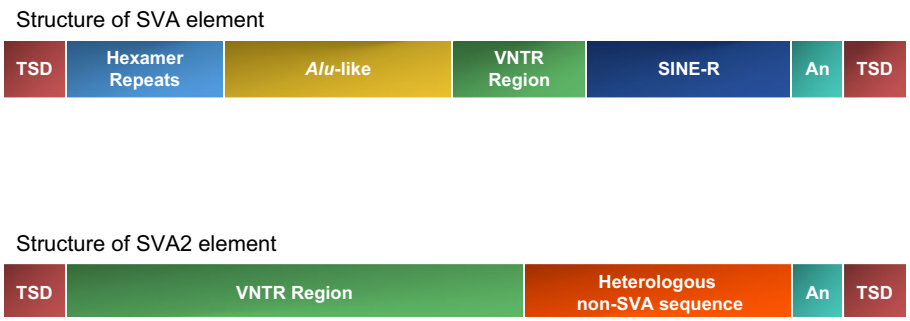

Fig. 1 Typical structure of SVA and SVA2 elements. A full-length SVA element includes hexamer repeats, *Alu*-like, VNTR, and SINE-R sequences, flanked by TSDs. SVA2 element consists of VNTR and heterologous non-SVA sequence followed by a poly-a tail and is also flanked by TSDs

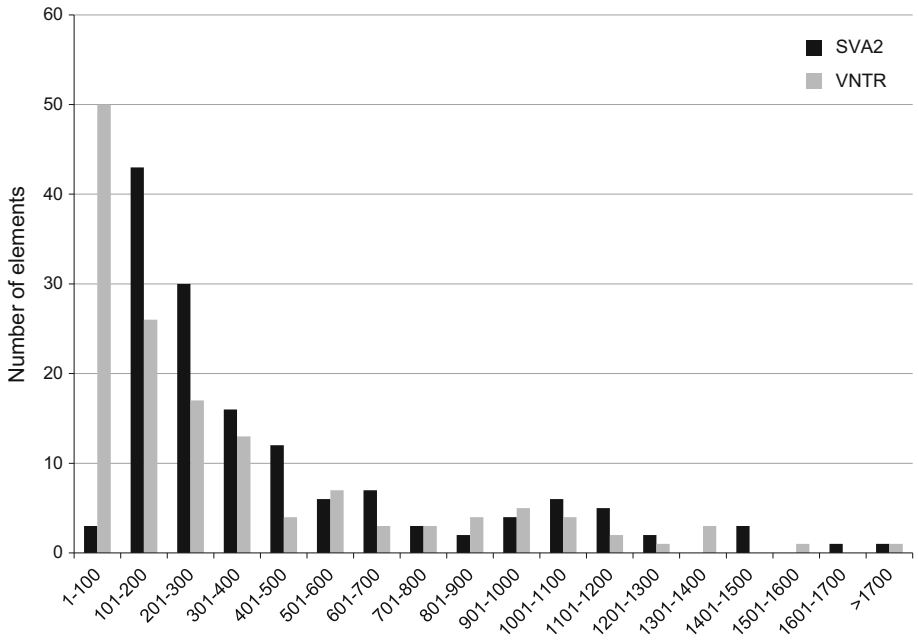

Fig. 2 Size distribution of human SVA2 elements. About half of human SVA2 elements belong to a size bin ranging from 101 to 200 bp. This figure shows that the sizes of SVA2 elements and SVA2 VNTR are skewed to the small size

Genomic environment of human SVA2 elements

To study the genomic environments of the human SVA2 elements, we estimated the density of SVA2 elements per the human chromosome (the number of SVA2 elements/the length of each chromosome). The identified SVA2 elements were randomly distributed across the human chromosomes but chromosomes 18 and 21 contain no SVA2 element (Fig. 3). SVA2 elements were present in high copy number on chromosomes 1, 2, and 7 but rare on chromosomes 15, 16, and Y. The chromosome with the highest density of SVA2 element was chromosome 19; although chromosome 19 has less copy number of SVA2 elements compared with chromosomes 1, 2, and 7, its length is much shorter than the three chromosomes. On the other hand, human chromosome 15 exhibited the lowest density of SVA2 elements (Table S2) except for chromosomes 18 and 21. The chromosomal position of human SVA2 elements was visualized by using Idiographica which is a web base tool (Fig. 4).

The genomic environment flanking each human SVA2 element was analyzed, focusing on GC content and gene density. We calculated the GC content in 20 kb of flanking sequences centered on a SVA2 element. The average GC content was 41.9 % which is similar to that of the human reference genome sequence. We calculated the gene density in 2 Mb of flanking sequences centered on a SVA2 element. The average gene density was about 22 per 2 Mb. We also analyzed genetic position of the human SVA2 elements. The result showed that 90 of the 144 SVA2 elements exist in the intergenic region while 50 SVA2 elements are located in the intronic region. Interestingly, four SVA2 elements were embedded in the exonic region. The genes containing SVA element in their exonic region are *ABHD2*, *TMEM69*, *CCDC137*, and *C1QTNF7*, which encode abhydrolase domain-containing protein 2, transmembrane protein 69, coiled-coil domain-containing protein 137, and complement C1q tumor necrosis factor-related protein 7 isoform, respectively (Table S3). All four SVA2 elements were located in 3′ UTR of the respective gene (Fig. S2). We examined if the SVA2 elements have a miRNA target site by using microRNA database (http://www.microrna.org). SVA2 element in the 3′ UTR of *TMEM69* had seven miRNA target sites and the element residing on the 3′UTR of *CCDC137* had 15 miRNA target sites which are SVA2-specific (Table S4). The other two exonic SVA2 elements did not have miRNA target sites due to the direction of the two SVA2 elements; SVA2 elements within *TMEM69* and *CCDC137* are in the sense direction (5′–3′) while the SVA2 elements within *ABHD2* and *C1QTNF7* are in the antisense direction (3′–5′). The SVA2 elements locating on the exonic region has a potential to involve in causing human phenotypic changes or diseases. In addition, the SVA2 elements in genes, *TMEM69* and *CCDC13,* could affect the expression levels of those genes by providing the miRNA target sites.

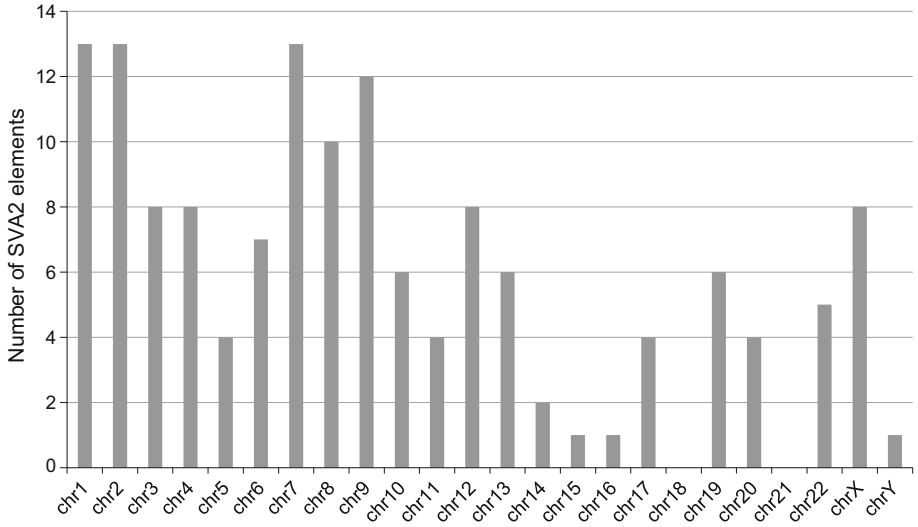

Fig. 3 Chromosomal density of human SVA2 elements. Human SVA2 elements are dispersed throughout all of the chromosomes but not in chromosome 18 and 21. Chromosome 1, 2 and 7 show a relatively high density of SVA2 elements, compared to the other chromosomes

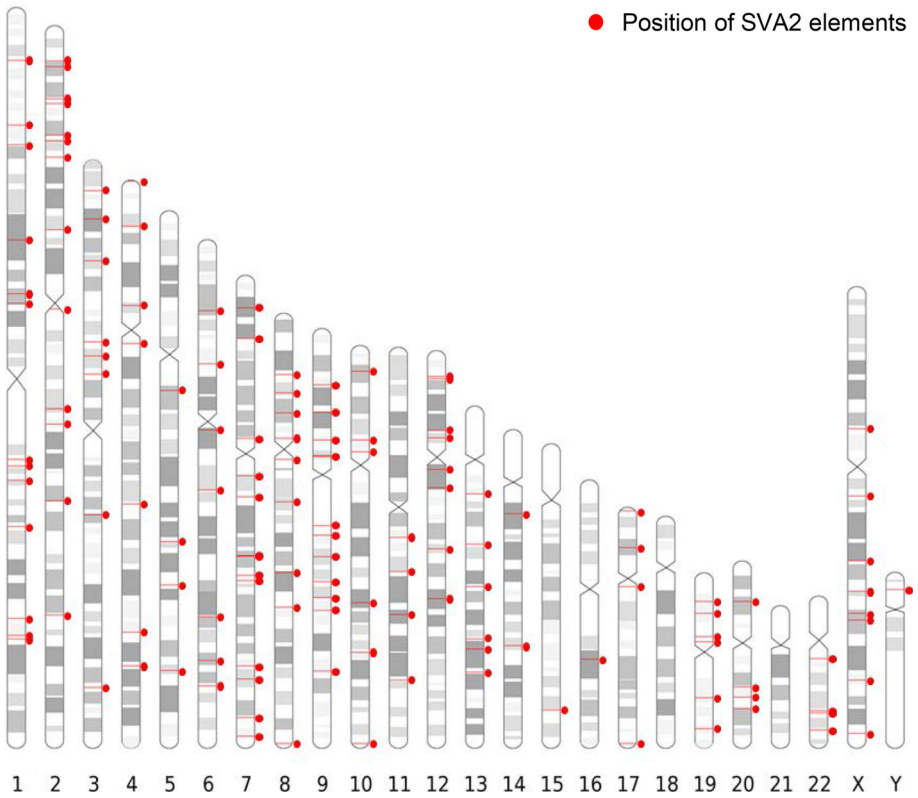

Fig. 4 Visualization of SVA2 elements in the human genome. The Idiographica figure (Kin and Ono 2007) visualizes the chromosomal position of human SVA2 elements

Evolution of SVA2 elements

Approximately 100 precursors of SVA were previously identified in the rhesus macaque genome (Han et al. 2007). The SVA precursors (now known as SVA2 elements) had VNTR and non-homologous SVA sequences. SVA2 VNTR region is similar to the SVA VNTR region. In addition, SVA elements propagated in the hominid

Table 2 Summary of evolutionary state for SVA2 elements in the primate genomes

Classification	Number of loci				
	Human	Chimpanzee	Gorilla	Orangutan	Rhesus macaque
Evolutionary state	133	130	131	114	82
Incomplete lineage sorting	11	7	2	9	7
Not present in the human genome	–	2	3	16	27
Total	144	139	136	139	116

genomes after the divergence of hominid and Old World primates. However, the suggestion that SVA2 element appeared before speciation event between Old World primates and great apes supported that SVA2 could be a precursor of SVA elements. Therefore, the SVA2 elements are found not only in the rhesus macaque genome but also in hominid genomes. Therefore, we wanted to examine whether the SVA2 elements are still retrotranspostionally active in the primate genomes. We tried to identify species-specific SVA2 elements because the existence of species-specific SVA2s indicate recent retrotransposition events. All of the SVA2 elements identified from the five different primate genomes were compared with their orthologs. For example, human SVA2 was compared with its counterparts from the other four non-human primates. Through this analysis, we found that 82 SVA2 elements were inserted into the primate genome before the divergence of Old World monkey and great apes. The 26 and 11 SVA2 elements were propagated in the host genome before and after the divergence of orangutan and gorilla, respectively. In the beginning of this study, we considered that 11 human SVA2 elements are "incomplete lineage sorting loci". The incomplete lineage sorting is generated by the presence of a polymorphic allele in a common ancestor that becomes either fixed or extinct in the genomes of daughter species (Ray et al. 2006). The 11 human SVA2 elements were actually inserted into the host genome before the divergence of human and other apes but became extinct in one or two genomes of daughter species (Table 2). In contrast to the events, species-specific segmental duplication increases the copy number of SVA2 elements in the respective genome. Those species-specific events are able to generate false species-specific SVA2 elements. However, through a thorough examination, we found that there is no species-specific SVA2 element in primate genomes, implying that SVA2 elements are currently inactive.

Conclusions

We, for the first time, conducted a comprehensive study on SVA2 elements in five different primate genomes. SVA2 element is considered a precursor of SVA elements because the two elements share a VNTR region. Unlike SVA element which was originated after the divergence of Old World monkey and great apes, SVA2 elements emerged before the divergence. Since then, SVA elements had been successfully propagated in a common ancestor of humans, chimpanzees, and gorillas. However, the result of this study suggests that SVA2 element is currently retrotranspositionally inactive in the primate genomes. Although SVA2 elements are not able to produce their new progenies, they have a potential to cause genomic rearrangement due to the characteristics of repeat elements. In addition, we found four human SVA2 elements existing in the exonic region (i.e. 3′ UTR) of human genes. Interestingly, two of the SVA2 elements contained miRNA target sites. Thus, the two SVA2 elements could regulate gene expressions and involve in the gene-related human diseases via the binding to miRNAs.

Acknowledgments The present work was conducted with funding from the Research Fund of Dankook University in 2013.

Conflict of interest The authors declare that there is no conflict of interests exists in this paper.

References

Batzer MA, Deininger PL (2002) Alu repeats and human genomic diversity. Nat Rev Genet 3:370–379

Boeke JD (1997) LINEs and Alus—the polyA connection. Nat Genet 16:6–7

Callinan PA, Batzer MA (2006) Retrotransposable elements and human disease. Genome Dyn 1:104–115

Christensen SM, Eickbush TH (2005) R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. Mol Cell Biol 25:6617–6628

Cost GJ, Feng Q, Jacquier A, Boeke JD (2002) Human L1 element target-primed reverse transcription in vitro. EMBO J 21:5899–5910

Damert A, Raiz J, Horn AV, Löwer J, Wang H, Xing J, Batzer MA, Löwer R, Schumann GG (2009) 5′-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. Genome Res 19:1992–2008

Goodier JL, Kazazian HH Jr (2008) Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell 135:23–35

Han K, Konkel MK, Xing J, Wang H, Lee J, Meyer TJ, Huang CT, Sandifer E, Hebert K, Barnes EW et al (2007) Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. Science 316:238–240

Hancks DC, Mandal PK, Cheung LE, Kazazian HH Jr (2012) The minimal active human SVA retrotransposon requires only the 5'-hexamer and Alu-like domains. Mol Cell Biol 32:4718–4726

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467

Kin T, Ono Y (2007) Idiographica: a general-purpose web application to build idiograms on-demand for human, mouse and rat. Bioinformatics 23:2945–2946

Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotranscription. Cell 72:595–605

Ostertag EM, Kazazian HH Jr (2001) Biology of mammalian L1 retrotransposons. Annu Rev Genet 35:501–538

Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr (2003) SVA elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet 73:1444–1451

Ray DA, Xing J, Salem AH, Batzer MA (2006) SINEs of a nearly perfect character. Syst Biol 55:928–935

Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, Zipf WB, Yu CY (1994) structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene duplication. J Biol Chem 269:8466–8476

Strichman-Almashanu LZ, Lee RS, Onyango PO, Perlman E, Flam F, Frieman MB, Feinberg AP (2002) A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. Genome Res 12:543–554

Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA (2005) SVA elements: a hominid-specific retroposon family. J Mol Biol 354:994–1007