Cross-Kingdom Commonality of a Novel Insertion Signature of RTE-Related Short Retroposons

Eri Nishiyama and Kazuhiko Ohshima*

Graduate School of Bioscience, Nagahama Institute of Bio-Science and Technology, Shiga, Japan

*Corresponding author: E-mail: k_ohshima@nagahama-i-bio.ac.jp.

Abstract

In multicellular organisms, such as vertebrates and flowering plants, horizontal transfer (HT) of genetic information is thought to be a rare event. However, recent findings unveiled unexpectedly frequent HT of RTE-clade LINEs. To elucidate the molecular footprints of the genomic integration machinery of RTE-related retroposons, the sequence patterns surrounding the insertion sites of plant Au-like SINE families were analyzed in the genomes of a wide variety of flowering plants. A novel and remarkable finding regarding target site duplications (TSDs) for SINEs was they start with thymine approximately one helical pitch (ten nucleotides) downstream of a thymine stretch. This TSD pattern was found in RTE-clade LINEs, which share the 3′-end sequence of these SINEs, in the genome of leguminous plants. These results demonstrably show that Au-like SINEs were mobilized by the enzymatic machinery of RTE-clade LINEs. Further, we discovered the same TSD pattern in animal SINEs from lizard and mammals, in which the RTE-clade LINEs sharing the 3′-end sequence with these animal SINEs showed a distinct TSD pattern. Moreover, a significant correlation was observed between the first nucleotide of TSDs and microsatellite-like sequences found at the 3′-ends of SINEs and LINEs. We propose that RTE-encoded protein could preferentially bind to a DNA region that contains a thymine stretch to cleave a phosphodiester bond downstream of the stretch. Further, determination of cleavage sites and/or efficiency of primer sites for reverse transcription may depend on microsatellite-like repeats in the RNA template. Such a unique mechanism may have enabled retroposons to successfully expand in frontier genomes after HT.

Key words: SINEs, LINEs, target site duplication, horizontal gene transfer.

Introduction

Eukaryotic genomes contain an extraordinary number of retroposons such as long terminal repeat (LTR) retrotransposons, long interspersed repetitive elements (LINEs) or non-LTR retrotransposons, and short interspersed repetitive elements (SINEs) (Weiner et al. 1986; Brosius 1991; Kazazian 2004; Jurka et al. 2005; Bennetzen and Wang 2014). Because of the insertion mechanism of LINEs: target DNA-primed reverse transcription (TPRT) (Luan et al. 1993; Cost et al. 2002; Eickbush and Eickbush 2015), DNA cleavage specificity of endonuclease (EN) domain primarily determines the site of LINE insertion (Luan et al. 1993; Feng et al. 1996; Maita et al. 2007). Apurinic/apyrimidinic EN (APE)-like ENs are encoded by over 20 clades of LINEs that insert at many different loci within their host genome, some of which have shown weak target site preferences (Szak et al. 2002; Zingler et al. 2005; Bringaud et al. 2006); although only two clades, Tx1 and R1, contain site-specific LINEs (Fujiwara 2015; Nichuguti et al. 2016). Integration at a specific site also depends on other factors, such as the structural parameters of the target DNA and interactions between the mRNA and the target DNA (Cost and Boeke 1998; Repanas et al. 2007; Monot et al. 2013; Fujiwara 2015).

Human L1 preferentially inserts at 5′-TT|AAAA-3′, where "|" indicates the site of insertion (Szak et al. 2002; Morrish et al. 2002, 2007), and its EN cleaves the TpA bond in 5′-TTTTAA-3′ on the complementary strand (Feng et al. 1996; Cost and Boeke 1998). TPRT usually results in the duplication of a short stretch of nucleotides (mostly no >20 bp) resulting from integration at staggered chromosomal breaks. Thus, each newly inserted element is typically flanked by short direct repeats, which are also known as a target site duplication (TSD) (Beck et al. 2011). To date, the analysis of TSDs from LINEs is largely confined to mammalian L1s. Using target analysis of nested transposons for genomic copies, Ichiyanagi and Okada (2008) studied TSDs for a variety of vertebrate LINEs, including those of the L1, L2, CR1, and RTE clades in mammalian, chicken, and zebrafish genomes.

SINEs are nonautonomous retroposons, the 5′-end sequences of which are derived from tRNA, 5S rRNA, or 7SL RNA with promoter activity for RNA polymerase III (Okada 1991; Batzer and Deininger 2002; Kapitonov and Jurka 2003; Ohshima 2013; Vassetzky and Kramerov 2013; Ahl et al. 2015). Mammalian L1s mobilize nonautonomous sequences such as SINE RNA and cytosolic mRNA by recognizing the 3′-poly(A) tail of the template RNA (Doucet et al. 2015), resulting in enormous SINE amplification and processed pseudogene formation. The 3′-end sequences of various SINEs originated from corresponding LINEs other than L1 (Ohshima et al. 1996), however, and to date, ~60 of these SINE/LINE pairs have been identified (Ohshima 2012; Vassetzky and Kramerov 2013). As the 3′-UTRs of several LINEs have been shown to be essential for retroposition, these LINEs presumably require stringent recognition of the 3′-end sequence of the RNA template (Okada et al. 1997; Kajikawa and Okada 2002; Eickbush and Eickbush 2012; Hayashi et al. 2014). The analyses of TSDs from SINEs have provided valuable clues to the enzymatic source for SINE retroposition (Jurka 1997; Lenoir et al. 2001; Wenke et al. 2011; Noll et al. 2015; Schwichtenberg et al. 2016).

AfroSINEs (Nikaido et al. 2003) are a SINE family in the genomes of afrotherians, which are African endemic mammals, proposed to be derived from and have been mobilized by RTE-clade LINE (Bov-B) because these two elements share a highly similar sequence (Gogolevsky et al. 2008). Because AfroSINEs and known elephant RTE-clade LINE are not terminated by the same tandem repeat motifs, Gilbert et al. (2008) proposed that these differences reflect constraints imposed by base pairing interactions between the mRNA 3′ terminal tandem repeats and the target DNA at the initiation of TPRT.

Plant genomes harbor a wide variety of SINE families (Mochizuki et al. 1992; Yoshioka et al. 1993; Deragon et al. 1994; Yasui et al. 2001; Xu et al. 2005; Deragon and Zhang 2006; Cognat et al. 2008; Tsuchimoto et al. 2008; Baucom et al. 2009; Gadzalski and Sakowicz 2011; Wenke et al. 2011; Schwichtenberg et al. 2016). Only three SINE/LINE pairs have been discovered: namely, maize ZmSINE2 and ZmSINE3 (LINE1-1_ZM: Baucom et al. 2009) and tobacco TS SINE (SolRTE-I_Nt: Wenke et al. 2011; RTE-1_STu: Ohshima 2012). High similarity of the Au SINE family between distantly related plant species has been reported (Fawcett et al. 2006). Although their phylogenetic distribution was patchy, Fawcett and Innan (2016) identified several copies present in the orthologous regions of various species, including species that diverged 90 Ma, thereby confirming the presence of Au SINE at multiple evolutionary time points. Therefore, the Au SINE appears to have been present in the common ancestor of all angiosperms being retained in some lineages while lost from others.

In multicellular organisms, such as vertebrates and flowering plants, horizontal transfer (HT) of genetic information is thought to be a rare event (Kidwell 1993). However, the number of well-supported cases of transfer from eukaryotes is now expanding rapidly (Bock 2010; Schaack et al. 2010; Wallau et al. 2012; Ivancevic et al. 2013; Fuentes et al. 2014; Peccoud et al. 2017). Recently, unexpectedly frequent HT of RTE-clade LINEs was reported. Walsh et al. (2013) showed that HT of Bov-B LINEs (Kordiš and Gubenšek 1998; Malik and Eickbush 1998; Župunski et al. 2001) was significantly more widespread than believed, and they demonstrated the existence of two plausible arthropod vectors, specifically reptile ticks. Their analysis indicated that at least nine HT events are required to explain the observed topology. Suh et al. (2016) showed that the genomes of nematodes and seven tropical bird lineages exclusively share a novel LINE, AviRTE, which resulted from HT. The HTs between bird and nematode genomes were estimated to have taken place 25–22 and 20–17 Ma.

In the present study, to elucidate the molecular footprints of the genomic integration machinery of RTE-related retroposons, the sequence patterns surrounding insertion sites of plant Au-like SINE families were analyzed in the genomes of a wide variety of flowering plants. There was a remarkable tendency of TSDs in SINEs, and moreover, the same TSD pattern was also found in plant RTE-clade LINEs and even in animal SINEs. Based on these observations, a model for the initial process of genomic integration of these retroposons is proposed, and the relationship between rampant HTs of RTE-clade LINEs and the mechanism is discussed.

Materials and Methods

Genomic Sequences

Plant genome sequences were obtained from Ensembl Plants (Bolser et al. 2017) and the Genome Database for Rosaceae (Jung et al. 2014). Animal genome sequences were obtained from Ensembl (Aken et al. 2017). supplementary table S1, Supplementary Material online shows a list.

Construction of Consensus Sequences

The consensus sequences (CONS) for 1) the RTE from common wheat (*Triticum aestivum*; TAe) and SINEs from 2) barrel clover (*Medicago truncatula*; MT), 3) purple false brome (*Brachypodium distachyon*; BDi), and 4) sorghum (*Sorghum bicolor*; SBi) were constructed from BLAST searches (Altschul et al. 1990) using an *E*-value of 5E-10. 1) BLAST against the common wheat genome using RTE-1_TD from durum wheat (*Triticum durum*) as the query resulted in ca. 6,000 hits, of which 30 randomly chosen sequences over 3,000 bases in length were used to construct the CONS (supplementary fig. S9, Supplementary Material online). 2) BLAST against the barrel clover genome using SINE2-1_TAe from common wheat as the query resulted in six hits, and the CONS from these sequences detected 374 sequences. Thirty randomly chosen sequences and the initial six sequences were used to

derive the final CONS (supplementary fig. S10, Supplementary Material online). 3) BLAST against the purple false brome genome using Au SINE from *Aegilops umbellulata* as the query resulted in 24 hits. CONS from these sequences detected 43 sequences from which the final CONS was generated (supplementary fig. S11, Supplementary Material online). 4) BLAST against sorghum genome using SINE2-1_ZM from maize as the query resulted in 25 hits. CONS from 16 sequences with high scores detected 26 higher-quality sequences that were used in the final CONS (supplementary fig. S12, Supplementary Material online). Regarding the soybean Au-like SINE, the sequence reported by Shu et al. (2011) (GmAu1) was used as the consensus sequence. The sequence of the Sauria SINE of green anole (clone ACA-1-15; GenBank: FJ158974) was obtained from Piskurek et al. (2009). The sequence of an *Oryzias* RTE of medaka fish (clone OlRTE-a03; GenBank: AB021490) was obtained from Župunski et al. (2001), and the sequence of a lizard RTE of green anole (clone AcRTE-a01; GenBank: AAWZ01014759) was obtained from Tay et al. (2010). All remaining sequences were obtained from Repbase (Jurka et al. 2005; Bao et al. 2015).

Search for TSDs

Using the CONS as queries, a series of BLAST searches were performed against the respective genomes with an *E*-value of 5*E*-10 used in all cases. Detected sequences plus 200 bases of their 5′ and 3′ flanking sequence were extracted from genomic sequences. Within these sequences, we searched for TSDs with a Python script using the following criteria: 1) TSD length is between 10 and 49 bases inclusive, 2) the 5′ and 3′ TSD sequences are perfectly matched, and 3) the 5′ and 3′ TSD sequences are separated by at least 100 bases. The copy numbers of LINEs and SINEs, and the number of TSDs detected are shown in table 1 for the respective species. It is possible that they are subsets of the copies (young family members) since we used a stringent parameter for BLAST search (for potato Au-like SINEs, see Wenke et al. 2011 and Seibt et al. 2016).

Analysis of Nucleotide Compositions and Motif Discovery

The 5′ TSD sequences with their flanking sequences from respective copies of SINE and LINE families were extracted from the genomic sequences of the corresponding species. The nucleotide composition of each family was plotted on a chart for every nucleotide position. To test whether there was a biased composition between two consecutive nucleotides, the χ^2 test was performed according to Jurka (1997) (supplementary fig. S1, Supplementary Material online; 15 degrees of freedom, significant level of 0.005). The nucleotide composition was also represented graphically by WebLogo (Crooks et al. 2004) (supplementary fig. S2, Supplementary Material online). The MEME motif discovery algorism (Bailey

and Elkan 1994) was applied to the TSD data sets. The MEME suite 4.11.2 (Bailey et al. 2015) was used with the following parameters by 'Terminal client': minimum motif width, 15; maximum motif width, 30; minimum sites per motif, N (number of analyzed TSDs) × 0.25; maximum sites per motif, N. The most statistically significant (low *E*-value) motifs were used for further analyses (supplementary table S2, Supplementary Material online).

Estimating the Occurrences of a Specific Trinucleotide near the 3′-Ends of Each Copy

To estimate the association of each copy with microsatellite-like sequence at the 3′-ends, the occurrences of a specific trinucleotide near the 3′-ends of each copy were examined. Ten bases of 3′-ends of BLAST-detected sequences plus ten bases of their 3′ flanking sequences were extracted from genomic sequences. Within these sequences, a specific trinucleotide was searched for with a Python script. The results are summarized in supplementary table S3, Supplementary Material online.

3D Model of RTE EN

The 3D structure of the EN domain from the LINEs with indiscriminate integration sites was previously determined for only human L1. Using human L1-EN (Protein Data Bank ID: 1vyb) as a template, 3D models of soybean RTE-EN were constructed with MODELLER (Fiser and Šali 2003) in Chimera (Pettersen et al. 2004). Of the five models generated, the model with the highest scores (GA341 = 1.00, zDOPE = −0.28) was selected for further analyses.

Results

Plant Au-like SINEs and RTE-Clade LINEs Share 3′-Terminal Sequences

We analyzed the characteristics of Au-like SINE sequences from various angiosperms identified based on sequence similarity to known Au SINEs. Figure 1 shows sequence comparisons of the full-length Au-like SINEs and the 3′-terminal sequence of a potato RTE (RTE-1_STu). Nucleotide sequences of the 3′-terminal region of the RTE (positions 3991–4069; supplementary figs. S6–S8, Supplementary Material online) and Au-like SINEs (positions 69–144) were very similar (pairwise distances: 0.135–0.362), a finding which suggests this region is essential for retroposition. Nucleotide positions 127–144 of the SINEs and the corresponding region of the RTE-clade LINEs were predicted to form a hairpin-like RNA secondary structure, which was conserved with several compensatory mutations (fig. 2). Since the RNA secondary structures of the 3′-terminal region from several LINEs are essential to initiate reverse transcription, it is highly plausible that Au-like SINEs have retrotransposed with the RTE-clade LINE machinery.

Table 1

Copy Numbers of LINEs and SINEs and the Number of Analyzed TSDs

	RTE-Clade LINEs			RTE-Related SINEs		
	Family	# of Copies	TSD	Family	# of Copies	TSD
Glycine max	RTE-1_GM	1,120	813	GmAu1	1,451	1,044
Medicago truncatula	RTE1_MT	667	305	MT_AUlikeSINE_cons	374	224
Malus domestica	RTE-1_Mad	856 (21,691)[a]	423	SINE-5_Mad	147 (2,025)[a]	97
	RTE-1B_Mad	714 (9,890)[a]	304			
Solanum tuberosum	RTE-1_STu	743	315	SINE2-2_STu	62	24
	RTE-2_STu	70	24			
Brachypodium distachyon	RTE-1_BDi	60	23	BDi_consensus_24	43	27
Triticum aestivum	TAe_RTE_cons	6,222	2,486	SINE2-1_TAe	2,308	1,062
Sorghum bicolor	RTE-1_SBi	95	30	SBi_AU_cons	26	12
Zea mays	RTE1_ZM	996	518	RST_ZmSINE1	268	180
	RTE2_ZM	596	416	RST_AU	16	6
				SINE2-1_ZM	200	85
Equus caballus	RTE-1_EC	606	340	SINE2-1_EC	4,712	1,613
Bos taurus	Bov-B	359,044	218,458	BOVTA	362,502	201,054
Loxodonta africana	RTE1_LA	193,947	124,680	AFROSINE-1_LA	6,877 (9,862)[b]	2,075
				AFROSINE-2_LA	10,315	2,983
				AFROSINE	135,168	54,407
				AFROSINE1B	14,921	6,166
				AFROSINE2	6,353 (34,868)[c]	2,042
				AFROSINE3	19,686	5,185
Procavia capensis	RTE1_Pca	297 (1160)[a]	188	PSINE1	164	66
				SINE2-1_Pca	141	26
Echinops telfairi	RTE1_ET	280 (950)[a]	187			
Ornithorhynchus anatinus	Plat_RTE1	369	78			
Anolis carolinensis	RTE_BOV_B_AC_1	15,625	7,122	Sauria SINE	78,442	33,597
	RTE-1_AC_1	10,450	5,671			
	AcRTE-a01	26	11			
Oryzias latipes	RTE-1_OL	3,650	1,229			
	RTE-2_OL	2,839	811			
	RTE-3_OL	449	187			
	OlRTE-a03	2,753	974			
Takifugu rubripes	Expander	345	93			
	EXPANDER2	209	63			
Caenorhabditis elegans	RTE-1	53	30			

[a]The number of copies analyzed with the total number of copies shown in parentheses.

[b]The number of copies following exclusion of those with hits to AFROSINE-1_LA and AFROSINE-2_LA. The total number of hits is shown in parentheses.

[c]The number of copies following exclusion of those with hits to AFROSINE2 and AFROSINE or AFROSINE1B. The total number of hits is shown in parentheses.

A Novel Insertion Signature of Plant RTE-Related Retroposons

We conducted TSD analyses for Au-like SINEs and RTE-clade LINEs from different flowering plants and found a novel insertion signature that is specific to these retroposons. Figure 3A shows the nucleotide composition of the genomic sequences surrounding the first nucleotide (P1) of the 5′ TSD of Au-like SINEs (left) and RTE-clade LINEs (right) from soybean (upper) and *Medicago* (lower), respectively. The P1 was frequently thymine (T) for both Au-like SINEs and RTE-clade LINEs, and moreover, we observed a prominent excess of T, often a stretch of ~5 Ts, near P−10 (refer to supplementary fig. S2,

Supplementary Material online for sequence logos). Such a feature at a remote position has not been reported for L1-clade LINEs. Figure 3B shows the nucleotide motifs found by the MEME motif discovery algorism in the same soybean data sets. Consistently, remarkable motifs which consist of a stretch of T and single T were found in both data sets from Au-like SINE (upper) and RTE-clade LINE (lower) (for statistical information, see supplementary table S2, Supplementary Material online). The same profile was also found in Au-like SINEs from other flowering plants, such as wheat, corn, and apples (supplementary fig. S1, Supplementary Material online). These results indicate that Au-like SINEs were amplified via reverse transcription with a unique machinery of RTE-clade LINEs.
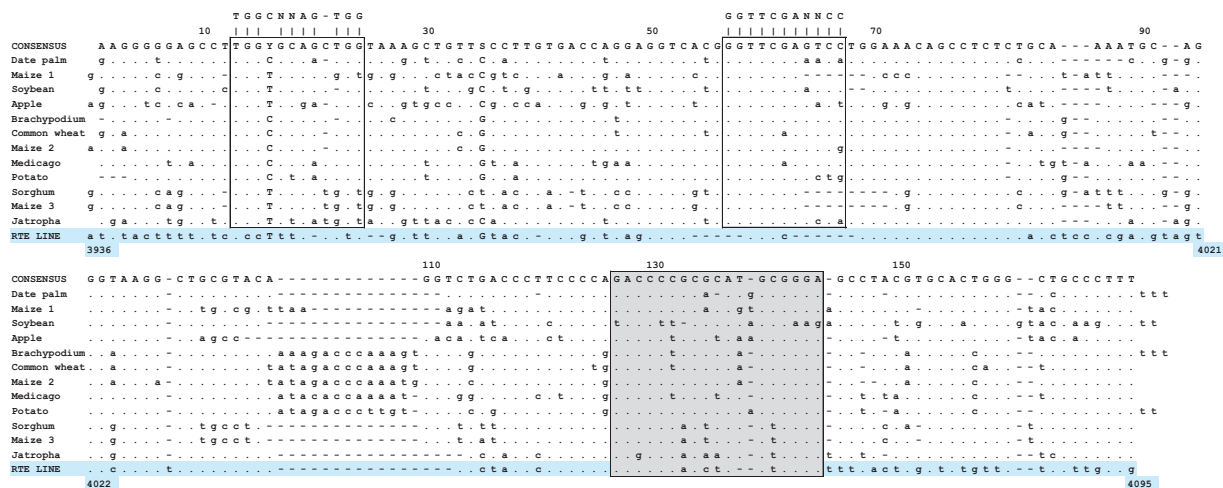
Fig. 1.—Sequence comparisons of Au-like SINEs and the 3′-terminal sequence of an RTE. The entire sequence of Au-like SINEs and the 3′-terminal sequence (~160 nucleotides) of a potato RTE-clade LINE (RTE-1_STu) (light blue) are aligned. Dots and hyphens represent identical nucleotides to the consensus sequence (shown at top) and gaps, respectively. Nucleotide positions of the SINEs and the LINE are shown on the top and bottom, respectively. The two internal promoters for RNA polymerase III (box A: positions 13–24; box B: 57–67) are shown in open boxes with the consensus sequences. Nucleotide positions (127–144) predicted to form a hairpin-like RNA secondary structure are shown in the grey box.
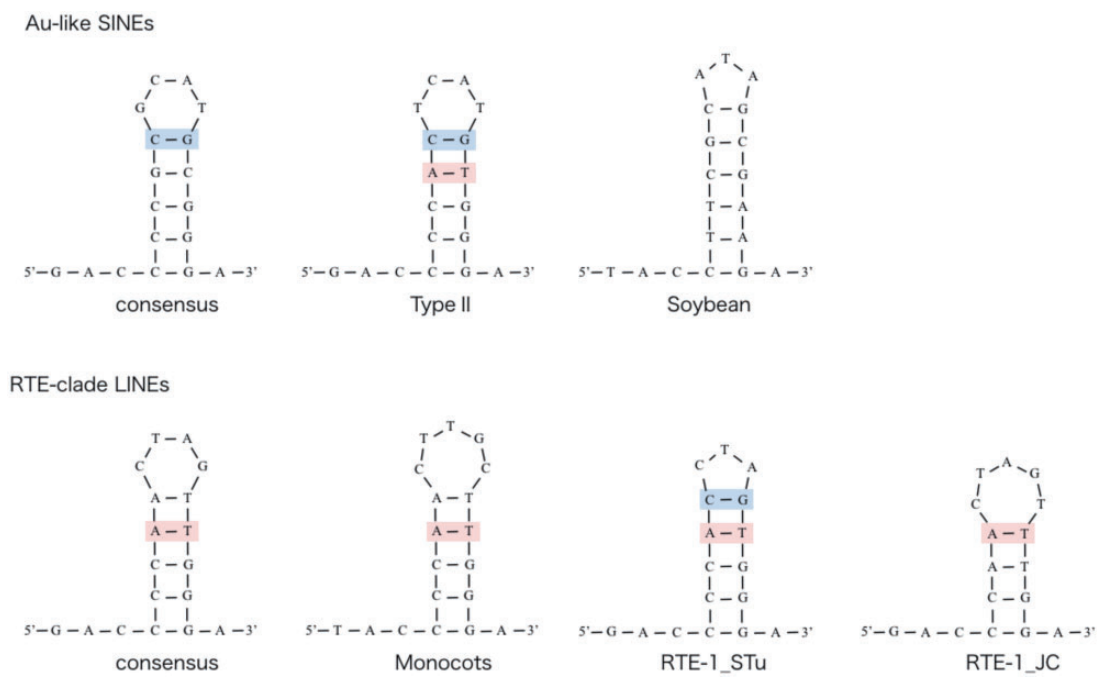

Fig. 2.—Secondary structure models for the 3′-terminal sequences of Au-like SINEs and RTE-clade LINEs. Transcripts from this region may form putative hairpin structures. Compensatory mutations, (A: T) ↔ (G: C) or (C: G) ↔ (A: T), are shown by pink and blue rectangles, respectively.

Characteristics of the EN Domain of Plant RTE-Clade LINEs

To understand the molecular basis of the unique TSD pattern of plant RTE-clade LINEs, we investigated characteristics of the EN domain of plant RTE-clade LINEs. Figure 4A shows comparisons of essential amino acid residues for EN activity (Weichenrieder et al. 2004) between RTE-clade LINEs and other LINEs. These amino acid residues are highly conserved among plant RTE-clade LINEs and other LINEs. Interestingly, residue 229 of plant RTEs was substituted to glutamine, whereas the residue at this position is aspartic acid in every other LINE including animal RTEs (fig. 4A). Since this amino acid residue does not participate in coordinating magnesium ions (Beernink et al. 2001; Weichenrieder et al. 2004), we posit that this D229Q substitution does not dramatically
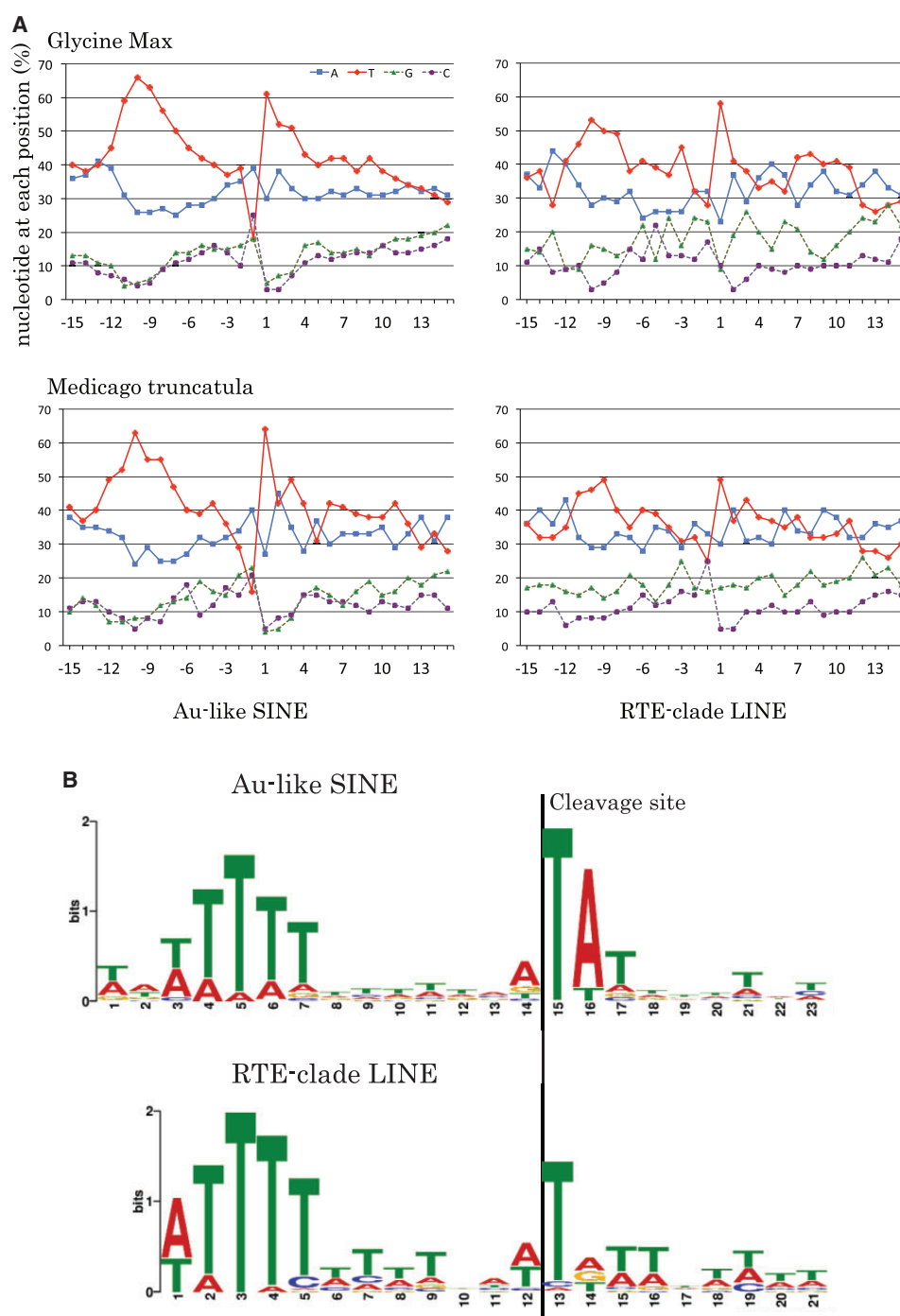
A

Glycine Max

Medicago truncatula

Au-like SINE   RTE-clade LINE

B

Au-like SINE

RTE-clade LINE

Fig. 3.—Nucleotide composition and motifs surrounding the first nucleotide of 5′ TSDs from plant retroposons. (*A*) Nucleotide composition. Thirty nucleotide positions are shown with the first nucleotide of the 5′ TSD at the center (position 1: P1). Nucleotide compositions at respective positions are represented graphically: T (red), A (blue), G (green), and C (purple). Au-like SINEs (left) and RTE-clade LINEs (right) are shown from soybean (upper: $n = 1,044$; 813, respectively) and *Medicago* (lower: $n = 224$; 305). Note that P1 is frequently T and a prominent excess of T is found at approximately P−10. The same profile is also found in other plants (supplementary fig. S1, Supplementary Material online). (*B*) Discovered motifs for soybean SINE and LINE. The MEME motif discovery algorism, which uses a finite mixture model, was applied to the same data set as (A) (supplementary table S2, Supplementary Material online). Au-like SINE (upper) and RTE-clade LINE (lower) from soybean are shown.

A

Clade	Name	Species	14	43	115	145	147	192	205	228	229	230
L1	L1HS	Human	N	E	Y	D	N	T	D	S	**D**	H
R1	R1Bm	Silkworm	N	E	Y	D	N	T	D	S	**D**	H
R1	TRAS1	Silkworm	N	E	Y	D	N	T	D	S	**D**	H
	BovB	Cow	N	E	Y	D	N	T	D	S	**D**	H
	RTE1_LA	Elephant	N	E	Y	D	N	T	D	T	**D**	H
	RTE-1_MD	Opossum	N	E	Y	D	N	T	D	T	**D**	H
	RTE-1	Nematode	N	E	Y	D	N	T	D	S	**D**	H
	RTE-1_GM	Soybean	N	E	Y	D	N	T	D	T	**Q**	H
RTE	RTE-1_Alp	Peanut	N	E	Y	D	N	T	D	T	**Q**	H
	RTE-1_MN	Mulberry	N	E	Y	D	N	T	D	T	**Q**	H
	RTE-1_Mad	Apple	N	E	Y	D	N	T	D	N	**Q**	H
	RTE-1B_Mad	Apple	N	E	Y	D	N	T	D	N	**Q**	H
	RTE-1_STu	Potato	N	E	Y	D	N	T	D	T	**Q**	H
	RTE-1_TD	Durum wheat	N	E	Y	D	N	T	D	T	**Q**	H
	RTE2_ZM	Maize	N	E	Y	D	N	T	D	T	**Q**	H

B

Clade	Name	Species					195						200					205
L1	L1HS	Human	T	F	F	S	A	P	-	-	H	H	T	Y	S	K	I	D
R1	R1Bm	Silkworm	T	F	S	T	A	N	G	-	-	-	-	E	S	Y	V	D
R1	TRAS1	Silkworm	T	F	D	T	I	R	G	G	K	R	Y	Q	S	R	V	D
	BovB	Cow	T	W	T	S	P	D	-	-	G	Q	H	R	N	Q	I	D
	RTE1_LA	Elephant	T	W	T	S	P	D	-	-	G	T	H	R	N	Q	I	D
	RTE-1_MD	Opossum	T	W	M	H	P	R	-	-	S	K	Q	W	H	L	I	D
	RTE-1	Nematode	T	F	V	S	P	D	-	-	G	N	H	R	H	E	I	D
	RTE-1_GM	Soybean	T	Y	K	S	-	-	-	-	G	G	T	C	S	Q	I	D
	RTE-1_Alp	Peanut	T	Y	K	S	-	-	-	-	G	M	T	S	S	Q	I	D
RTE	RTE-1_MN	Mulberry	T	F	K	N	-	-	-	-	G	R	N	L	S	Q	I	D
	RTE-1_Mad	Apple	T	Y	K	S	-	-	-	-	G	S	S	K	T	Q	I	D
	RTE-1B_Mad	Apple	T	Y	K	S	-	-	-	-	G	S	S	K	T	Q	I	D
	RTE-1_STu	Potato	T	F	R	S	-	-	-	-	P	V	A	K	T	Q	I	D
	RTE-1_TD	Durum wheat	T	F	S	S	-	-	-	-	G	Q	H	S	S	Q	I	D
	RTE1_ZM	Maize	T	F	S	S	-	-	-	-	G	H	R	S	S	Q	I	D
	RTE2_ZM	Maize	T	F	S	S	-	-	-	-	G	Q	H	T	S	Q	I	D

FIG. 4.—Comparisons of critical amino acids for the APE-like EN of LINEs. (*A*) Comparisons of essential amino acids for LINE EN activity. Essential amino acid residues for EN activity (Weichenrieder et al. 2004) are compared between RTE-clade LINEs and other LINEs. Among highly conserved residues, residue 229 (highlighted in black) is substituted only in plant RTEs. (*B*) Amino acid sequences of the EN beta hairpin loop, which probes the DNA minor groove. Amino acid substitutions proposed to either alter cleavage pattern (R1Bm) or decrease nicking activity (TRAS1 and L1) are shown in red. Plant RTEs are two amino acids shorter compared with other LINEs.

decrease endonucleolytic activity, although it is located adjacent to the active center of the EN. Figure 4B shows the amino acid sequences of the betaB6–betaB5 hairpin loop region of EN from animal and plant LINEs. Amino acid substitutions at positions shown in red either alters the cleavage pattern such as at R1Bm (Maita et al. 2007) or decreases nicking activity as demonstrated in TRAS1 (Maita et al. 2004) and L1 (Repanas et al. 2007). For the L1-EN, it is suggested that the conformational flexibility of the beta-hairpin loop probing the DNA minor groove may be much more important than its sequence (Repanas et al. 2007). The beta-hairpin loop of plant RTEs are two amino acids (residues 196–197) shorter than that of other LINEs (fig. 4B). Figure 5 shows the predicted three-dimensional (3D) structure of EN from soybean RTE (RTE-1_GM). Consistently, the beta-hairpin loop of soybean RTE (fig. 5 right, shown in cyan) is smaller than that found in L1 (fig. 5 left, shown in light brown). This region is predicted to overhang the minor groove of the DNA when the EN is in contact. Therefore, it is plausible that a change in the length of the beta-hairpin loop in conjunction with

the D229Q substitution could impact the specificity of plant RTEs to cleave DNA.

Identical Insertion Signature from Plant Retroposons Found in Several Animal RTE-Related SINEs

Different kinds of SINE families share 3′-terminal sequences with various RTE-clade LINEs in the genome of vertebrates (supplementary fig. S3, Supplementary Material online). Our analyses of animal SINEs with RTE-related 3′-tails revealed that the identical TSD pattern found in plants, which starts with T approximately ten nucleotides downstream of a stretch of Ts, was also found in animal SINEs from lizard and mammals (fig. 6A and B and supplementary table S2, Supplementary Material online). Analysis of green anole and elephant demonstrably showed an excess of T at P1, with a stretch of ~3 Ts at approximately P−10. Intriguingly, a horse SINE showed an excess of adenine (A) at P1 (T at P−1) with a stretch of ~3 Ts at approximately P−10 (fig. 6A and B). In contrast, RTE-clade LINEs sharing 3′-end sequences with animal SINEs start with A (P1) in many cases (fig. 6A and table 2).
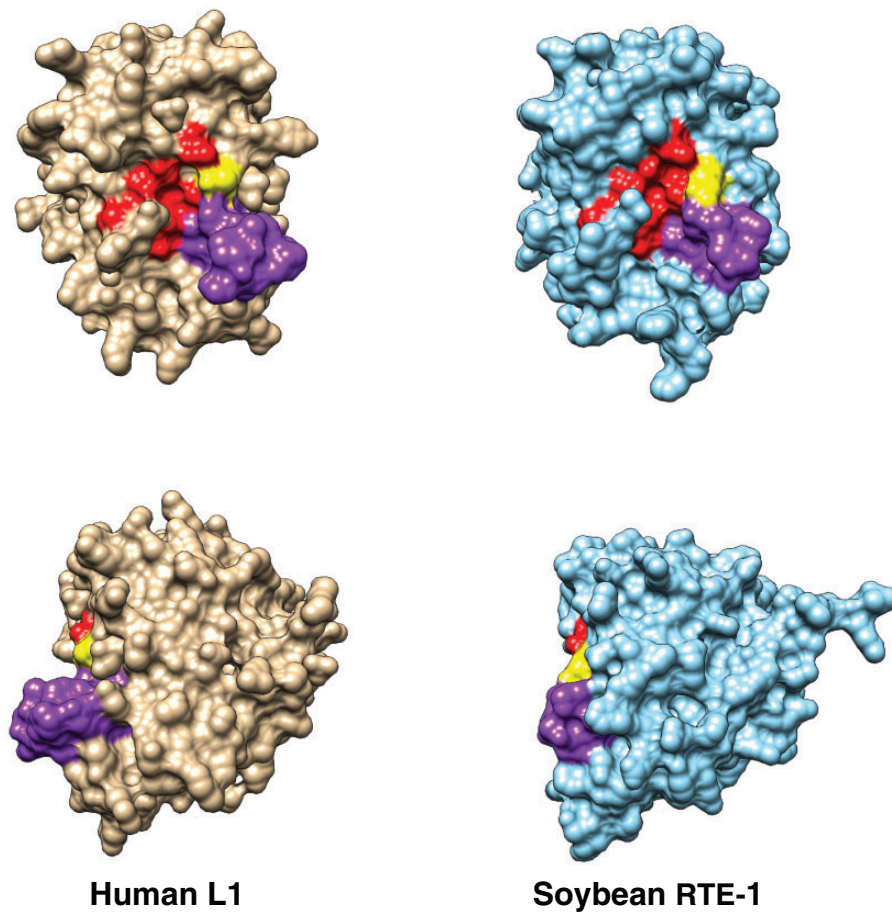
Human L1   **Soybean RTE-1**

Fɪɢ. 5.—Comparison of the 3D structure of EN domains from soybean RTE and human L1. Space-filling representation of a 3D model of soybean RTE-EN constructed using human L1-EN as template. The beta-hairpin loop of soybean RTE (cyan; right) and L1 (light brown; left) is represented in purple. The catalytic core and D229Q substitution are denoted in red and yellow, respectively. The lower images show left side views of the upper images. For reference, the DNA cleavage strand would be positioned vertically with the 5′-end at the top and the 3′-end at the bottom. Ribbon representation is available in supplementary fig. S5, Supplementary Material online.

For example, an RTE-clade LINE of green anole had an excess of A at P1 with a slight excess of T at approximately P−10.

The TSD lengths of given LINEs fall within clade-specific ranges regardless of their hosts (Ichiyanagi and Okada 2008). The majority of the TSDs for mammals and zebrafish L1-clade LINEs were 7–18 bp in length with 13–15 bp being the most abundant, whereas the majority of RTE-clade LINEs were 7–15 bp with 10–12 bp being the most abundant (Ichiyanagi and Okada 2008). We discovered that the majority of the TSDs for animal retroposons analyzed in this study were not >13 bp in length for both RTEs and SINEs (supplementary fig. S4, Supplementary Material online), and this finding further supports the possibility that in combination with common 3′-end sequences (supplementary fig. S3, Supplementary Material online), these SINEs are dependent on the RTE-clade LINEs for their retroposition. The TSD pattern for animal retroposons (fig. 6A) indicates that RTE-clade LINEs and the related SINEs show distinct TSD patterns in some cases.

Global Correlation of 3′-Microsatellite-like Sequences and TSD Profile in Plant and Animal Retroposons

The 3′-end sequences of LINEs and SINEs often terminate in microsatellite-like sequences, such as (GTT)n, (CAA)n, (AT)n, and (A)n. During the course of our TSD analysis, we observed an inconsistent tendency between plants and animals as well as RTEs and SINEs. Our analysis of the relationship between microsatellite-like sequences at the 3′-end and the first nucleotide of the TSD revealed several interesting correlations (table 2 and supplementary table S3, Supplementary Material online).

Plant RTE-clade LINEs end in (GTT)n, and the first nucleotide of their TSD is often T. Au-like SINEs, which share a specific nucleotide sequence of the 3′-terminal region with plant RTE-clade LINEs, end in a stretch of Ts and the first nucleotide of the TSD is definitively T. Animal RTE-clade LINEs often end in a microsatellite-like sequence with a repeated A such as (CAA)n and the first nucleotide of their TSD
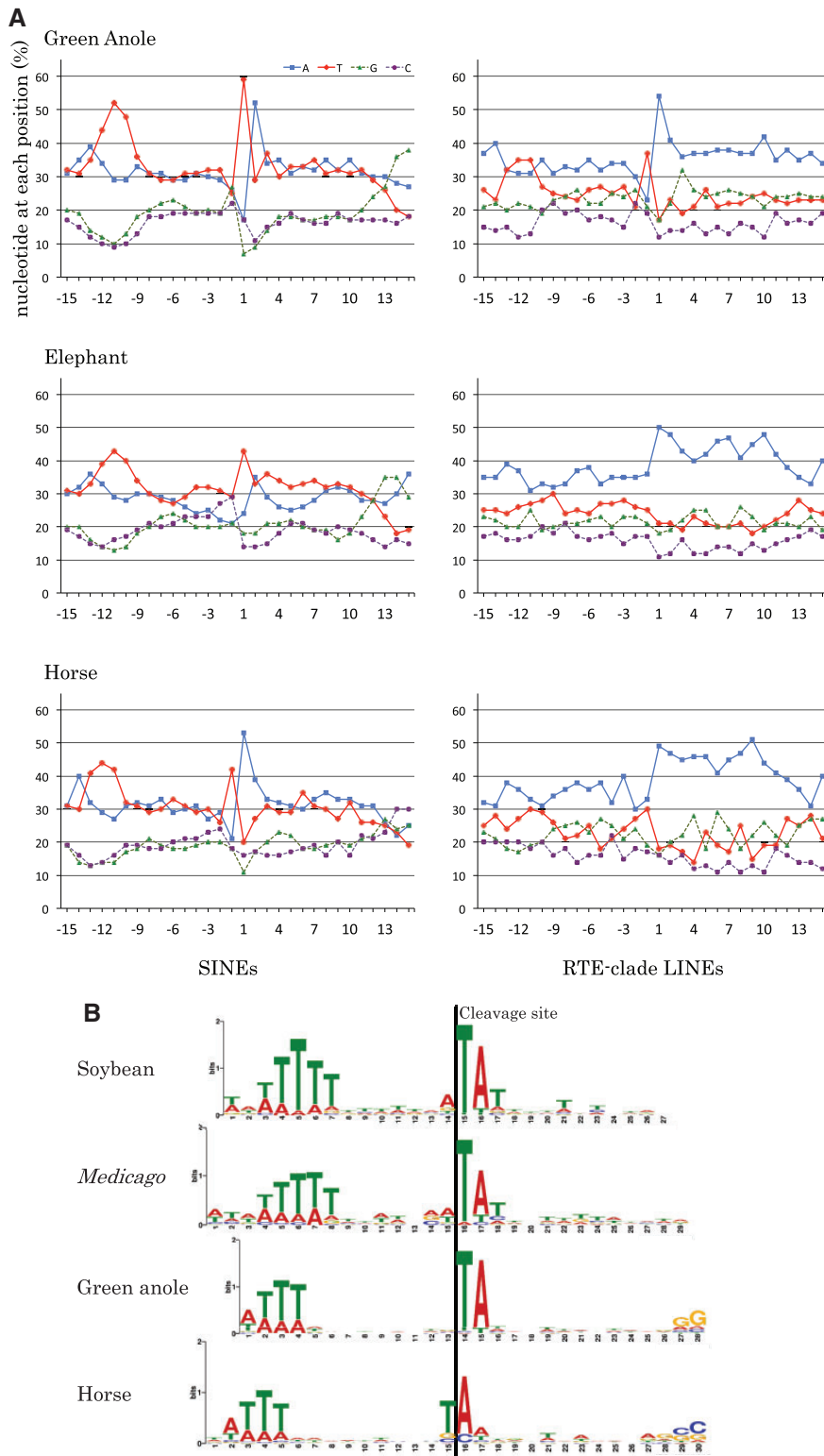
FIG. 6.—Nucleotide composition surrounding the first nucleotide of 5′ TSDs from animal retroposons and comparisons of the discovered SINE motifs between animals and plants. (A) Thirty nucleotide positions are shown with the first nucleotide of the 5′ TSD at the center (position 1: P1). Animal SINEs with an RTE-related 3′-tail (left) and RTE-clade LINEs sharing a 3′-end sequence with animal SINEs (right) from green anole (top: $n = 33{,}597$; 7,122, respectively), elephant (middle: $n = 13{,}097$; 124,680), and horse (bottom: $n = 1{,}613$; 340). The identical TSD pattern in plants, where P1 is frequently T and a prominent excess of Ts are located at approximately $P-10$, is also found in lizard and elephant SINEs. Note that RTE-clade LINEs start with adenine. Nucleotide

Table 2

Correlation of 3′-Microsatellite-Like Sequences and the First Nucleotide of TSDs[a]

	Name	Species	3′ Repeat	TSD
RTE	RTE-1_GM	Soybean	(GTT)n	T
	RTE1_MT	Medicago	(GTT)n	T
	RTE-1_Mad	Apple	(GTT)n	(A)
	RTE-1B_Mad	Apple	(GTT)n	(A)
	RTE-1_STu	Potato	(GTT)n	T
	TAe_RTE_cons	Common wheat	(GTT)n	(T/G)
	RTE-1_SBi	Sorghum	(GTT)n	T
	RTE1_ZM	Maize	(GATGTT)n	(G)
	RTE2_ZM	Maize	(GTT)n	(G)
	RTE-1_EC	Horse	(CAA)n	A
	BovB	Cow	(CTGAA)n	A
	RTE1_LA	Elephant	(CAA)n	A
	RTE1_Pca	Hyrax	(CAA)n	A
	Plat_RTE1	Platypus	(TA)n	A
	RTE_BOV_B_AC_1	Green anole	(CGA)n	A
	RTE-1_AC_1	Green anole	(GTAA)n	A
	RTE-1_OL	Medaka	(ATGG)n	(G)
	RTE-3_OL	Medaka	(TAG)n	(A/T)
SINE	GmAu1	Soybean	TTTTT	T
	MT_AUlikeSINE_cons	Medicago	TTT	T
	SINE-5_Mad	Apple	TTT	T
	SINE2-2_STu	Potato	TTTTT	T
	BDi_consensus_24	Purple false brome	T-rich	T
	SINE2-1_TAe	Common wheat	TTT	T
	RST_ZmSINE1	Maize	TTT	T
	SINE2-1_ZM	Maize	TTT	T
	SINE2-1_EC	Horse	(CAA)n	A
	BOVTA	Cow	(CA)n	(A)
	AFROSINE-2_LA	Elephant	(CAA)n	A
	AFROSINE2	Elephant	(CAA)n	(T/A)
	AFROSINE	Elephant	(GGTTT)n	T
	AFROSINE3	Elephant	(GGTTTT)n	T
	AFROSINE-1_LA	Elephant	(GGTTTT)n	(T/A)
	AFROSINE1B	Elephant	T-rich	T
	Sauria SINE	Green anole	(ACCTTT)n	T

[a]Microsatellite-like sequence at 3′-ends of SINEs and LINEs consist of a stretch of T or A plus other nucleotides. The first nucleotide of TSDs and the repeated nucleotide within the microsatellite-like sequence are consistent in many cases. In the cases where the first nucleotide of TSDs is not obvious, the nucleotides are in parentheses.

is frequently A. Animal SINEs, which share a specific nucleotide sequence of the 3′-terminal region with animal RTE-clade LINEs, were two types: one that ends in (CAA)n and has A as the first nucleotide of its TSD, and the other that ends in T-rich repeats and has T as the first nucleotide of its TSD. Interestingly, these two types of SINEs coexist in the elephant

genome (table 2 and supplementary table S3, Supplementary Material online; Gilbert et al. 2008; Bao et al. 2015). These results demonstrate that microsatellite-like terminal sequences were critically involved in determining the insertion sites of RTE-related retroposons (see Discussion).

Discussion

Genomic Integration Machinery of RTE-Related Retroposons

In this study, we found a remarkable consistency of the TSDs for plant Au-like SINEs to start with a T approximately ten nucleotides downstream of a stretch of Ts. The same TSD pattern was also found in RTE-clade LINEs, which share 3′-end sequences with Au-like SINEs, in the genome of leguminous plants. Further, animal SINEs from lizard and mammals with the RTE-related 3′-tail have the same TSD pattern, which was originally discovered in plants. Such a split signature for insertion has never been previously reported for L1-clade LINEs. Moreover, a significant correlation was observed between the first nucleotide of TSDs and the microsatellite-like sequence at the 3′-ends of SINEs and LINEs.

To explain these results comprehensively, we propose the following model (fig. 7). At the beginning of reverse transcription, the RTE protein binds to the DNA region containing a stretch of Ts upstream of the cleavage site, and cuts a phosphodiester bond at the site approximately one helical pitch downstream of the stretch of Ts. Microsatellite-like sequences such as (GGUUUU)n in the 3′-end of the template RNA for reverse transcription may influence selection of the cleavage site of the RTE EN on the first DNA cleavage strand (e.g., A on the complementary strand of T). Regarding SINEs, for non-autonomous retroposons from animal genomes, green anole and elephant SINEs tend to be cleaved at T, whereas horse and some elephant SINEs tend to be cleaved at A (fig. 6A and B and table 2 and supplementary table S3, Supplementary Material online). The observation that these elephant SINEs are largely identical with the exception of microsatellite-like sequences like (GGTTTT)n or (CAA)n suggests that the RTE-clade LINE in the elephant genome generated distinct TSD patterns depending on the different microsatellite-like sequences (Gilbert et al. 2008). Microsatellite-like sequence at the 3′-ends of animal SINEs and LINEs consist of a stretch of Ts or As plus other nucleotides. The concordance of the first nucleotide of TSDs and the repeated nucleotide within the microsatellite-like sequence indicates that the repeated nucleotide at the 3′-ends of template RNA increases the opportunity of the RTE protein to cleave the DNA strand

Fig. 6.—Continued

compositions at the respective positions are graphically represented: T (red), A (blue), G (green), and C (purple). (B) Comparisons of the discovered SINE motifs between animals and plants. MEME was applied to the animal and plant data sets (supplementary table S2, Supplementary Material online). Plant Au-like SINEs (soybean and Medicago) and animal RTE-related SINEs (green anole and horse) are shown.
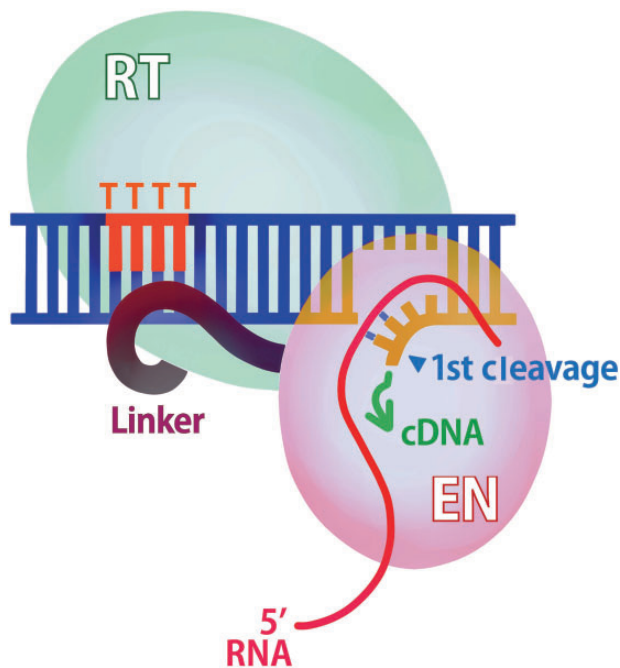
Fig. 7.—Model of the genomic integration machinery of RTE-related retroposons. The RTE protein binds to a DNA region containing a stretch of Ts upstream of the cleavage site, and cuts a phosphodiester bond approximately one helical pitch downstream of the stretch of Ts. Microsatellite-like sequences in the 3′-end of the template RNA for reverse transcription influence cleavage site selection by the RTE EN and/or facilitate the initiation of reverse transcription through base-pairing.

complementary to the repeated nucleotide (Zingler et al. 2005; Jinek et al. 2012). Alternatively, the microsatellite-like sequences could facilitate the initiation of reverse transcription through base-pairing. The 3′-terminal sequence of mammalian L1s (several bp in length) and that of the CR1, L2, and RTE clades of LINEs (one to several bp) overlaps with the 5′-end of the target sequence (Ostertag and Kazazian 2001; Ichiyanagi and Okada 2008). The overlaps between the LINE and target sequences at the 3′ junctions of retrotransposed copies are proposed to be generated by retrotransposition reactions in which the LINE RNA becomes base paired with the EN-cleaved strand of the target duplex DNA to facilitate the initiation of reverse transcription (Ostertag and Kazazian 2001; Ichiyanagi et al. 2007). Base pairing between the target DNA and the 3′-end of the mRNA may either be required for or at least facilitate the initiation of TPRT for I factor, R1Bm, and R2Ol (Chaboissier et al. 2000; Anzai et al. 2005; Fujiwara 2015). However, these interactions are not required for TPRT for some LINEs such as R2Bm (Luan and Eickbush 1995). Global correlation between the first nucleotide of TSDs and the microsatellite-like sequence at the 3′-ends of RTE-clade LINEs observed in this study is consistent with these previous observations, although animal 3′-microhomology was limited to one or two bases. Further, these two possible roles of microsatellite-like sequences may not be mutually exclusive.

Molecular Adaptation after Horizontal Transfer

This study also provides the first evidence for cross-kingdom (i.e., plant-animal) commonality of a novel insertion signature of SINEs and LINEs. Since all LINE families are evolutionarily long hitchhikers in the eukaryotic genome with ~30 clades of LINEs divided in early eukaryotes (Malik et al. 1999), they may share the same machinery from the common ancestor of plants and animals. An alternative possibility is that our observed plant-animal commonality resulted from HT events of RTE-clade LINEs between ancient plants and animals through plant-animal interactions such as between flowering plants and pollinators (e.g., insects and birds). In support, a strong similarity of some fish LINEs to plant RTE-clade LINEs have been reported (Župunski et al. 2001; Tay et al. 2010). A recent study showed unexpectedly frequent HT of RTE-clade LINEs in which HT of the Bov-B LINE was significantly more widespread than believed, and at least nine HT events were required to explain the observed topology (Walsh et al. 2013). Similarly, the genomes of the nematodes and seven tropical bird lineages exclusively shared an AviRTE LINE resulting from HT (Suh et al. 2016). The cross-kingdom commonality of the novel insertion signature found in this study could be a footprint of such a complex trajectory of genetic materials between species.

Among the various LINE clades, why the RTE-clade LINEs frequently undergo HT is not known. Our study revealed that animal RTE-clade LINEs may switch their integration site depending on their 3′ microsatellite-like sequences. Because the microsatellite contents of eukaryotic genomes are taxon-specific (Tay et al. 2010) such a simple and flexible integration mechanism of RTE-clade LINEs may have contributed to the successful expansion of RTEs and the associated SINEs in frontier genomes after HT. If RTE-clade LINEs could capture a novel microsatellite-like sequence in their 3′-end, the novel repeats may have extended the opportunity of RTEs to integrate their copies into frontier genomes, an integration that corresponds to the microsatellite environment in the genome. Further investigation is required for a better understanding of the detailed mechanism that underlies molecular adaptation after HT and the precise history of cross-kingdom HT.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Literature Cited

Ahl V, Keller H, Schmidt S, Weichenrieder O. 2015. Retrotransposition and crystal structure of an Alu RNP in the ribosome-stalling conformation. Mol Cell 60(5):715–727.

Aken BL, et al. 2017. Ensembl 2017. Nucleic Acids Res. 45(D1):D635–D642.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Anzai T, Osanai M, Hamada M, Fujiwara H. 2005. Functional roles of 3′-terminal structures of template RNA during in vivo retrotransposition of non-LTR retrotransposon, R1Bm. Nucleic Acids Res. 33(6):1993–2002.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 2:28–36.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. Nucleic Acids Res. 43(W1):W39–W49.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:11.

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat Rev Genet. 3(5):370–379.

Baucom RS, et al. 2009. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet. 5(11):e1000732.

Beck CR, Garcia-Perez JL, Badge RM, Moran JV. 2011. LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet. 12:187–215.

Beernink PT, et al. 2001. Two divalent metal ions in the active site of a new crystal form of human apurinic/apyrimidinic endonuclease, Ape1: implications for the catalytic mechanism. J Mol Biol. 307(4):1023–1034.

Bennetzen JL, Wang H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol. 65:505–530.

Bock R. 2010. The give-and-take of DNA: horizontal gene transfer in plants. Trends Plant Sci. 15(1):11–22.

Bolser DM, Staines DM, Perry E, Kersey PJ. 2017. Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data. Methods Mol Biol. 1533:1–31.

Bringaud F, et al. 2006. The Trypanosoma cruzi L1Tc and NARTc non-LTR retrotransposons show relative site specificity for insertion. Mol Biol Evol. 23(2):411–420.

Brosius J. 1991. Retroposons—seeds of evolution. Science 251(4995):753.

Chaboissier MC, Finnegan D, Bucheton A. 2000. Retrotransposition of the I factor, a non-long terminal repeat retrotransposon of Drosophila, generates tandem repeats at the 3′ end. Nucleic Acids Res. 28(13):2467–2472.

Cognat V, et al. 2008. On the evolution and expression of Chlamydomonas reinhardtii nucleus-encoded transfer RNA genes. Genetics 179(1):113–123.

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

Cost GJ, Boeke JD. 1998. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry 37(51):18081–18093.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14(6):1188–1190.

Deragon JM, et al. 1994. An analysis of retroposition in plants based on a family of SINEs from Brassica napus. J Mol Evol. 39(4):378–386.

Deragon JM, Zhang X. 2006. Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers. Syst Biol. 55(6):949–956.

Doucet AJ, Wilusz JE, Miyoshi T, Liu Y, Moran JV. 2015. A 3′ poly(A) tract is required for LINE-1 retrotransposition. Mol Cell 60(5):728–741.

Eickbush DG, Eickbush TH. 2012. R2 and R2/R1 hybrid non-autonomous retrotransposons derived by internal deletions of full-length elements. Mob DNA 3(1):10.

Eickbush TH, Eickbush DG. 2015. Integration, regulation, and long-term stability of R2 retrotransposons. Microbiol Spectr. 3(2):MDNA3-0011-2014.

Fawcett JA, Innan H. 2016. High similarity between distantly related species of a plant SINE family is consistent with a scenario of vertical transmission without horizontal transfers. Mol Biol Evol. 33(10):2593–2604.

Fawcett JA, Kawahara T, Watanabe H, Yasui Y. 2006. A SINE family widely distributed in the plant kingdom and its evolutionary history. Plant Mol Biol. 61(3):505–514.

Feng Q, Moran JV, Kazazian HH, Boeke JD. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87(5):905–916.

Fiser A, Šali A. 2003. MODELLER: generation and refinement of homology-based protein structure models. Methods Enzymol. 374:461–491.

Fuentes I, Stegemann S, Golczyk H, Karcher D, Bock R. 2014. Horizontal genome transfer as an asexual path to the formation of new species. Nature 511(7508):232–235.

Fujiwara H. 2015. Site-specific non-LTR retrotransposons. Microbiol Spectr. 3(2): MDNA3-0001-2014.

Gadzalski M, Sakowicz T. 2011. Novel SINEs families in Medicago truncatula and Lotus japonicus: bioinformatic analysis. Gene 480(1-2):21–27.

Gilbert C, Pace JK II, Waters PD. 2008. Target site analysis of RTE1_LA and its AfroSINE partner in the elephant genome. Gene 425(1-2):1–8.

Gogolevsky KP, Vassetzky NS, Kramerov DA. 2008. Bov-B-mobilized SINEs in vertebrate genomes. Gene 407(1-2):75–85.

Hayashi Y, Kajikawa M, Matsumoto T, Okada N. 2014. Mechanism by which a LINE protein recognizes its 3′ tail RNA. Nucleic Acids Res. 42(16):10605–10617.

Ichiyanagi K, Nakajima R, Kajikawa M, Okada N. 2007. Novel retrotransposon analysis reveals multiple mobility pathways dictated by hosts. Genome Res. 17(1):33–41.

Ichiyanagi K, Okada N. 2008. Mobility pathways for vertebrate L1, L2, CR1, and RTE clade retrotransposons. Mol Biol Evol. 25(6):1148–1157.

Ivancevic AM, Walsh AM, Kortschak RD, Adelson DL. 2013. Jumping the fine LINE between species: horizontal transfer of transposable elements in animals catalyses genome evolution. Bioessays 35(12):1071–1082.

Jinek M, et al. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816–821.

Jung S, et al. 2014. The Genome Database for Rosaceae (GDR): year 10 update. Nucleic Acids Res. 42(Database issue):D1237–D1244.

Jurka J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci U S A. 94(5):1872–1877.

Jurka J, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 110(1-4):462–467.

Kajikawa M, Okada N. 2002. LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell 111(3):433–444.

Kapitonov VV, Jurka J. 2003. A novel class of SINE elements derived from 5S rRNA. Mol Biol Evol. 20(5):694–702.

Kazazian HH Jr. 2004. Mobile elements: drivers of genome evolution. Science 303(5664):1626–1632.

Kidwell MG. 1993. Lateral transfer in natural populations of eukaryotes. Annu Rev Genet. 27:235–256.

Kordiš D, Gubenšek F. 1998. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. Proc Natl Acad Sci U S A. 95(18):10704–10709.

Lenoir A, et al. 2001. The evolutionary origin and genomic organization of SINEs in *Arabidopsis thaliana*. Mol Biol Evol. 18(12):2315–2322.

Luan DD, Eickbush TH. 1995. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol Cell Biol. 15(7):3882–3891.

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 retrotransposition. Cell 72(4):595–605.

Maita N, Anzai T, Aoyagi H, Mizuno H, Fujiwara H. 2004. Crystal structure of the endonuclease domain encoded by the telomere-specific long interspersed nuclear element, TRAS1. J Biol Chem. 279(39):41067–41076.

Maita N, Aoyagi H, Osanai M, Shirakawa M, Fujiwara H. 2007. Characterization of the sequence specificity of the R1Bm endonuclease domain by structural and biochemical studies. Nucleic Acids Res. 35(12):3918–3927.

Malik HS, Eickbush TH. 1998. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. Mol Biol Evol. 15(9):1123–1134.

Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol. 16(6):793–805.

Mochizuki K, Umeda M, Ohtsubo H, Ohtsubo E. 1992. Characterization of a plant SINE, p-SINE*1*, in rice genomes. Jpn J Genet. 67(2):155–166.

Monot C, et al. 2013. The specificity and flexibility of L1 reverse transcription priming at imperfect T-tracts. PLoS Genet. 9(5):e1003499.

Morrish TA, et al. 2007. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. Nature 446(7132):208–212.

Morrish TA, et al. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet. 31(2):159–165.

Nichuguti N, Hayase M, Fujiwara H. 2016. Both the exact target site sequence and a long poly(A) tail are required for precise insertion of the 18S ribosomal DNA-specific non-long terminal repeat retrotransposon R7Ag. Mol Cell Biol. 36(10):1494–1508.

Nikaido M, Nishihara H, Hukumoto Y, Okada N. 2003. Ancient SINEs from African endemic mammals. Mol Biol Evol. 20(4):522–527.

Noll A, Raabe CA, Churakov G, Brosius J, Schmitz J. 2015. Ancient traces of tailless retropseudogenes in therian genomes. Genome Biol Evol. 7(3):889–900.

Ohshima K. 2012. Parallel relaxation of stringent RNA recognition in plant and mammalian L1 retrotransposons. Mol Biol Evol. 29(11):3255–3259.

Ohshima K. 2013. RNA-mediated gene duplication and retroposons: retrogenes, LINEs, SINEs, and sequence specificity. Int J Evol Biol. 2013:424726.

Ohshima K, Hamada M, Terai Y, Okada N. 1996. The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. Mol Cell Biol. 16(7):3756–3764.

Okada N. 1991. SINEs: short interspersed repeated elements of the eukaryotic genome. Trends Ecol Evol. 6(11):358–361.

Okada N, Hamada M, Ogiwara I, Ohshima K. 1997. SINEs and LINEs share common 3′ sequences: a review. Gene 205(1-2):229–243.

Ostertag EM, Kazazian HH Jr. 2001. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. Genome Res. 11(12):2059–2065.

Peccoud J, Loiseau V, Cordaux R, Gilbert C. 2017. Massive horizontal transfer of transposable elements in insects. Proc Natl Acad Sci U S A. 114(18):4721–4726.

Pettersen EF, et al. 2004. UCSF Chimera: a visualization system for exploratory research and analysis. J Comput Chem. 25(13):1605–1612.

Piskurek O, Nishihara H, Okada N. 2009. The evolution of two partner LINE/SINE families and a full-length chromodomain-containing Ty3/

Gypsy LTR element in the first reptilian genome of *Anolis carolinensis*. Gene 441(1-2):111–118.

Repanas K, et al. 2007. Determinants for DNA target structure selectivity of the human LINE-1 retrotransposon endonuclease. Nucleic Acids Res 35(14):4914–4926.

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(9):537–546.

Schwichtenberg K, et al. 2016. Diversification, evolution and methylation of short interspersed nuclear element families in sugar beet and related Amaranthaceae species. Plant J. 85:229–244.

Seibt KM, Wenke T, Muders K, Truberg B, Schmidt T. 2016. Short interspersed nuclear elements (SINEs) are abundant in Solanaceae and have a family-specific impact on gene structure and genome organization. Plant J. 86(3):268–285.

Shu Y, et al. 2011. Identification and characterization of a new member of the SINE Au retroposon family (GmAu1) in the soybean, *Glycine max* (L.) Merr., genome and its potential application. Plant Cell Rep. 30(12):2207–2213.

Suh A, et al. 2016. Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. Nat Commun. 7:11396.

Szak ST, et al. 2002. Molecular archeology of L1 insertions in the human genome. Genome Biol. 3(10):research0052.

Tay WT, Behere GT, Batterham P, Heckel DG. 2010. Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes. BMC Evol Biol. 10:144.

Tsuchimoto S, Hirao Y, Ohtsubo E, Ohtsubo H. 2008. New SINE families from rice, *OsSN*, with poly(A) at the 3′ ends. Genes Genet Syst. 83(3):227–236.

Vassetzky NS, Kramerov DA. 2013. SINEBase: a database and tool for SINE analysis. Nucleic Acids Res. 41(Database issue):D83–D89.

Wallau GL, Ortiz MF, Loreto EL. 2012. Horizontal transposon transfer in eukarya: detection, bias, and perspectives. Genome Biol Evol. 4(8):801–811.

Walsh AM, Kortschak RD, Gardner MG, Bertozzi T, Adelson DL. 2013. Widespread horizontal transfer of retrotransposons. Proc Natl Acad Sci U S A. 110(3):1012–1016.

Weichenrieder O, Repanas K, Perrakis A. 2004. Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon. Structure 12(6):975–986.

Weiner AM, Deininger PL, Efstratiadis A. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem. 55:631–661.

Wenke T, et al. 2011. Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. Plant Cell 23(9):3117–3128.

Xu JH, Osawa I, Tsuchimoto S, Ohtsubo E, Ohtsubo H. 2005. Two new SINE elements, *p-SINE2* and *p-SINE3*, from rice. Genes Genet Syst. 80(3):161–171.

Yasui Y, Nasuda S, Matsuoka Y, Kawahara T. 2001. The Au family, a novel short interspersed element (SINE) from *Aegilops umbellulata*. Theor Appl Genet. 102(4):463–470.

Yoshioka Y, et al. 1993. Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc Natl Acad Sci U S A. 90(14):6562–6566.

Zingler N, Weichenrieder O, Schumann GG. 2005. APE-type non-LTR retrotransposons: determinants involved in target site recognition. Cytogenet Genome Res. 110(1–4):250–268.

Župunski V, Gubenšek F, Kordiš D. 2001. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. Mol Biol Evol. 18(10):1849–1863.

Associate editor: Dan Graur