Structural features and mechanism of translocation of non-LTR retrotransposons in Candida albicans

Jingchen Jiang1,†, Liuya Zhao2,†, Lan Yan2, Lulu Zhang2, Yingying Cao2, Yan Wang2, Yuanying Jiang2, Tianhua Yan1,* and Yongbing Cao2,*

1Department of Pharmacology; School of Pharmacy; China Pharmaceutical University; Nanjing, PR China; 2R & D Center of New Drug; School of Pharmacy; Second Military Medical University; Shanghai, PR China

†These authors contributed equally to this work.

Keywords: Candida albicans, non-LTR retrotransposons, LINE, Zorro, target-primed reverse transcription (TPRT)

A number of abundant mobile genetic elements called retrotransposons reverse transcribe RNA to generate DNA for insertion into eukaryotic genomes. Non-long-terminal repeat (non-LTR) retrotransposons represent a major class of retrotransposons, and transposons that move by target-primed reverse transcription lack LTRs characteristic of retroviruses and retroviral-like transposons. Yeast model systems in Candida albicans and Saccharomyces cerevisiae have been developed for the study of non-LTR retrotransposons. Non-LTR retrotransposons are divided into LINEs (long interspersed nuclear elements), SINEs (short interspersed nuclear elements), and SVA (SINE, VNTR, and Alu). LINE-1 elements have been described in fungi, and several families called Zorro elements have been detected from C. albicans. They are all members of L1 clades. Through a mechanism named target-primed reverse transcription (TPRT), LINEs transpose the new copy into the target site to initiate DNA synthesis primed by the 3′ OH of the broken strand. In this article, we describe some advances in the research on structural features and origin of non-LTR retrotransposons in C. albicans, and discuss mechanisms underlying their reverse transcription and integration of the donor copy into the target site.

Introduction

Candida albicans is a major human fungal pathogen. With the spread of AIDS and the increased use of invasive surgical techniques, C. albicans infections have become more of a problem in recent years.1 C. albicans is an asexual eukaryote. However, Saccharomyces cerevisiae can also reproduce sexually.2 Several laboratories have devoted considerable efforts over recent years toward understanding the genomic organization of C. albicans and how it varies among strains. Several results to date include the construction of a SfiI restriction map of the complete genome3 and a detailed physical map of chromosome 7.4 C. albicans is an important model system for studying pathogenic fungi and interactions between these species and their hosts. Several researchers5,6 reported the existence of a large number of families of retrotransposons in C. albicans. Retrotransposons should be are transcribed into mRNA molecules and then be reverse transcribed into double stranded cDNA by their own reverse transcriptase before the potential mobility of retrotransposons can be approximately predicted by the presence of their mRNA transcript.7 Retrotransposons are a significant component of many eukaryote genomes; for instance, L1 retrotransposon comprises 15% of the human genome,8 and is known to cause mutations and promote genomic alterations.9 It is widespread in multicellular eukaryotes, and has an important effect on the structure of eukaryotic genomic and genetic evolution. Two types of transposons have been classified: transposons that encode a transposase required for transposition, and retroposons that use a retrotranscriptase encoded in their genome for retrotransposition. Transposons are found in a large variety of eukaryotes, and retrotransposons are part of different subfamilies of transposons. It is remarkable that retrotransposons are highly related to animal retroviruses with respect to gene organization and expression strategies.10,11

Retrotransposons are divided into two major categories. First, long-terminal-repeat (LTR) retrotransposons have structures and mechanisms similar to those of vertebrate retroviruses. The integrated forms of LTR retrotransposons are flanked by LTR at the end of both sides. Second, non-long-terminal repeat (non-LTR) retrotransposons that move by target-primed reverse transcription (TPRT), which emerged from the biochemical work of Luan and Eickbush12 using the R2Bm model of Bombyx mori lacking the LTR retrotransposons characteristic of retroviruses and retroviral-like transposons. Non-LTR retrotransposons are divided into LINEs (long interspersed nuclear elements), SINEs (short interspersed nuclear elements), and SVA (SINE, VNTR, and Alu). Non-LTR retrotransposons also contain a reverse transcriptase domain. Unlike LTR retrotransposons, they have no LTR retrotransposons, either direct or indirect. This review summarizes

*Correspondence to: Tianhua Yan; Email: yth0001@126.com; Yongbing Cao; Email: ybciao@vip.sina.com
Submitted: 09/30/2013; Revised: 11/14/2013; Accepted: 11/20/2013
http://dx.doi.org/10.4161/viru.27278
that influence the evolution of eukaryote genomes. Complete elements are one of the three classes of non-LTR retrotransposons. The rapid accumulation of more sequences analyzed by the retrotransposition machinery of LTR retrotransposons. The first indication is that these elements were categorized with the LINE transposition intermediate. ORF2 encodes endonuclease (endo), reverse transcriptase (RT), zinc finger domain (zf), and RNase H domains in some cases (not shown). Arrows are TSDs. A represents poly-A tail. (B) Structure of Zorro3 in C. albicans: ORF1 contains two zinc knuckle (zk) motifs called type I ORF1, while human L1s contains a type II ORF1. Zorro3 has no TSDs, with poly-A tract flanking both ends. (C) Structure of Zorro1 in C. albicans. The end of 5′UTR cannot be identified. Unlike another non-LTR retrotransposons, neither a poly-A tract nor a 3′ tandem repeat is apparent at the 3′ end of Zorro1.

**Figure 1.** Structure of non-LTR retrotransposons. (A) Structure of LINEs: LINEs family consists of two open reading frames, ORF1 and ORF2. ORF1 encodes a RNA-binding protein that associates with the LINE transposition intermediate. ORF2 encodes endonuclease (endo), reverse transcriptase (RT), zinc finger domain (zf), and RNase H domains in some cases (not shown). Arrows are TSDs. A represents poly-A tail. (B) Structure of Zorro3 in C. albicans: ORF1 contains two zinc knuckle (zk) motifs called type I ORF1, while human L1s contains a type II ORF1. Zorro3 has no TSDs, with poly-A tract flanking both ends. (C) Structure of Zorro1 in C. albicans. The end of 5′UTR cannot be identified. Unlike another non-LTR retrotransposons, neither a poly-A tract nor a 3′ tandem repeat is apparent at the 3′ end of Zorro1.

the past and recent advances in the study of non-retrotransposon elements in *C. albicans*. Further delineation and comparison of non-LTR retrotransposons in *C. albicans* may provide interesting insights into more general aspects of the genome structure, function, and mechanism, though the integrated structure and mechanism remain unclear.

**LINEs Elements Found in *C. albicans***

As we described above, LINEs (long interspersed nuclear elements) are one of the three classes of non-LTR retrotransposons that influence the evolution of eukaryote genomes. Complete mechanistic details of how LINEs duplicate and retrotranspose are unclear; however, a mechanism of the reverse transcription, termed target-primed reverse transcription (TPRT), has been reported. In history, these elements which are called LINEs in generally today have been referred to by a variety of names, including poly A retrotransposons, nonviral retroposons, or simple retroposons. The first indication is that these elements were catalyzed by the retrotransposition machinery of LTR retrotransposons or retroviruses. The rapid accumulation of more sequences eventually leads to the recovery of elements from different animals and plants with ORFs that encode intact RT domains. Phylogenetic comparison of these RT sequences with that of all other RT sequences revealed that they represented a distinct class of retrotransposons. It soon became known that RT domains of several elements could encode authentic RT DNA polymerase activity. These elements are called LINEs retrotransposons today. The structure of LINEs is shown in Figure 1A. LINEs are 4–6 kbnp and bounded by an untranslated region (UTR) at both ends of the element. LINEs are characterized by 3′ poly-A tails or 3′ tandem repeats as other non-retrotransposons and transcribed from a promoter within the first few nucleotides of the element. Active LINEs frequently result in 5′ truncated LINE copies. Most LINE elements are inactivated because of inefficiency of reverse transcription that is error-prone, so that ORFs encoding the transposition machinery are likely to be disabled by mutations, and not highly processive, so that 5′ truncation of the elements often occurs during transposition. A typical LINEs family consists of two open reading frames, ORF1 and ORF2 (Fig. 1A). ORF1 encodes a RNA-binding protein that associates with the LINEs transposition intermediate and nucleic acid chaperone activity, both of which are important for LINEs activity. ORF2 encodes endonuclease, reverse transcriptase activity, zinc finger domain, and RNase H domains in some cases. Genomic LINEs, like human L1, are typically flanked by target site duplications (TSDs) as LTR-retrotransposons. ORF1 and ORF2 proteins assemble with LINEs RNA into a ribonucleoprotein (RNP) complex, which is presumably transported into the nucleus.

Multiple retrotransposons, consisting of non-LTR retrotransposons and LTR-retrotransposons, are flanked by 4–5 bp short direct repeats representing TSDs at 5′ and 3′ ends. For instance, 36% of the total *S. cerevisiae* Ty 1–4 elements were flanked by TSDs, and it is reported that Tca elements are also typically flanked by TSDs. Analyzing the sequences of all the perfect TSDs of Tca elements in *C. albicans* (Fig. 2A) and Ty elements in *S. cerevisiae* (Fig. 2B) to derive a 4–5 bp TSDs target site sequence, a strong bias for A and T: in the internal position 2 (72%), position 3 (76%), position 4 (78%), is shown in Figure 2A, a bias for A and G is found in position 1 (92%), a bias for T and C is shown in position 4 (71%). Recombination or mutation may result in the exchange of target site sequences between the elements.

Many non-retrotransposons have been found in vertebrates, insects, and fungi. Human L1 element has affected both the size and complexity of the human genome, and varietal plant non-LTR retrotransposons have been reported, for instance, Cin4 in maize and BLIN (6.3 kbp in length) from barley. So far, phylogenetic analysis of non-LTR retrotransposons based on the reverse transcriptase domains has allowed for distinguishing 21 clades. Three clades (Tad, L1, and CRE) of non-LTR retrotransposons are known in fungi. L1 clade elements were described from the genomes of *C. albicans*, a basidiomycete *Microbotryum violaceum*, and a glomeromycete *Gigaspora*. Unfortunately, *S. cerevisiae* appears to lack non-LTR retrotransposons.

The existence of non-LTR retrotransposons in *C. albicans* has been reported. Subsequently, Goodwin et al. used a series of TBLASTN (protein query vs. nucleotide database) and BLASTN search to screen non-LTR retrotransposons in assembling 5 of the Stanford *C. albicans* sequence database, and identified only...
a small number of sequences corresponding to non-LTR retrotansposons. Only three of them appear to be full-length or nearly full-length: Zorro1, Zorro2, and Zorro3 with 25–40% amino acid identity.6 Zorro elements are widespread in *C. albicans* giving low copy numbers (data not shown by the original authors).6

The structures of the Zorro elements are shown in Figure 1B and C. The structure of Zorro2 is similar to that of Zorro1, except that the ORFs have suffered several nonsense frameshift mutations and highly conserved residues can be identified. The intact Zorro1 element (Fig. 1C) contains two ORFs, like many non-LTR retrotansposons. ORF1 contains two zinc-finger motifs potentially considered as putative nucleic acid-binding domains. ORF2 encodes a reverse transcriptase (RT), and a C-terminal. Upstream of ORF1 is a 5′ untranslated region (5′UTR), and the end of 5′UTR cannot be identified. Comprising with 5′UTR, downstream of ORF2 is a 3′ untranslated region (3′UTR). The end of this 3′UTR can be tentatively identified; however, neither a poly-A tract nor a 3′ tandem repeat is apparent at the 3′ end of Zorro1. The Zorro3 element is a structurally intact element.48,49 It contains ORF1 and ORF2, the first of which encodes two zinc-finger motifs (considered as putative nucleic acid-binding domains). ORF2 of Zorro3 encodes an endonuclease (EN), a reverse transcriptase (RT), and a C-terminal. Zorro3 is bounded by 5′UTR at upstream of ORF1 and 3′UTR at downstream of ORF2. The end of 5′UTR of Zorro3 is characterized by a series of A residues, and the end of 3′UTR can be identified as a short poly-A tract (itself bordered by poly-A). Interestingly, ORF2 of Zorro3 is separated from a feature-like stop codon that contains four in-phase stop codons. But it was reported50 that the gag and pol ORFs were separated by a UGA stop codon (gag-UGA-pol junction) in the *C. albicans* retrotansposon Tca2. Forbes and Gibson et al.50 demonstrated that the LTR promoter directed Tca2 pol protein expression and suggested that there was a non-canonical mechanism underlying gag UGA bypass in Tca2. Unfortunately, whether or not Zorro3’s ORF2 directly translates stop codon bypass remains unclear.

**Figure 2.** TSDs target site sequence flanking by retrotansposons. The direction of TSDs is 5′ to 3′. (A) Four base pairs TSDs target site sequence analyzed by Tca families in *Candida albicans*. Sample capacity, *n* = 24. (B) Five base pairs TSDs target site sequence analyzed by Ty families in *Saccharomyces cerevisiae*. Sample capacity, *n* = 118.
L1-like non-LTR retrotransposons were described for all eukaryotic groups: Protista, Plantae, Fungi, and Metazoa. The neighbor-joining (NJ) phylogenetic tree based on reverse transcriptase of non-LTR retrotransposons reveals the position of the Zorro elements in L1 non-LTR retrotransposons. Figure 3 shows that the phylogenetic tree in distinct families is inside L1 clade based on RT domain. Subsequently, three Zorro elements emerge as a monophyletic group shown in Figure 4. These assignments are well supported by bootstrap re-sampling. It is remarkable that the three families of Zorro elements have been evolving independently for a very long time, and that they are probably extremely ancient components of the Candida genome.

As we described below, Zorro elements in C. albicans are intact elements consisting of two ORFs, and ORF2 encodes an endonuclease (EN), a reverse transcriptase (RT), and a C-terminal. An UTR is bounded at both ends of Zorro elements in C. albicans. However, comparing with another L1 non-LTR retrotransposons, for instance, human L1, which is a classical structure, these are series of differences between Zorro elements and another L1 non-LTR retrotransposons (Fig. 1). Unlike human L1 elements, in Figure 5. Recombination begins with nicking of the target DNA by the element-encoded EN that preferentially cleaves A/T rich sequences, with nicking occurring mainly at the TpA and flanking phosphodiesteres. The target DNA 3’ OH exposed by endonuclease cleavage then acts as a primer for the synthesis of a new line DNA strand by reverse transcriptase using the line mRNA as a template. Thus a new line DNA strand is produced at the insertion site. And then, the nuclease makes a break in the opposite strand of chromosomal DNA a few nucleotides from the first. Template RNA is removed by RNase H allowing the new 3’ OH to prime synthesis of the second DNA strand and host repair enzymes to complete integration. Finally, a second DNA strand is synthesized, and the target DNA at each end is filled in to generate the TSDs.

Translocation of LINEs Using Target-Primed Reverse Transcription

The process of how LINE elements retrotranspose is called target-primed reverse transcription (TPRT), which is a mode of duplication and transposition of non-LTR retrotransposons by spreading through reverse transcription of retrotransposon RNA primed by DNA at the target site. By extension, it is likely that this mechanism applies to numerous LINEs found in diverse lineages, like human L1. At first, a RNA binding protein with endonuclease (EN) activity encoded by ORF1, a multifunctional protein with reverse transcriptase (RT) activity encoded by ORF2, and the L1 RNA transcribed from its internal RNA polymerase II promoter located within the 5’UTR, to compose a compound called L1 ribonucleoprotein particle (RNP). RNP enters the nucleus and nicks a chromosomal target site for integration. The sequence of events in translocation is shown in Figure 5.
DNA moves to the RT active site and the newly generated 3′ OH primes reverse transcription and double-strand breaks (DSBs) generated in trans.

Goodwin et al. developed a yeast model system using the Zorro3 element from *C. albicans* for the study of non-LTR retrotransposons. This system called retrotransposition assay for Zorro3 is outlined in Figure 6, in which the ORF of the *C. albicans* URA3 gene and its promoter sequence, with the ORF disrupted by an antisense intron inserted into 3′ UTR of Zorro3 element, as the indicator gene. When Zorro3 is transcribing to give a full-length mRNA, and then the intron would be removed by splicing. Thus, retrotransposition events can be detected by the appearance of URA3+ colonies on the appropriate selective media. After retrotransposition assay, 30 independent transposed copies were amplified to reveal not only the 3′ and 5′ ends but their 3′ and 5′ flanking sequences of retrotransposed Zorro3 elements. Several findings from these sequences indicate that the target site of Zorro3 elements which is inserted very close to coding regions specifically integrated at poly-A sequences, and there seemed to be a bias toward promoter regions. In addition, Goodwin et al. suggested that the transposable events in Zorro3 of *C. albicans* are similar to TPRT in mammalian cells.

As we described above, non-LTR retrotransposons have never been found in *S. cerevisiae* that has no endogenous L1 homologs or remnants. However, Poulter et al. established a model system of *S. cerevisiae* called retrotransposition assay for scZorro3 (Zorro3 in *S. cerevisiae* named ScZorro3), which has a similar process of retrotransposition assay for Zorro3 in *C. albicans* except using mHIS3AI as an indicator gene to confirm Zorro3 retrotransposition, and found that *S. cerevisiae* unexpectedly retained the basal host machinery required for L1 retrotransposition. Through this model system called scZorro3 that recapitulates the non-LTR retrotransposition process in *S. cerevisiae*, they found several differences between Zorro3 of *C. albicans* retrotransposition and scZorro3 of *S. cerevisiae* retrotransposition. For instance, the reverse transcription complex searches for sequences with homology to the minus strand to enable the template to jump during minus strand synthesis. In Zorro3 of *C. albicans* retrotransposition, this search is largely restricted to regions around the target site. In scZorro3 of *S. cerevisiae* retrotransposition, this search space is relaxed, and template jumping occurs in other RNAs/DNAs at a higher frequency. In addition, scZorro3 can generate a circular and episomal retrotransposition products in *S. cerevisiae*. Previously, circular products derived from retroviruses or LTR-retrotransposons were observed. Han and Shao suggested that these products are likely to be formed via a variation of TPRT. For simplicity, bottom chromosome strand nick at first, and then LINE mRNA annealing, minus strand synthesis finally. Subsequently, the top strand chromosomal nick and the template jump to the top strand and then re-cleave the top and bottom strands to release the retrotransposition intermediate. These episomal products may represent an unexpected source for de novo retrotransposition. Yeast model systems of *C. albicans* and *S. cerevisiae* have been principally described, which have been developed for studying Zorro family elements and TPRT emerged by biochemical experiments with human L1 and Zorro3 retrotransposon. However, complete mechanistic details

![Figure 5](https://www.landesbioscience.com/Virulence/249/images/11558573.jpg)

**Figure 5.** The mechanism of target-primed reverse transcription (TPRT). Transposition begins with the transcription of the LINE element (red) into RNA (blue) which encodes an RNA binding protein and a multifunctional protein with endonuclease and reverse transcriptase activity. These proteins (not shown) associate with the LINE RNA, and the endonuclease nicks the DNA at the target site, which contains a poly T tract, which base-pairs with the poly A sequence in the LINE RNA. The LINE RNA is then copied by the reverse transcriptase into a DNA copy (green), which is covalently attached to the target DNA. A second DNA strand is then synthesized on the template of the DNA copy, and the target DNA at each end is filled in to generate the TSDs that flank these elements.
of how Zorro families of LINE elements retrotranspose remain unclear.

TPRT is a process spreading through reverse transcription of retrotransposon RNA primed by DNA, effectively welding the new copy into the target site as it is made. It is a complicated process that LTR retrotransposons can move from place to place in a genome by reverse transcription of an RNA transposition mediated in cells (i.e., we do not describe in details). Distinguishing features of TPRT (as compared with the process that LTR retrotransposons transpose) are the RNP, consisting of L1 RNA, proteins encoded by ORF1 and ORF2, enters the nucleus and nicks a chromosomal target site as the first step; however, no compound similar to RNP have been found in the process that LTR retrotransposons transpose. The target DNA 3′ OH acts as a primer for the synthesis of a new line DNA strand in TPRT, whereas a tRNA base-paired to a sequence near 5′ end of the genomic RNA, as a primer to anneals to binding site on retroviral RNA for the synthesis of minus strand DNA; retroviral RNA ends in direct repeats (R), and results that a linear double-stranded DNA with an LTR at each end.

Non-LTR Retrotransposons Play an Important Role in Evolutionary Dynamics of C. albicans

The evolutionary history of a particular or related species, the population structure, ecological aspects, and the mating mode could affect the diversity of non-LTR retrotransposons and copy numbers. For instance, L1 elements play an important role in the evolution of the structure and activity of the remainder of the genome by providing dispersed sites of sequence similarity at which recombination can occur, by inserting into genes altering their structure and/or regulation, and by carrying flanking sequences with them during transposition (L1-mediated sequence transduction). In addition, there are other processes that could affect the copy number and diversity of non-LTR retrotransposons in fungi: stochastic loss of non-LTR retrotransposons, burst of retrotransposition, the limitation of copy number increase by natural selection which removes deleterious insertions, horizontal transfer, passive and active inactivation of repetitive sequences, and self-regulation of transposition. Low copy numbers of non-LTR retrotransposons could cause a loss of retrotransposons-like elements as a result of genetic drift, especially when the population is small and non-LTR retrotransposons degenerate copies. It is reported that the presence of retrotransposons and their large copy numbers can cause mutations and genomic rearrangements. These discoveries indicate that non-LTR retrotransposons and the transposition play an important role in evolutionary dynamics of C. albicans.

The inactivation of repeated sequences is a very important factor, which leads to the shift in diversity and copy number of non-LTR retrotransposons. For instance, non-LTR retrotransposons represented only by degenerate copies in Drosophila could lose these elements as a result of genetic drift, especially if the population is small. In bacteria, Tn retrotransposons are likely to be principal players in the formation of tetracycline resistance by spreading drug resistance gene during genetic transfer. In addition, the relationship between resistance and virulence with reverse transposition of retrotransposons is rarely reported, but in our original research, the transposition of Zorro2 and Zorro3 in strains that are resistant to miconazole and the strains show low virulence in a systemic murine candidiasis model, have been observed (unpublished).

SINEs and SVA Elements Are Rarely Reported in C. albicans

We have summarized several past and recent advances in the study of LINEs including Zorro families in C. albicans. Unfortunately, little has been known about the distribution and properties of SINES and SVA elements in C. albicans as compared with LINEs elements. However, much has been disclosed about the biology and function of SINES and SVA elements since these elements were discovered.

SINEs are genomic sequences derived from tRNA genes or 7SL RNA, and they spread non-autonomously in the genome by TPRT mediated by LINE-encoded recombination proteins. The first described SINES were mouse B1 and B2 and human Alu. Today these elements are also found existing in other organisms, including fungi, insects, birds, and plants. SINEs are similar to LINEs in that both move via TPRT. SINE elements are much shorter (100–300 bp) than LINEs. A typical SINE consists of three parts: 5′ ends of all SINES families originating from one of the three types of short pol III transcripts: tRNAs, 5S tRNA, or 7SL RNA. The 3′ ends consist of poly A tails flanked by TSDs. The internal domain of the SINEs family
SVA elements for another group of non-autonomous retroelements in humans and non-human primates, and are present in fish, deuterostomes, and cephalopods. It is named “SVA” after its main components (SINE, VNTR, and Alu) with a stretch of sequence that shares sequence similarity with Alu sequences. The 3′ ends of full-length SVA have the human endogenous retrovirus HERV-K, including the LTR and a 3′ poly A tail, and TSDs flanking both ends of SVA elements. A (CCCTCT)n hexamer simple repeat region that is located at the 5′ end. The internal domain is composed of an Alu-like sequence, a VNTR (variable number of tandem repeats) region, and a SINE region (SINE-R) about 490 bp. It is proposed that SVA elements are non-autonomous retrotransposons that are mobilized by L1 encoded proteins in trans.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
This project was supported by the National Natural Science Foundation of China (81173100) and Shanghai Basic Research Project (11JC1415400).

References
1. Oddis FC. A Review and Bibliography. Candida and Candidosis. London: Bailliere Tindall; 1988.
2. Scherer S, Magee PT. Genetics of Candida albicans. Microbiol Rev 1990; 54:226-41; PMID:2215421.
3. Chu WS, Magee BB, Magee PT. A comprehensive survey of retrotransposons in the yeast Candida albicans. Genome Res 2000; 10:174- 91; PMID:10673276; http://dx.doi.org/10.1011/gr.10.2.174.
4. Goodwin TJ, Ormandy JE, Poulter RTM. Multiple LTR-retrotransposon families in the asexual yeast Crithidia fasciculata. Proc Natl Acad Sci U S A 1991; 88:9794-8; PMID:1719539; http://dx.doi.org/10.1073/pnas.88.21.9794.
5. Mathias SL, Scott AF, Kazazian HH Jr., Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. Science 1991; 254:1808- 10; PMID:1714378; http://dx.doi.org/10.1126/science.1722352.
6. Kulpa DA, Moran JV. Ribonucleoprotein particle with LINE-1 RNA in mouse embryonal carcinoma cells. Mol Cell Biol 1991; 11:4804-7; PMID:1715025; http://dx.doi.org/10.1016/j.gene.2008.11.030.
7. Boeke JD, Garfinkel DJ, Styles CA, Fink GR. Ty elements transpose through an RNA intermediate. Cell 1985; 40:491-500; PMID:2982495; http://dx.doi.org/10.1016/S0092-8674(00)81997-2.
8. McCallum J, Li P, Kwok SC, Hodges RS, Williams MC. Low-affinity, non-LTR retrotransposons on the human genome. Nat Genet 2006; 38:549-61; PMID:16582663; http://dx.doi.org/10.1038/ng.2005.303.
9. SVA elements were originally named SINE-R. It is named “SVA” after its main components (SINE, VNTR, and Alu) with a stretch of sequence that shares sequence similarity with Alu sequences. The 3′ ends of full-length SVA have the human endogenous retrovirus HERV-K, including the LTR and a 3′ poly A tail, and TSDs flanking both ends of SVA elements. A (CCCTCT)n hexamer simple repeat region that is located at the 5′ end. The internal domain is composed of an Alu-like sequence, a VNTR (variable number of tandem repeats) region, and a SINE region (SINE-R) about 490 bp. It is proposed that SVA elements are non-autonomous retrotransposons that are mobilized by L1 encoded proteins in trans.
53. Cost GJ, Eckbish TH, NeSoL-1, an ancient lineage of site-specific non-LTR retrotransposons from Caranorbidiidae elegans. Genetics 2000; 154:193-203; PMID:10862980

50. Forbes EM, Nieduszynska SR, Brunton FK, Gibson TJD. The DIRS1 group of retrotransposons. Mol Biol Evol 2002; 19:664-77; PMID:11961100; http://dx.doi.org/10.1093/oxfordjournals.molbev.a041215

48. Khazina E, Weichenrieder O. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of ancient mobile DNA. BMC Biol 2003; 20:1811-25; PMID:12826326; http://dx.doi.org/10.1039/molmsb17897

45. Gollotte A, L’Haridon F, Chatagnier O, Wettstein JD, Tu Z. Non-LTR retrotransposons. Res Microbiol 2007; 158:124-34; PMID:17683538; http://dx.doi.org/10.1016/j.resmic.2006.11.001

43. Biedler J, Tu Z. Non-LTR retrotransposons involved in target site recognition. Cytogenet Genome Res 2005; 110:426-40; PMID:16093695; http://dx.doi.org/10.1159/000084958

41. Arkhipova IR, Morrison HG. Three retrotransposon families in the genome of *Giardia lamblia*: two telomeric, one dead. Proc Natl Acad Sci U S A 2001; 98:14497-502; PMID:11734649; http://dx.doi.org/10.1073/pnas.231494798

40. Lander ES, Linton LM, Birren B, Nusbaum C, et al.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921; PMID:11327001; http://dx.doi.org/10.1038/35057062

39. Volff JN, Körting C, Scharr M. Multiple lineages of mouse non-LTR retrotransposon Rex with varying success in invading fish genomes. Mol Biol Evol 2000; 17:1673-84; PMID:11700955; http://dx.doi.org/10.1093/oxfordjournals.molbev.a026266

38. Malik HS, Eckbish TH, NeSoL-1, an ancient lineage of site-specific non-LTR retrotransposons from Caranorbidiidae elegans. Genetics 2000; 154:193-203; PMID:10862980

37. Zingler N, Weichenrieder O, Schumann GG. APE-1 involved in target site recognition. Cytogenet Genome Res 2005; 110:426-40; PMID:16093695; http://dx.doi.org/10.1159/000084958

36. Lander ES, Linton LM, Birren B, Nusbaum C, et al.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921; PMID:11327001; http://dx.doi.org/10.1038/35057062

35. Sver Disability identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol 1990; 10:6718-29; PMID:1710022

34. Ogiwara I, Miya M, Ohshima K, Okada N. V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. Genomics 2002; 72:313-24; PMID:11827951; http://dx.doi.org/10.1016/S0887-358X(01)00158-2

33. Flavell AJ, Ish-Horowicz D. Extrachromosomal retrotransposition in Saccharomyces cerevisiae. EMBO J 1989; 8:2869-75; PMID:29491182; http://dx.doi.org/10.1002/str.3460080901

32. Nishihara H, Smit AF, Okada N. Functional annotation of a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. Genomics 2002; 72:313-24; PMID:11827951; http://dx.doi.org/10.1016/S0887-358X(01)00158-2

31. Khazina E, Weichenrieder O, Non-LTR retrotransposon Rex involved in retrotransposition. BMC Genomics 2007; 8:263; PMID:17683538; http://dx.doi.org/10.1039/molmsb17897

28. Forbes EM, Nieduszynska SR, Brunton FK, Gibson TJD. Basic local alignment search tool. J Mol Biol 1999; 215:403-10; PMID:2237112

27. Khazina E, Weichenrieder O. Non-LTR retrotransposons in dimorphic yeast. Yarrowia lipolytica. Mol Biol Evol 2002; 19:664-77; PMID:11961100; http://dx.doi.org/10.1093/oxfordjournals.molbev.a041215

25. Hood ME. Repetitive DNA in the automictic fungus *Microbotryum violaceum*. Genetics 2002; 162:2041-51; PMID:11700955; http://dx.doi.org/10.1038/35057062

24. Goodwin TJ, Busby JN, Poulter RT. A yeast model for targeted (non-LTR) retrotransposition. BMC Genomics 2007; 8:263; PMID:17683538; http://dx.doi.org/10.1039/molmsb17897

23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403-10; PMID:2237112

22. Khazina E, Weichenrieder O. Non-LTR retrotransposons in dimorphic yeast. Yarrowia lipolytica. Mol Biol Evol 2002; 19:664-77; PMID:11961100; http://dx.doi.org/10.1093/oxfordjournals.molbev.a041215

20. Hood ME. Repetitive DNA in the automictic fungus *Microbotryum violaceum*. Genetics 2002; 162:2041-51; PMID:11700955; http://dx.doi.org/10.1038/35057062

19. Khazina E, Weichenrieder O, Non-LTR retrotransposons in dimorphic yeast. Yarrowia lipolytica. Mol Biol Evol 2002; 19:664-77; PMID:11961100; http://dx.doi.org/10.1093/oxfordjournals.molbev.a041215

18. Goodwin TJ, Busby JN, Poulter RT. A yeast model for targeted (non-LTR) retrotransposition. BMC Genomics 2007; 8:263; PMID:17683538; http://dx.doi.org/10.1039/molmsb17897

17. Khazina E, Weichenrieder O, Non-LTR retrotransposons in dimorphic yeast. Yarrowia lipolytica. Mol Biol Evol 2002; 19:664-77; PMID:11961100; http://dx.doi.org/10.1093/oxfordjournals.molbev.a041215

16. Hood ME. Repetitive DNA in the automictic fungus *Microbotryum violaceum*. Genetics 2002; 162:2041-51; PMID:11700955; http://dx.doi.org/10.1038/35057062