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

Jingchen Jiang[1,†], Liuya Zhao[2,†], Lan Yan[2], Lulu Zhang[2], Yingying Cao[2], Yan Wang[2], Yuanying Jiang[2], Tianhua Yan[1,*], and Yongbing Cao[2,*]

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

[†]These authors contributed equally to this work.

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 translocate 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 genome[3] 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 researchers[5,6] 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 Eickbush[12] 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
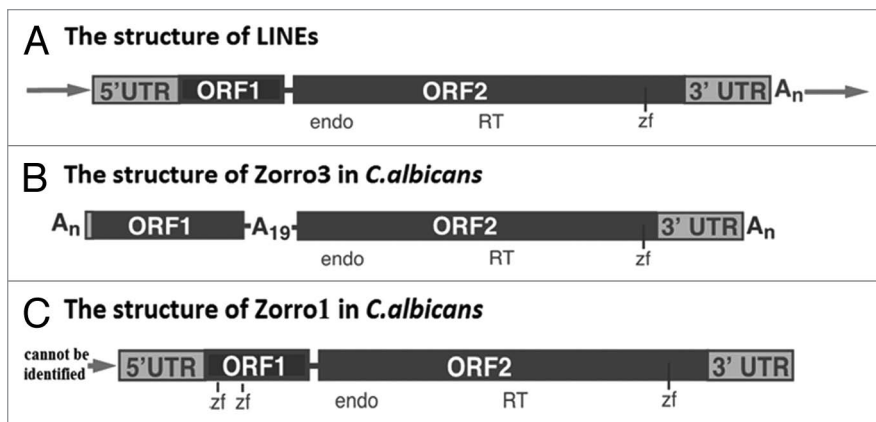
Figure 1. Structure of non-LTR retrotransposons. (**A**) Structure of LINEs: LINEs family consists of two open reading frames, ORF1 and ORF2. ORF1 encodes a RNA-binding protein that associates with the LINE transposition intermediate. ORF2 encodes endonuclease (endo), reverse transcriptase (RT), zinc 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.[13] 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.[12] 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.[14] 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.[15,16] It soon became known that RT domains of several elements could encode authentic RT DNA polymerase activity.[17-19] These elements are called LINEs retrotransposons today. The structure of LINEs is shown in **Figure 1A**. LINEs are 4–6 kbp in length and bounded by an untranslated region (UTR) at both ends of the element.[20] 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.[21] 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,[22-25] both of which are important for LINEs activity.[26,27] ORF2 encodes endonuclease,[28] reverse transcriptase activity,[29] 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,[30] which is presumably transported into the nucleus.[31,32]

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,[33] and it is reported that Tca elements are also typically flanked by TSDs.[5] Analyzing the sequences of all the perfect TSDs of Tca elements in *C. albicans*[5] (**Fig. 2A**) and Ty elements in *S. cerevisiae*[33] (**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 2B**. 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,[34] and varietal plant non-LTR retrotransposons have been reported, for instance, Cin4 in maize[35] and BLIN (6.3 kbp in length) from barley.[36] So far, phylogenetic analysis of non-LTR retrotransposons based on the reverse transcriptase domains has allowed for distinguishing 21 clades.[37-43] 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*,[44] a basidiomycete *Microbotryum violaceum*,[45] and a *glomeromycete Gigaspora*.[46] Unfortunately, *S. cerevisiae* appears to lack non-LTR retrotransposons.[33]

The existence of non-LTR retrotransposons in *C. albicans* has been reported.[4,5] Subsequently, Goodwin et al.[6] used a series of TBLASTN (protein query vs. nucleotide database) and BLASTN search[47] to screen non-LTR retrotransposons in assembling 5 of the Stanford *C. albicans* sequence database, and identified only
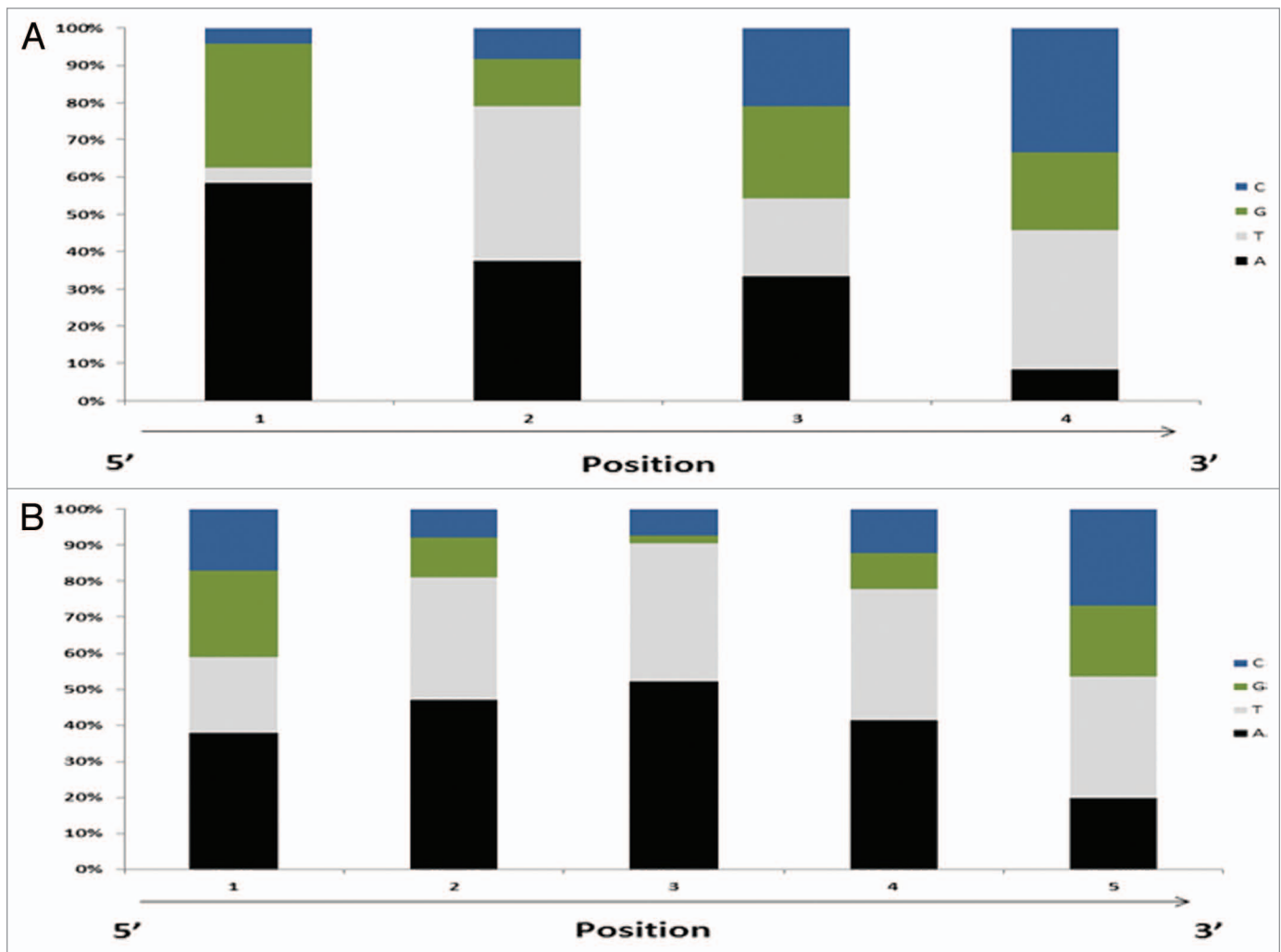
Figure 2. TSDs target site sequence flanking by retrotransposons. The direction of TSDs is 5′ to 3′. (**A**) Four base pairs TSDs target site sequence analyzed by Tca families in *Candida albicans*. Sample capacity, *n* = 24. (**B**) Five base pairs TSDs target site sequence analyzed by Ty families in *Saccharomyces cerevisiae*. Sample capacity, *n* = 118.

a small number of sequences corresponding to non-LTR retrotransposons. 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 retrotransposons. ORF1 containing two zinc-finger motifs potentially considered as putative nucleic acid-binding domains. ORF2 encodes a potential endonuclease (EN), 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 reported[50] that the gag and pol ORFs were separated by a UGA stop codon (gag-UGA-pol junction) in the *C. albicans* retrotransposon 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.
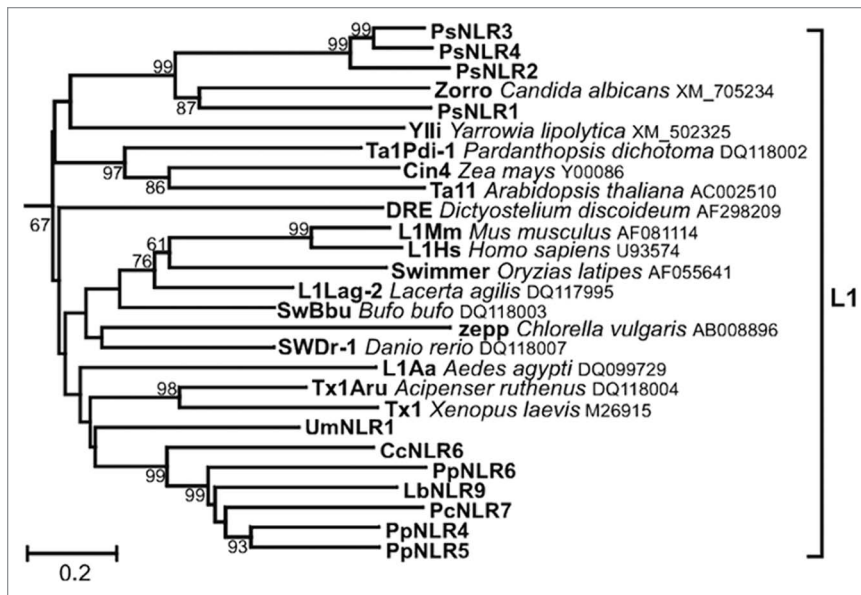
Figure 3. The neighbor-joining (NJ) phylogenetic tree based on RT amino acid sequences of L1 elements from fungi. The percentage of bootstrap support for major branches is indicated. The clade and families are shown on the right. The distance is the categories distance of the PROTDIST program of PHYLIP.[62]
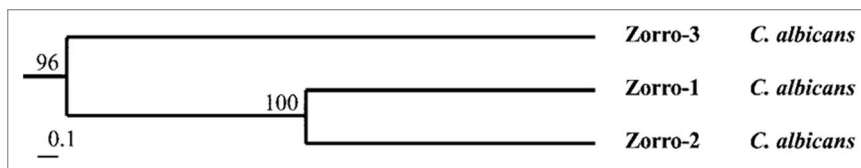
Figure 4. The neighbor-joining (NJ) phylogenetic tree based on RT sequences of Zorro elements from *C. albicans*. The percentage of bootstrap support for major branches is indicated. The clade and families are shown on the right. The distance is the categories distance of the PROTDIST program of PHYLIP.[62]

L1-like non-LTR retrotransposons were described for all eukaryotic groups: Protista, Plantae, Fungi, and Metazoa.[36,43,51,52] 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 elemnets,

neither a poly-A tract nor a 3′ tandem repeat is apparent at the 3′ end of this copy of Zorro1 and Zorro2. However, Zorro3 has a short poly-A tract at the end of 3′UTR. Another distinguishing feature between Zorro3 and human L1 elements is that Zorro elements contain two zinc-finger motifs in ORF1 instead of the conserved mammalian C-terminal domain. ORF1 contains two zinc knuckle motifs called type I ORF1, while human L1s contains a type II ORF1.[48] Another distinguishing feature is a 19-bp poly-A tract in the inter ORF region.

## 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.[53] 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,[54] to compose a compound called L1 ribonucleoprotein particle (RNP).[21,55] RNP enters the nucleus and nicks a chromosomal target site for integration. The sequence of events in translocation is shown 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 phosphodiesters. 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.[56] 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.[57-58] In addition, TPRT is mediated by the activities of both EN and RT domains; however, whether EN and RT are competitive inhibitors or non-competitive inhibitors remains unclear.[26,59] In RNP, the poly G RNA could inhibit L1 EN activity. By DNA binding or action of L1 ORF1,[22,60] the poly G RNA may be removed from the L1 EN domain, and L1 EN activates to nick the chromosome at the target site. The nicked

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.[49] 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.[61] 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*.[62] Previously, circular products derived from retroviruses or LTR-retrotransposons were observed.[63-65] Han and Shao suggested that these products are likely to be formed via a variation of TPRT.[63] 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
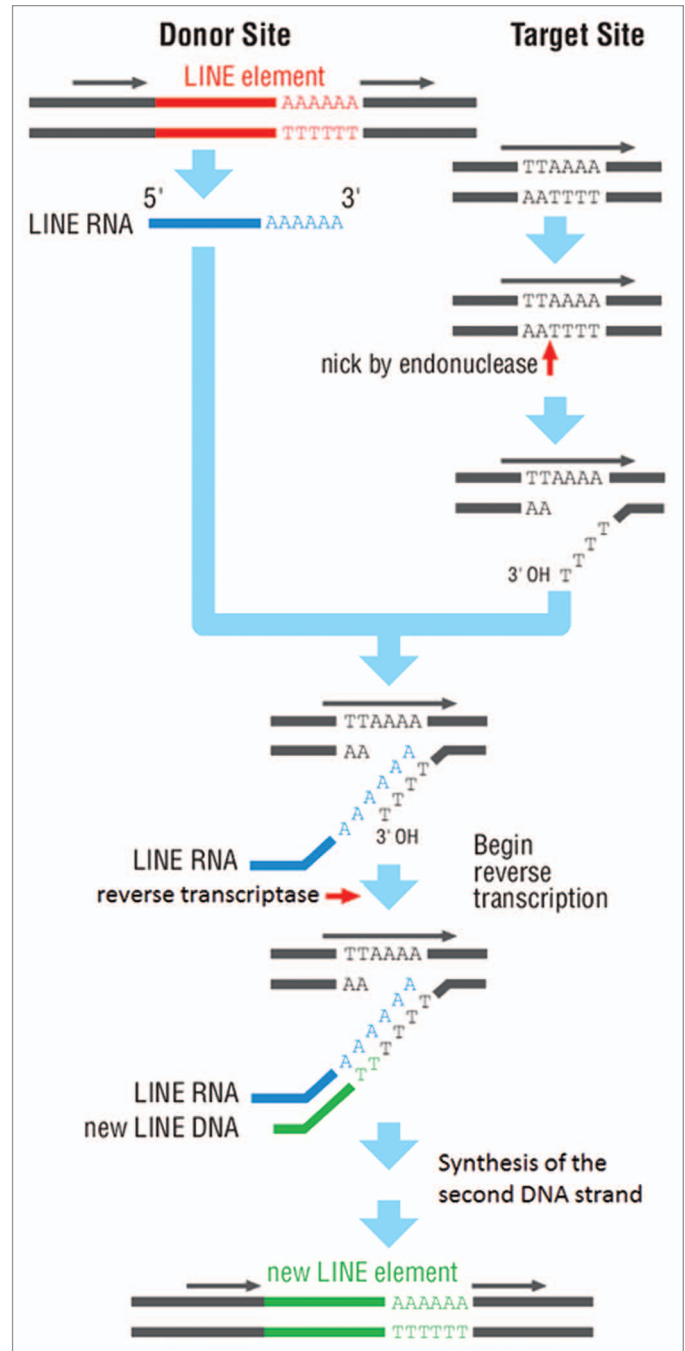
Figure 5. The mechanism of target-primed reverse transcription (TPRT). Transposition begins with the transcription of the LINE element (red) into RNA (blue) which encodes an RNA binding protein and a multifunctional protein with endonuclease and reverse transcriptase activity. These proteins (not shown) 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.

TPRT emerged by biochemical experiments with human L1 and Zorro3 retrotransposon. However, complete mechanistic details
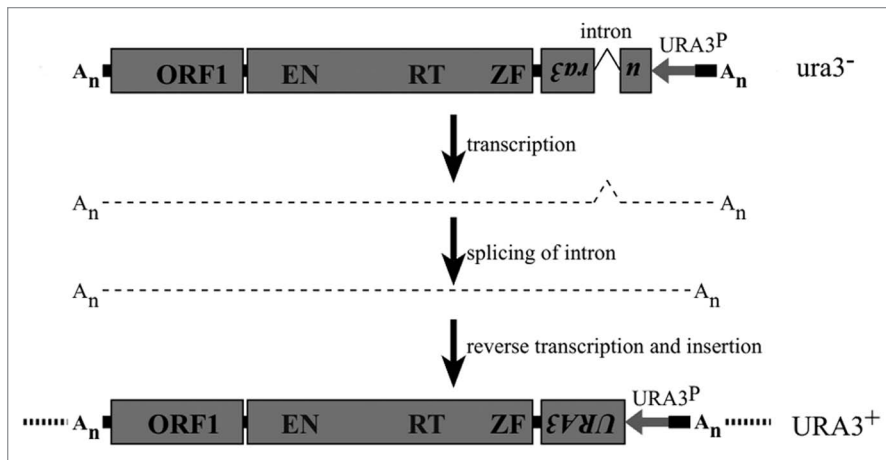
Figure 6. An assay for Zorro3 retrotransposition. The cloned Zorro3 element has a retrotransposition indicator gene (URA3 promoter, and URA3 ORF, disrupted by an antisense intron) inserted into its 3′ UTR. Reverse transcription and integration of the spliced RNA results in a functional and stably integrated URA3 gene and confers a URA3+ phenotype on the host cell.

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 complicate process that LTR retrotransposons can move from place to place in a genome by reverse transcription of an RNA transposition mediated in cells (in this study, we do not describe in details).[66] 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;[9] 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.[67,68] 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).[69] 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.[67,70,71] 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.[72] 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 shifts in diversity and copy number of non-LTR retrotransposons. For instance, non-LTR retrotrasposorons represented only by degenerate copies in *Drosophila* could lose these elements as a result of genetic drift, especially if the population is small.[72] In bacteria, Tn retrotransposons are likely to be principal players in the formation of tetracycline resistance by spreading drug resistance gene during genetic transfer.[73] 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[74,75] and human Alu.[76] 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.[77] SINE elements are much shorter (100–300 bp) than LINEs. A typical SINE consists of three parts:[78] 5′ ends of all SINEs families originating from one of the three types of short pol III transcripts: tRNAs, 5S rRNA, or 7SL RNA. The 3′ ends consist of poly A tails flanked by TSDs. The internal domain of the SINEs family

is usually unique and has no coding capacity. To date, 4 such domains have been described: CORE domain in vertebrates,[79] V-domain in fish,[80] Deu-domain in deuterostomes,[81] and Ceph-domain in cephalopods.[82]

SVA elements for another group of non-autonomous retro-elements in humans and non-human primate, and are present at a relatively low copy number of a few thousand per genome. The SVA elements were originally named SINE-R.[83] It is named "SVA" after its main components (SINE, VNTR, and Alu) by Shen et al.,[84] who identified the SINE-R element together 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 tails, 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 trans2.[85]

References

1. Odds 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. Construction of an *SfiI* macrorestriction map of the *Candida albicans* genome. J Bacteriol 1993; 175:6637-51; PMID:8407841

4. Chibana H, Magee BB, Grindle S, Ran Y, Scherer S, Magee PT. A physical map of chromosome 7 of Candida albicans. Genetics 1998; 149:1739-52; PMID:9691033

5. Goodwin TJ, Poulter RTM. Multiple LTR-retrotransposon families in the asexual yeast *Candida albicans*. Genome Res 2000; 10:174-91; PMID:10673276; http://dx.doi.org/10.1101/gr.10.2.174

6. Goodwin TJD, Ormandy JE, Poulter RTM. L1-like non-LTR retrotransposons in the yeast *Candida albicans*. Curr Genet 2001; 39:83-91; PMID:11405100; http://dx.doi.org/10.1007/s002940000181

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/10.1016/0092-8674(85)90197-7

8. Kazazian HH Jr, Moran JV. The impact of L1 retrotransposons on the human genome. Nat Genet 1998; 19:19-24; PMID:9590283; http://dx.doi.org/10.1038/ng0598-19

9. Zou S, Kim JM, Voytas DF. The Saccharomyces retrotransposon Ty5 influences the organization of chromosome ends. Nucleic Acids Res 1996; 24:4825-31; PMID:8972872; http://dx.doi.org/10.1093/nar/24.23.4825

10. Suzanne B. Sandmeyer. Yeast retrotransposons. Curr Opin Genet 1992; 2:705-11; http://dx.doi.org/10.1016/S0959-437X(05)80130-3

11. Uren AG, Kool J, Berns A, van Lohuizen M. Retroviral insertional mutagenesis: past, present and future. Oncogene 2005; 24:7656-72; PMID:16299527; http://dx.doi.org/10.1038/sj.onc.1209043

12. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 1993; 72:595-605; PMID:7679954; http://dx.doi.org/10.1016/0092-8674(93)90078-5

13. McDonald JF. Evolution and consequences of transposable elements. Curr Opin Genet Dev 1993; 3:855-64; PMID:8118210; http://dx.doi.org/10.1016/0959-437X(93)90005-A

14. Weiner AM, Deininger PL, Efstratiadis A. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 1986; 55:631-61; PMID:2427017; http://dx.doi.org/10.1146/annurev.bi.55.070186.003215

15. Xiong Y, Eickbush TH. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol Biol Evol 1988; 5:675-90; PMID:2464735

16. Doolittle RF, Feng DF, Johnson MS, McClure MA. Origins and evolutionary relationships of retroviruses. Q Rev Biol 1989; 64:1-30; PMID:2469098; http://dx.doi.org/10.1086/416128

17. Gabriel A, Boeke JD. Reverse transcriptase encoded by a retrotransposon from the trypanosomatid 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

18. Ivanov VA, Melnikov AA, Siunov AV, Fodor II, Ilyin YV. Authentic reverse transcriptase is coded by jockey, a mobile Drosophila element related to mammalian LINEs. EMBO J 1991; 10:2489-95; PMID:1714378

19. 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:1722352; http://dx.doi.org/10.1126/science.1722352

20. Cheng XD, Ling HQ. [Non-LTR retrotransposons: LINEs and SINEs in plant genome]. Yi Chuan 2006; 28:731-6; PMID:16818439

21. Piskurek O, Nishihara H, Okada N. 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 2009; 441:111-8; PMID:19118606; http://dx.doi.org/10.1016/j.gene.2008.11.030

22. Hohjoh H, Singer MF. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J 1996; 15:630-9; PMID:8599946

23. Hohjoh H, Singer MF. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. EMBO J 1997; 16:6034-43; PMID:9312060; http://dx.doi.org/10.1093/emboj/16.19.6034

24. Martin SL, Bushman FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol Cell Biol 2001; 21:467-75; PMID:11134335; http://dx.doi.org/10.1128/MCB.21.2.467-475.2001

25. Kolosha VO, Martin SL. High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). J Biol Chem 2003; 278:8112-7; PMID:12506113; http://dx.doi.org/10.1074/jbc.M210487200

26. Kulpa DA, Moran JV. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. Hum Mol Genet 2005; 14:3237-48; PMID:16183655; http://dx.doi.org/10.1093/hmg/ddi354

27. Martin SL, Cruceanu M, Branciforte D, Wai-Lun Li P, Kwok SC, Hodges RS, Williams MC. LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. J Mol Biol 2005; 348:549-61; PMID:15826653; http://dx.doi.org/10.1016/j.jmb.2005.03.003

28. Feng Q, Moran JV, Kazazian HH Jr., Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 1996; 87:905-16; PMID:8945517; http://dx.doi.org/10.1016/S0092-8674(00)81997-2

29. 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:1722352; http://dx.doi.org/10.1126/science.1722352

30. Martin SL. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol Cell Biol 1991; 11:4804-7; PMID:1715025

31. Kinsey JA. Tad, a LINE-like transposable element of *Neurospora*, can transpose between nuclei in heterokaryons. Genetics 1990; 126:317-23; PMID:2174012

32. Kubo S, Seleme MC, Soifer HS, Perez JL, Moran JV, Kazazian HHJr, Kasahara N. L1 retrotransposition in nondividing and primary human somatic cells. Proc. Natl. Acad 2006; 103:8036-41; http://dx.doi.org/10.1073/pnas.0601954103

33. Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res 1998; 8:464-78; PMID:9582191

34. Kazazian HH Jr., Moran JV. The impact of L1 retrotransposons on the human genome. Nat Genet 1998; 19:19-24; PMID:9590283; http://dx.doi.org/10.1038/ng0598-19

35. Schwarz-Sommer Z, Leclercq L, Göbel E, Saedler H. Cin4, an insert altering the structure of the A1 gene in Zea mays, exhibits properties of nonviral retrotransposons. EMBO J 1987; 6:3873-80; PMID:16453815

36. Vershinin AV, Druka A, Alkhimova AG, Kleinhofs A, Heslop-Harrison JS. LINEs and gypsy-like retrotransposons in Hordeum species. Plant Mol Biol 2002; 49:1-14; PMID:12008894; http://dx.doi.org/10.1023/A:1014469830680

37. Malik HS, Burke WD, Eickbush TH. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 1999; 16:793-805; PMID:10368957; http://dx.doi.org/10.1093/oxfordjournals.molbev.a026164

38. Malik HS, Eickbush TH. NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis* elegans. Genetics 2000; 154:193-203; PMID:10628980

39. Volff JN, Körting C, Schartl M. Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. Mol Biol Evol 2000; 17:1673-84; PMID:11070055; http://dx.doi.org/10.1093/oxfordjournals.molbev.a026266

40. Lovšin N, Gubenšek F, Kordi D. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in *Deuterostomia*. Mol Biol Evol 2001; 18:2213-24; PMID:11719571; http://dx.doi.org/10.1093/oxfordjournals.molbev.a003768

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

42. Burke WD, Malik HS, Rich SM, Eickbush TH. Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia.* Mol Biol Evol 2002; 19:619-30; PMID:11961096; http://dx.doi.org/10.1093/oxfordjournals.molbev.a004121

43. Biedler J, Tu Z. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity. Mol Biol Evol 2003; 20:1811-25; PMID:12832632; http://dx.doi.org/10.1093/molbev/msg189

44. Casaregola S, Neuvéglise C, Bon E, Gaillardin C. Ylli, a non-LTR retrotransposon L1 family in the dimorphic yeast *Yarrowia lipolytica.* Mol Biol Evol 2002; 19:664-77; PMID:11961100; http://dx.doi.org/10.1093/oxfordjournals.molbev.a004125

45. Hood ME. Repetitive DNA in the automictic fungus *Microbotryum violaceum.* Genetica 2005; 124:1-10; PMID:16010998; http://dx.doi.org/10.1007/s10709-004-6615-y

46. Gollotte A, L'Haridon F, Chatagnier O, Wettstein G, Arnould C, van Tuinen D, Gianinazzi-Pearson V. Repetitive DNA sequences include retrotransposons in genomes of the Glomeromycota. Genetica 2006; 128:455-69; PMID:17028973

47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403-10; PMID:2231712

48. Khazina E, Weichenrieder O. Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. Proc Natl Acad Sci U S A 2009; 106:731-6; PMID:19139409; http://dx.doi.org/10.1073/pnas.0809964106

49. Goodwin TJ, Busby JN, Poulter RT. A yeast model for target-primed (non-LTR) retrotransposition. BMC Genomics 2007; 8:263; PMID:17683538; http://dx.doi.org/10.1186/1471-2164-8-263

50. Forbes EM, Nieduszynska SR, Brunton FK, Gibson J, Glover LA, Stansfield I. Control of gag-pol gene expression in the Candida albicans retrotransposon Tca2. BMC Mol Biol 2007; 8:94; PMID:17961216; http://dx.doi.org/10.1186/1471-2199-8-94

51. Goodwin TJD, Poulter RTM. The DIRS1 group of retrotransposons. Mol Biol Evol 2001; 18:2067-82; PMID:11600703; http://dx.doi.org/10.1093/oxfordjournals.molbev.a003748

52. Zingler N, Weichenrieder O, Schumann GG. APE-type non-LTR retrotransposons: determinants involved in target site recognition. Cytogenet Genome Res 2005; 110:250-68; PMID:16093679; http://dx.doi.org/10.1159/000084959

53. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. EMBO J 2002; 21:5899-910; PMID:12411507; http://dx.doi.org/10.1093/emboj/cdf592

54. Kazazian HH Jr. Mobile elements: drivers of genome evolution. Science 2004; 303:1626-32; PMID:15016989; http://dx.doi.org/10.1126/science.1089670

55. Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol 1990; 10:6718-29; PMID:1701022

56. Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. Annu Rev Genet 2001; 35:501-38; PMID:11700292; http://dx.doi.org/10.1146/annurev.genet.35.102401.091032

57. Pickeral OK, Makałowski W, Boguski MS, Boeke JD. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res 2000; 10:411-5; PMID:10779482; http://dx.doi.org/10.1101/gr.10.4.411

58. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921; PMID:11237011; http://dx.doi.org/10.1038/35057062

59. Martin SL, Li J, Weisz JA. Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1. J Mol Biol 2000; 304:11-20; PMID:11071806; http://dx.doi.org/10.1006/jmbi.2000.4182

60. Ostertag EM, Kazazian HH Jr.. Biology of mammalian L1 retrotransposons. Annu Rev Genet 2001; 35:501-38; PMID:11700292; http://dx.doi.org/10.1146/annurev.genet.35.102401.091032

61. Dong C, Poulter RT, Han JS. LINE-like retrotransposition in Saccharomyces cerevisiae. Genetics 2009; 181:301-11; PMID:18957700; http://dx.doi.org/10.1534/genetics.108.096636

62. Han JS, Shao S. Circular retrotransposition products generated by a LINE retrotransposition. Nucleic Acids Res 2012; 40:10866-77; PMID:22977178; http://dx.doi.org/10.1093/nar/gks859

63. Stanfield S, Helinski DR. Small circular DNA in Drosophila melanogaster. Cell 1976; 9:333-45; PMID:824055; http://dx.doi.org/10.1016/0092-8674(76)90123-9

64. Flavell AJ, Ish-Horowicz D. Extrachromosomal circular copies of the eukaryotic transposable element copia in cultured Drosophila cells. Nature 1981; 292:591-5; PMID:6265802; http://dx.doi.org/10.1038/292591a0

65. Whitcomb JM, Hughes SH. Retroviral reverse transcription and integration: progress and problems. Annu Rev Cell Biol 1992; 8:275-306; PMID:1282352; http://dx.doi.org/10.1146/annurev.cb.08.110192.001423

66. Boeke JD, Corces VG. Transcription and reverse transcription of retrotransposons. Annu Rev Microbiol 1989; 43:403-34; PMID:2552899; http://dx.doi.org/10.1146/annurev.mi.43.100189.002155

67. Arkhipova IR. Mobile genetic elements and sexual reproduction. Cytogenet Genome Res 2005; 110:372-82; PMID:16093689; http://dx.doi.org/10.1159/000084969

68. Johnson LJ. The genome strikes back: the evolutionary importance of defense against mobile elements. Evol Biol 2007; 34:121-9; http://dx.doi.org/10.1007/s11692-007-9012-5

69. Pickeral OK, Makałowski W, Boguski MS, Boeke JD. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res 2000; 10:411-5; PMID:10779482; http://dx.doi.org/10.1101/gr.10.4.411

70. Hua-Van A, Le Rouzic A, Maisonhaute C, Capy P. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. Cytogenet Genome Res 2005; 110:426-40; PMID:16093695; http://dx.doi.org/10.1159/000084975

71. Le Rouzic A, Capy P. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. Genetics 2005; 169:1033-43; PMID:15731520; http://dx.doi.org/10.1534/genetics.104.031211

72. Brookfield JF, Badge RM. Population genetics models of transposable elements. Genetica 1997; 100:281-94; PMID:9440281; http://dx.doi.org/10.1023/A:1018310418744

73. Rice LB. Tn916 family conjugative transposons and dissemination of antimicrobial resistance determinants. Antimicrob Agents Chemother 1998; 42:1871-7; PMID:9687377

74. Krayev AS, Kramerov DA, Skryabin KG, Ryskov AP, Bayev AA, Georgiev GP. The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. Nucleic Acids Res 1980; 8:1201-15; PMID:7433120; http://dx.doi.org/10.1093/nar/8.6.1201

75. Krayev AS, Markusheva TV, Kramerov DA, Ryskov AP, Skryabin KG, Bayev AA, Georgiev GP. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. Nucleic Acids Res 1982; 10:7461-75; PMID:6296779; http://dx.doi.org/10.1093/nar/10.23.7461

76. Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW. Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J Mol Biol 1981; 151:17-33; PMID:6276559; http://dx.doi.org/10.1016/0022-2836(81)90219-9

77. Prak ET, Kazazian HH Jr. Mobile elements and the human genome. Nat Rev Genet 2000; 1:134-44; PMID:11253653; http://dx.doi.org/10.1038/35038572

78. Kramerov DA, Vassetzky NS. SINEs. Wiley Interdiscip Rev RNA 2011; 2:772-86; PMID:21976282; http://dx.doi.org/10.1002/wrna.91

79. Gilbert N, Labuda D. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. Proc Natl Acad Sci U S A 1999; 96:2869-74; PMID:10077603; http://dx.doi.org/10.1073/pnas.96.6.2869

80. 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. Genome Res 2002; 12:316-24; PMID:11827951; http://dx.doi.org/10.1101/gr.212302

81. Nishihara H, Smit AF, Okada N. Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res 2006; 16:864-74; PMID:16717141; http://dx.doi.org/10.1101/gr.5255506

82. Akasaki T, Nikaido M, Nishihara H, Tsuchiya K, Segawa S, Okada N. Characterization of a novel SINE superfamily from invertebrates: "Ceph-SINEs" from the genomes of squids and cuttlefish. Gene 2010; 454:8-19; PMID:19914361; http://dx.doi.org/10.1016/j.gene.2009.11.005

83. Ono M, Kawakami M, Takezawa T. A novel human nonviral retroposon derived from an endogenous retrovirus. Nucleic Acids Res 1987; 15:8725-37; PMID:2825118; http://dx.doi.org/10.1093/nar/15.21.8725

84. Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, Zipf WB, Yu CY. 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 1994; 269:8466-76; PMID:8132574

85. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. SVA elements: a hominid-specific retroposon family. J Mol Biol 2005; 354:994-1007; PMID:16288912; http://dx.doi.org/10.1016/j.jmb.2005.09.085