The structure and retrotransposition mechanism of LTR-retrotransposons in the asexual yeast

_Candida albicans_

**Lulu Zhang**1,†, Lan Yan1,†, Jingchen Jiang2, Yan Wang1, Yuanying Jiang1, Tianhua Yan2,*, and Yongbing Cao1,*

1Research and Develop Center of New Drug; School of Pharmacy; Second Military Medical University; Shanghai, PR China; 2Department of Pharmacology; School of Pharmacy; China Pharmaceutical University; Nanjing, PR China

*Correspondence to: Tianhua Yan; Email: 13016931191@126.com; Yongbing Cao; Email: ybcaco@vip.sina.com
Submitted: 12/09/2013; Revised: 07/25/2014; Accepted: 07/28/2014; Published Online: 08/07/2014
http://dx.doi.org/10.4161/viru.32180

**Keywords:** _Candida albicans_, LTR-retrotransposons, _Tea_, structure feature, retrotransposition mechanism

Retrotransposons constitute a major part of the genome in a number of eukaryotes. Long-terminal repeat (LTR) retrotransposons are one type of the retrotransposons. _Candida albicans_ have 34 distinct LTR-retrotransposon families. They respectively belong to the Ty1/copia and Ty3/gypsy groups which have been extensively studied in the model yeast _Saccharomyces cerevisiae_. LTR-retrotransposons carry two LTRs flanking a long internal protein-coding domain, open reading frames. LTR-retrotransposons use RNA as intermediate to synthesize double-stranded DNA copies. In this article, we describe the structure feature, retrotransposition mechanism and the influence on organism diversity of LTR retrotransposons in _C. albicans_. We also discuss the relationship between pathogenicity and LTR retrotransposons in _C. albicans_.

**Introduction**

As an opportunistic fungal pathogen of humans, _Candida albicans_ usually cause fungal infections such as thrush, vaginitis, and life-threatening bloodborne candidiasis.1 Of the 16 Mb haploid genome of _C. albicans_, 14.9 Mb has been assembled by Stanford University (Stanford DNA Sequencing and Technology Center), which facilitates understanding of the genetic structure, pathogenicity, adaptability, and evolution in _C. albicans_. Retrotransposons are mobile genetic elements capable of independent transposition through RNA intermediates.1 Retrotransposons are widespread transposable elements in eukaryotes, and constitute a major part of eukaryotic genomes in many cases.2-7 For instance, retrotransposons constitute 42% of human genomes8 and 75% of maize genomes.9

Two primary subclasses of retrotransposons are present in eukaryotic cells: long-terminal repeat (LTR) retrotransposons and non LTR retrotransposons (LINEs, long interspersed nuclear elements, and SINEs, short interspersed nuclear elements).10 LTR retrotransposons in the yeast _Saccharomyces cerevisiae_ have been extensively studied. According to the sequence similarity of reverse transcriptases (RTs)11,12 and the subunits of pol genes, there are two different types of LTR retrotransposons.1 One is the Ty1/copia type, where the subunits are protease (PR), integrase (IN), reverse transcriptase (RT), and RNase H (RH) in order; the other is the Ty3/gypsy type, where the subunits are PR, RT, RH, and IN (Fig. 1).13 In _S. cerevisiae_, six classes of retrotransposons, Ty1–5 and Ty3p, belong to 2 different types, respectively: Ty1/copia components (Ty1, Ty2, Ty4, and Ty5), and Ty3/gypsy components (Ty3 and Ty3p). In these retrotransposons families, Ty1, Ty2, Ty3, and Ty4 have transposition activity in _S. cerevisiae_ as known.14 This review summarizes the structure, mechanism, and influence on organism diversity of LTR-retrotransposons found in _C. albicans_.

**LTR-Retrotransposon Families in _C. albicans_**

Compared with the 6 families present in _S. cerevisiae_, 34 distinct LTR-retrotransposon families pertaining to the Ty1/copia and Ty3/gypsy groups have been found in _C. albicans_.15 They are named by the Greek alphabet letters (α, β, ..., iota.), archaic Greek letters (sampi, san, etc.), and phonetically similar names of New Zealand birds (mous, tara, weka, etc.).15 According to the internal regions, the first 16 LTRs have been defined as _Tea1–Tea16_ (Table 1).

By comparing the LTR retrotransposons families in _C. albicans_ and _S. cerevisiae_, we can find a great difference.15 The different number of families illustrates that the two species experience disparity in retrotransposons element evolution.

LTR retrotransposons contain two long-terminal repeats (LTRs), as shown in **Figure 1**, at their ends, typically 250–600 bp in length, flanking a 5–7 kb long internal protein-coding domain.16 Between the two LTRs are two open reading frames (ORFs): gag and pol. The gag ORF encodes the structural proteins that make up a virus like particle (VLP).17 The pol ORF encodes the enzymes required for reverse transcription and integration. The enzymes are PR, IN, RT, and RH.
The LTRs of retrotransposons have the transcription termination signals and promoter, and are divided into three functional areas: U3, R, and U5. U3 contains an enhancer and promoter region at the 3' end of transcript; R contains both the start and termination sites for transcription; U5 only exists in the transcript of 5' terminal. The process of transcription is from the left LTR U3/R boundary to the right LTR R/U5 boundary to produce a RNA molecule with both ends of R region (Fig. 2).

Thirty-four LTR retrotransposon families in *C. albicans* share the following characteristics: (1) the length of the retrotransposon LTRs ranges from 127 to 780 bp with a mean length of 359 bp; (2) most elements have the terminal dinucleotides 5'-TG…CA-3' and these dinucleotides tend to form part of larger terminal inverted repeats; (3) the total cope number of LTRs is 355; (4) both sides of more than half of the full-length LTR are short direct repeats representing target-site duplications (TSDs),

| LTR  | Length (bp) | TSD (bp) | Associated internal region | Copy number* | Accession number | Gag/pol ORFs* |
|------|-------------|----------|-----------------------------|--------------|------------------|---------------|
| alpha| 388         | 5        | Tca1                        | 10 (5–10)    | M94628           | No            |
| beta | 395         | 5        | Tca8                        | 10 (6–8)     | Y08494           | Partial       |
| gamma| 280         | 5        | Tca2                        | 9 (5–10)     | Y08494           | Intact        |
| kappa| 280         | 5        | Tca6                        | 20 (10–15)   | AF069450         | Frags         |
| zeta | 508         | 5        | Tca7                        | 19 (10–15)   | AF074943         | Partial       |
| san  | 381         | 5        | Tca4                        | 5 (1–4)      | AF074943         | Intact        |
| omega| 685         | 5        | Tca5                        | 3 (0–5)      | AF093417         | Intact        |
| nu   | 277         | 4        | Tca3                        | 11           | AF119344         | Partial       |
| psi  | 470         | 5        | Tca9                        | 30           | AF119344         | No            |
| chi  | 192         | 5        | Tca10                       | 11           | AF118059         | No            |
| eta  | 470         | 5        | Tca11                       | 13           | AF118059         | N.D.          |
| whio | 348         | 5        | Tca12                       | 8            | AF180289         | N.D.          |
| moo  | 507         | 5        | Tca13                       | 8            | AF180291         | No            |
| lambda| 512        | 5        | Tca14                       | 4            | AF180284         | N.D.          |
| kahu | 531         | 5        | Tca15                       | 17           | AF192278         | Partial       |
| huiia| 127         | 5        | Tca16                       | 6            | AF180285         | Frag.s        |
| omicron| 268        | 4        | -                           | 9            | -                |               |
| rho  | 275         | 4        | -                           | 13           | -                |               |
| pi   | 280         | 4/5      | -                           | 17           | -                |               |
| lota | 251         | 4        | -                           | 13           | -                |               |
| sampia| 324        | 5        | -                           | 9            | -                |               |
| theta| 366         | 5        | -                           | 7            | -                |               |
| upsilon| 264        | 5        | -                           | 6            | -                |               |
| kappa| 208         | 5        | -                           | 10           | -                |               |
| epsilon| 480        | 5        | -                           | 9            | -                |               |
| phi  | 194         | 5        | -                           | 14           | -                |               |
| episeman| 518       | 5        | -                           | 4            | -                |               |
| mu   | 780         | 5        | -                           | 4            | -                |               |
| xi   | 387         | 5        | -                           | 5            | -                |               |
| weka | 165         | 5/7      | -                           | 9            | -                |               |
| tui  | 199         | 5        | -                           | 19           | -                |               |
| titi | 336         | 5        | -                           | 3            | -                |               |
| tara | 285         | 5        | -                           | 9            | -                |               |
| tora | 282         | 5        | -                           | 11           | -                |               |

*Numbers in parentheses indicate the copy number in a variety of strains as estimated by Southern blotting. *Intact, long ORFs present containing all of the motifs characteristic of full-length, functional retrotransposons; Partial, long ORFs are present but no full-length ones discovered yet; Frags, no long ORFs, but short regions of homology to other retroelement ORFs are present; No, no long ORFs present nor any homology to other retroelements in databases.*
most of which are 5 bp in length showing a tendency toward a purine in the first position and a pyrimidine in the last (data not shown); and (5) these retrotransposon families in C. albicans generally vary widely in coding capacity as Ty families in S. cerevisiae and have a low copy number. Most families have the similar copy number to Ty5 (7 copies) LTR in S. cerevisiae, but considerably lower than that of Ty1/Ty2 (251 copies), Ty3 (41 copies), and Ty4 (32 copies). Most LTR families in C. albicans only contain solo LTRs or LTR remnants in which the LTR-retrotransposon sequence has been retained and the intervening sequence has been lost. So far, just 3 intact retrotransposons have been founded. They are (1) Tca2, an active retrotransposon element which has a stop codon between two ORFs, (2) Tca5, a Ty5 like retrotransposon, and (3) Tca4, which belongs to Ty1/copia group and resembles to Tca2. Several families of highly degenerate elements appear to be still capable of transposition, presumably via transactivation. Full-length retrotransposons are usually lost when two LTRs recombine with each other. This results in isolated solo LTRs at the original sites. A survey shows that 85% of Ty insertions in S. cerevisiae are solo LTRs or LTR fragments. In C. albicans, for retrotransposon element, β, there are some 395 bp solo LTRs. Both full-length retrotransposons and solo LTRs have short (4 or 5 bp) direct repeats on either end. Most LTR retrotransposons generate 5 bp short direct repeats representing TSDs, as the Ty elements in S. cerevisiae. For instance, 36% of the total S. cerevisiae Ty1–4 elements are flanked by TSDs. Except for Ty3/gypsy elements (Ty3) that are flanked by 4 bp TSDs, Ty5 elements have no TSDs. The condition of Ty5 suggests frequent recombination between these elements at the telomeres. Recombination or mutation may have resulted in exchange of target site sequences between elements.

A similar preference is apparent for the elements in C. albicans: 14 full-length LTRs are flanked by 4 bp direct repeats, like element pt, weka has a confirmed 7 bp TSD. With analysis of the 5 bp TSDs in C. albicans and S. cerevisiae, base bias can be found: purine highly is enriched in the first site (55%); and pyrimidine is enriched in the last site (53%) and there is a strong bias for A and T: in the internal position 2 (72%), position 3 (76%), and position 4 (78%) (data not shown). LTR retrotransposons have a potential tRNA primer-binding site. At the downstream from the left of 5′ LTR is a short poly-purine sequence (8–49 nt), termed as the primer-binding sites (PBSs) (Table 1), which is a 10–20 nucleotide sequence that can base-pair with cytoplasmic tRNA molecule partly. Fourteen different families of LTRs in C. albicans have connection with PBSs.

All PBSs share a common feature, the LTR terminus starting from 5′-TGG-3′ is base-pair with CCA sequence at the 3′ end of all tRNAs. Goodwin et al. had found the 3′ end of tRNA could be used as a primer to determine whether two LTRs were two independent insertions or a single retrotransposon. Furthermore, if the two LTRs belong to a family, they repeat in a direct orientation on a contig. According to the above mentioned common features, 14 families of LTRs in C. albicans have connection with PBSs. The condition of Ty5 suggests frequent recombination between these elements at the telomeres. Recombination or mutation may have resulted in exchange of target site sequences between elements.
A short gap between the LTR and the PBS is a common feature of *gypsy*-class retrotransposons.24

**Mechanisms of Retrotransposition**

A general phenomenon that resides in a number of eukaryotes is that adverse living conditions can activate retrotransposons. They can move from place to place in a genome by reverse transcription of a RNA transposition intermediate to enable the organism to adapt to the environment. It has been demonstrated that the retrotransposons generate new copies by reverse transcription of their RNA transcripts.3

Similar to retroviruses in function, the LTRs of retrotransposons play a border role in complex reverse transcription procedure.25 The two DNA strands are synthesized from opposite directions: tRNA binds to the site near the 3' end of left LTR when synthesizing the first strand, while the polypurine track (PPT), a
short purine-rich sequence immediately upstream of the right LTR, as a primer binds to the upstream of the right LTR when synthesizing the second strand.

Retrotransposons must synthesize mRNA at first, which can be translated into proteins related to replication, also acts as a replication template. At each end of the template, mRNA has a short repeat sequence (R, green) and U region (Fig. 4). Primer tRNA can base-pair with PBS sequence. When tRNA anneals to PBS, minus strand DNA starts synthesizing under the catalysis of reverse transcriptase. DNA synthesized can reach the U5 region (red) at the 5’ end of the template and continue as the last few bases. The newly synthesized DNA can base-pair with the repeat sequence at the 3’ end of the genomic RNA. Then, new strand DNA base-pair with 3’ end of the RNA template starting with U5 sequence, and continues until R and U5 at the 5’ end of RNA template. After that the RNase H degrades the RNA template leaving fragments at the poly-purine tract to prime second strand DNA synthesis. Using the new strand DNA as a template reverse transcriptase synthesizes another strand DNA until U5 and R. After that, the DNA fragment transfers to another end of new template. Synthesis then proceeds in both directions to give double-stranded DNA and LTR, making up of U5, R and U3, at each end. Integrase inserts this into chromosomal DNA, and transcription initiating in one LTR and terminating in the other generates genomic RNA with terminal repeats.

Specific Description

Of the 34 distinct LTR-retrotransposon families in C. albicans, 16 families still retain some internal sequences which have very different coding capacity (Tca1–Tca16). Tca2, Tca4, and Tca5 are intact and possess all the characteristic features of functional retrotransposons. Tca3 and Tca8 have no full-length elements, but still retain long and continuous ORFs like other retroposons; Tca6 still contains ORF fragments in the internal region; Tca9 and Tca13 have 5 bp direct repeats at each end, intact PBSs and PPTs and identical LTRs; Tca10 is a compound element, whose LTR shares 99.5% identity and is flanked by a 5 bp direct repeat. Some extensive ORFs, the PBS and PPT regions make up the ~2 kb long internal region. The other 18 LTR families do not have the sequences similar to retrotransposons internal regions. These retrotransposon families may remain solo LTRs and LTR fragments, resulting in some of these LTRs escaping from detection.
Many researchers have studied a few LTR retrotransposons in 

*C. albicans*. Now we summarize the characteristics of Tca1, Tca2, Tca3, Tca4, and Tca5.

**Tca1, an Inactive Element without ORF**

*Tca1*, as the first *C. albicans* retroelement, is of 5614 bp and has 388 bp LTRs. No significant ORF sequence presents in the internal region, suggesting that *Tca1* is a degenerate and inactive element. It has been demonstrated that *Tca1* has two loci in the strain SC5314. One is referred to as *Tca1–1*, and the other is designated as *Tca1–2*. The *Tca1* elements of the two loci have more than 99% similarity in sequence and almost the same structure. *Tca1–1* was isolated from a lambda phage genomic library clone called CJY-3. This clone contains 3.25 kb and 4.8 kb long fragments (Fig. 5). *Tca1–1* lacks any retrotransposon-like ORF. This differentiates *Tca1–1* from *Tca1–2*. *Tca1–2* was found in an approximately 6 kb clone, called CJY-4. It is highly similar to the insert of CJY-3, besides minor restriction site polymorphisms (Fig. 6).

Nucleotide sequence analysis suggests that the insert of CJY-4 is flanked by identical direct repeat sequences, which was 99% identical to the LTRs of *Tca1–1*. Each LTR of *Tca1–2* is flanked by 6 bp inverted repeat sequence (TGTTCG) like LTRs of other retrotransposons. The sequence in the front of 5′ LTR and behind of 3′ LTR of *Tca1–2* is ATTGC. This suggests that the sequence is a copy of integration target site. The sequence is different from the one of *Tca1–1*, TTGGT.

Sequence analysis of *Tca1–2* has not found any extended ORFs or potential splice sites. This suggests that *Tca1–2* has been degenerated highly. Compared with about 1400 bp of insertion region of *Tca1–1*, *Tca1–2* is just differing in less than 1% (data not shown).

Previous work indicated that the transcription of *Tca1–1* is affected by growth temperature. Northern blot analysis suggested that the transcription of *Tca1* is also influenced by temperature in strains lacking *Tca1–1* or *Tca1–2*. Thus, *Tca1* elements are temperature-regulated retrotransposons. The expression level of *Tca1* at 25 °C is higher than that at 37 °C.

It has been reported that the virulence determinants of a lot of pathogenic bacteria (*Shigella* app, *Yersinia* app, *Vibrio* cholerae, *Bordetella* pertussis, and *Staphylococcus* aureus) were regulated by temperature. Antley et al. found *C. albicans* cells grown at 25 °C are more virulent than those grown at 37 °C. As *Ty* elements can regulate adjacent gene transcription the *Tca1* may control adjacent genes temperature-dependently. The expression of *Tca1* could be similar to the invasion gene in *Yersinia pseudotuberculosis*.

Southern blot analysis suggests that deletion of *Tca1* elements does not result in obvious growth defects and the recombination between LTRs occurs readily. Furthermore, silencing a single element would lead to complete loss at that locus, suggesting that the *Tca1* elements at both loci are hemizygous. It is known that elements may be abundant in natural populations if they provide some advantage to the microorganism. However, there is no clear reason to maintain these degenerate elements. Neither copy of *Tca1* choose silent chromatin as preferential integration region as the *Ty5* retrotransposon in *S. cerevisiae*. Although nearly 40% of the strains lack *Tca1* elements, solo LTRs in the genome suggest that they have existed in these strains for a time, and moreover, it is not essential to maintain *Tca1* in natural populations.

**Tca2 is an Abundant, Extrachromosomal DNA Molecule with Intact ORF**

*Tca2* is widespread in *C. albicans* and was first identified in the strain hOG1042, known as pCal. The pCal was a distinct band when the uncult *C. albicans* DNA was examined on an agarose gel, so it is an extrachromosomal retrotransposon. *Tca2* belongs to *Ty1/copia*-type retrotransposon, and was originally identified as a linear double-stranded DNA molecule. It is the most abundant DNA copy (about 50 copies per cell). *Tca2* is 6426 bp long and has 280 bp LTRs flanked by a six-nucleotide imperfect inverted repeat, TGTGGG...CCATCA. The internal domain usually has two long *gag* and *pol*-like ORFs, ORF1 and ORF2 are separated by a stop-codon (UGA). The structure feature is similar to mammalian retroviruses but is unique in LTR retrotransposons. However, the *pol* gene can translate into protein. Forbes and coworkers demonstrated that the LTR promoter

---

**Figure 5.** Southern blot analysis of DNA from lambda clones CJY-3 and CJY-4. Genomic DNA from strain SC5314 or purified DNA from lambda clones CJY-3 and CJY-4 was digested with EcoRI and hybridized with the α-element probe.
is heat-shock induced. Like pol proteins, the expression of Tca2 pol protein is regulated by LTR promoter. There is a novel stop codon bypass mechanism to make gag-pol RNA translate pol proteins. The mechanism is read-through suppression of the UAG codon. In Tca2, the sequences between the gag and pol ORFs are an 8 bp purine-rich sequence, AAAACAGG, and a potential pseudoknot. They are tightly followed the UGA stop codon, which is essential to suppression.

Tca2 has retrotranspositional activity, and is strongly dependent on the growth temperature. There is substantially more extrachromosomal DNA at 37 °C than that at 27 °C. In addition, northern blot analysis on the transcriptional activity of Tca2 showed that the most prominent band is full-length transcripts. Interestingly, the number of full-length transcripts at 37 °C is more than that at 27 °C, suggesting that Tca2 has transposition activity and is in favor of higher temperature (37 °C). There is also a correlation between the level of Tca2 RNA and extrachromosomal DNA, suggesting that the synthesis of extrachromosomal DNA needs Tca2 RNA as intermediate. The level of transcription is probably affected by strains, sequences and genomic context. Thus, we speculate that there may be close relationship between the Tca2 expression and virulence in the process of C. albicans evolution.

Tca3 has a Partial Coding Region

Tca3 is a Ty3/gypsy-like retrotransposon widely existing in C. albicans. It has 277bp LTRs and contains a little part of a coding region.

There are two major data sets of genome sequence have been determined for C. albicans strain SC5314. The first is Assembly 6 and another is Assembly 19. Assembly 6 represents an initial draft of the diploid genome, sharing a similar Tca3 complement to Assembly 6. However, it has some information which Assembly 6 does not contain, such as a composite on contig 19–10140. Assembly 19 contains identical
LTRs flanking a target site duplication instead of the internal region. Data suggest that the sequence of internal region is substituted by a run of Ns. Highly similarity to other Tca3 elements may lead the internal sequence co-assemble with them mistakenly. Tca3 is roughly the same as other Ty3/gypsy elements in structure except for some special features. For example, elements of Tca3 seem to have a distinct mechanism for DNA synthesis which is involve in minus-strand DNA. Second, elements have a distant relationship with other Ty3/gypsy-like elements. Third, the ORFs of Tca3 elements have no obvious pol domain and shorter RT domain as typical Ty3/gypsy elements. Finally, losing the pol domain and large part of gag domain is another special feature.44

A copy of Tca3 obtained from ATCC10261 that retains the gag/pol region is referred to as Tca3F-10261 (Fig. 7B). It is found that Tca3A is more common than Tca3F. Tca3F is 6134 bp long. The internal region contains two long ORFs flanking by identical 313 bp LTRs. Tca3F-10261 has 99% identity with Assembly 6 Tca3A elements, except for the -1.6 kb indel covering the gag/pol region, a 42 bp indel upwards to the large indel, and some variation in the 5′ untranslated regions.

Goodwin et al.44 have found 5 of 6 strains emerged Tca3 bands, suggesting that Tca3 is quite common in C. albicans. In the five examined strains, only 1 to 3 bands were produced, which indicates that the elements are present in low copy number (data not shown). This might be the chance events in evolutionary process leading to excessive mutation. This might be the chance events in evolutionary process leading to excessive mutation. This might be the chance events in evolutionary process leading to excessive mutation. This might be the chance events in evolutionary process leading to excessive mutation. This might be the chance events in evolutionary process leading to excessive mutation. This might be the chance events in evolutionary process leading to excessive mutation.

Sequence comparison and Southern analysis have shown that Tca3 is much conserved in sequence and structure, either in loci or in strains, suggesting that the original deletion event has occurred not long ago.

**Tca4, an Active Element, Contains Intact ORF**

Tca4 is a Ty1/copia element closely related to Tca2,46 which is flanked by 381 bp LTRs and contains an intact gag/pol ORFs. Tca4 is simple in sequence, suggesting that it has recently emerged. Like Tca2, Tca4 is a retrotranspositionally active element that strongly depends on the growth temperature. High temperature can induce Tca4 transcription.

**Tca5, an Intact LTR Retrotransposon**

Tca5 is of ~5.6 kb and 4218 bp of internal sequence flanked by identical 685 bp LTRs.48 Only one single long ORF lies in the internal region. It is larger than most other yeast retrotransposon LTRs. Phylogenetic analysis and sequence comparison imply that Tca5 has close connection with Ty5 element in *S. cerevisiae*. The *S. cerevisiae* Ty5 LTRs range in length from 251 bp for Ty5Δ to 371bp for Ty4.49 The terminal inverted repeat sequences are the most conserved in Ty1-like elements, containing at least 5 nt: 5′-TGTTG...CAACA-3′. Therefore, the Tca5 retrotransponson begins with TG and ends with CA, as most retrotransposons do. Sequencing of the PCR product revealed the left LTR, which was identical to the right LTR, and these LTR are designated as omega.

Close behind the downstream of left LTR is PBS which is complementary to methionine tRNAiMet15 of initiator. Upstream of right LTR is a tract enriched in purines (ATGGGGAAG) and a similar sequence (ATGGGGAGG) at position 3132 bp.49

The internal region contains a single and long uninterrupted ORF with a relatively low copy number, and this ORF extends from within the left LTR (382 bp) to within 60 bp of the right LTR, with no inframe stop codon or frame shift, predicting a protein with motifs characteristic of gag, protease, integrase, reverse transcriptase, and RNAse H proteins of other retrotransposons.49

However, it is unclear where the ORF might initiate translation in vivo.49 There are several possible sites,49 all of which are ATG in sequence. The first one within the ORF (382 bp from the 5′ end of the left LTR) is the unusually deep inset within the LTR. The second is also within the LTR three codons beyond the first. The third one within the ORF is unusually far beyond the end of the LTR, the three ATG are in-phase. Besides, there are several out-of-phase ATG triplets 5′ to this candidate.49
heterogeneity of the Candida genome, either by recombination between the LTRs of one element or by chromosomal rearrangements between two elements. Tea5 may still be active which is supported by the detection of full-length transcript in northern blot. 19

In C. albicans, Tea5 is the only element showing significant similarity to Ty5. 16 It is a suitable candidate for comparisons because it has all the expected structures of an active element and is probably intact. Comparative genomic analysis between Tea5 and Ty5 elements will be helpful in understanding them.

**LTR-Retrotransposons Improve Genetic Diversity**

* C. albicans is a diploid organism. 50 Clinical isolates of *C. albicans* show variform phenotypic characteristics associated with infection and virulence. 31-37 These complex phenotypes suggest genetic diversity. However, *C. albicans* is an asexual fungus; it must have other mechanisms to generate various genomes. Retrotransposons have contributed to the evolution of genomes and genes in ways that go well beyond simply increasing genome size. It also affects genomes on a small scale, such as mobilization and integration at a locus directly, or recombination with heterotopic element indirectly. These changes contain gene inactivating, variance in transcription ability of gene, gene deletions and inversions. 29 It has been well documented that retrotransposons would mobilize in response to stress. 58-60 Servant et al. 61 have shown that the relationship between transposable activity of Ty1 LTR retrotransposons and genome expression in *S. cerevisiae*. It seems to have a connection between stress reaction and the diversity of retrotransposons in *C. albicans*. Many organisms have the ability to recruit LTR-retrotransposons as alternative exons or promoters to drive genome evolution. 62-65 The LTRs of Ty elements have the ability to modify genome transcription and transposable elements can influence the expression of neighboring host genes. 61,66 In gram-positive bacteria, circularized form of *Tn* retrotransposons can promote organisms resistance to antimicrobial. 67 There is an article 33 about the relationship between virulence and reverse transposition of LTR retrotransposons in *C. albicans*. Take the Tea1 for example, the expression level of Tea1 is related to temperature, suggesting that Tea1 may place adjacent to the temperature-dependent genes. Since Tea1 exhibits higher expression at 25 °C than at 37 °C, Tea1 could upregulate the expression of genes required for the establishment of infection. Furthermore, in our original research, we found the transposition phenomenon of LTR-retrotransposon in strains which are resistant to miconazole (unpublished observation). It is suggested that there is a relationship between virulence and LTR-retrotransposons.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

L.Y. is supported by Natural Science Foundation of China (31000079). Y.C is supported by Natural Science Foundation of China (81713100) and Shanghai Basic Research Project (No11JC145400). Y.J. is supported by China National 973 Program (2013CB531602) and Natural Science Foundation of China (81330083). The authors would like to apologize to all researchers whose important works were unable to be cited because of space limitations.

**References**

1. Odds FC. Pathogenesis of Candida infections. J Am Acad Dermatol 1994; 31:52-5; PMID:8077502; http://dx.doi.org/10.1067/mjd.1994.8077502

2. Fradin C, Hube B. Tissue infection and site-specific gene expression in Candida albicans. Adv Appl Microbiol 2003; 53:271-90; PMID:14606322; http://dx.doi.org/10.1016/S0065-2164(03)53008-8

3. Booke JD, Garfinkel DJ, Styler GA, 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

4. Le QH, Wright S, Yu Z, Bureau T. Transposon diversity in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 2000; 97:7576-81; PMID:10861007; http://dx.doi.org/10.1073/pnas.97.13.7576

5. Li WH, Gu Z, Wang H, Nekrutenko A. Evolutionary analyses of the human genome. Nature 2001; 409:847-8; PMID:11237011; http://dx.doi.org/10.1038/35057062

6. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408:796-814; PMID:11130711; http://dx.doi.org/10.1038/35057039

7. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000; 408:796-814; PMID:11130711; http://dx.doi.org/10.1038/35057039

8. 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

9. SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. The paleontology of intergene retrotransposons of maize. Nat Genet 1998; 20:43-5; PMID:9753528; http://dx.doi.org/10.1038/1695

10. Santana MF, Barista AD, Ribeiro LE, de Araujo EF, de Queiroz MV. Terminal repeat retrotransposons as DNA markers in fungi. J Basic Microbiol 2013; Forthcoming; PMID:23440766; http://dx.doi.org/10.1002/jobm.201200453

11. Xiong Y, Eckbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 1990; 9:3353-62; PMID:1689655

12. Eckbush TH, Jamburuthugoda VK. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res 2008; 134:21-34; PMID:18261821; http://dx.doi.org/10.1016/j.virusres.2007.12.010

13. Neveügloë C, Feldmann H, Bon E, Gaillardin C, Casarrega S. Genomic evolution of the long terminal repeat retrotransposons in *hemiascomycetes* yeasts. Genome Res 2002; 12:930-43; PMID:12045146; http://dx.doi.org/10.1094/gr.2002

14. Carr M, Benassou D, Bergman CM. Evolutionary genomics of transposable elements in *Sarcobasomyces cerevisiae*. PLoS One 2012; 7:e50978; PMID:22326439; http://dx.doi.org/10.1371/journal.pone.0050978

15. Goodwin TJ, Poulter RT. 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

16. Hudson NJ, Goodwin TJ, Butler MJ, Poulter RT. An active retrotransposon in *Candida albicans*. Nucleic Acids Res 2001; 29:4014-24; PMID:11574684

17. Matthews GD, Goodwin TJ, Butler MJ, Berryman TA, Poulter RT. pCal, a highly unusual *Ty1 copia* retrotransposon from the pathogenic yeast *Candida albicans*. J Bacteriol 1997; 179:718-28; PMID:9371461

18. Uen AG, Koel J, Berns A, van Lohuizen M. Retroviral inserional mutagenesis: past, present and future. Oncogene 2005; 24:7565-72; PMID:16299927; http://dx.doi.org/10.1038/sj.onc.1209043

19. Plant EP, Goodwin TJ, Poulter RT. Tea5, a *Ty5*-like retrotransposon from *Candida albicans*. Yeast 2000; 16:1509-18; PMID:11119373; http://dx.doi.org/10.1002/1097-0061(200012)16:16<1509::AID-TEA638-3.0.CO;2-R

20. Perreau VM, Santos MA, Tuite M, Betta, a novel repetitive DNA element associated with tRNA genes in the pathogenic yeast *Candida albicans*. Mol Microbiol 1997; 25:229-36; PMID:9282735; http://dx.doi.org/10.1046/j.1365-2958.1997.4521816.x
