SHORT COMMUNICATION

Consistent association between *beta* DNA element and transfer RNA genes in *Candida albicans*

Ekachai Chukeatirote

Abstract

Chukeatirote, E.
Consistent association between *beta* DNA element and transfer RNA genes in *Candida albicans*
Songklanakarin J. Sci. Technol., 2004, 26(3) : 403-409

The database of *Candida albicans* genome was screened to investigate the relationship between tRNA genes and retrotransposons. The results showed that, among 34 distinct families of LTR retrotransposons identified, only the *beta* element appeared to preferentially integrate near the tRNA genes. Specific site selection of the *beta* LTRs may have arisen from a coevolution process in which the host's genome used to control their transposition. It is possible that, due to this coevolution mechanism, integration of the *beta* elements become mutualistic, although not proven, to the *C. albicans* genome by acting as a modulator in tRNA gene transcription.

Key words : *beta*, LTR retrotransposon, tRNA gene, *Candida albicans* genome

Ph.D. (Biochemistry), Department of Biotechnology, School of Science, Mae Fah Luang University, Chiang Rai, 57100 Thailand.
Corresponding e-mail : ekachai@mfu.ac.th
Received, 31 October 2003 Accepted, 9 February 2004

บทคัดย่อ

เอกชัย ชูเกียรติโรจน์

ความสัมพันธ์ของการแทรกตำแหน่งอย่างเฉพาะเจาะจงของ *beta* DNA element และ tRNA genes ในยีสต์ *Candida albicans*

ว. สงขลานครินทร์ วทท. 2547 26(3) : 403-409

งานวิจัยนี้ได้ศึกษาถึงความสัมพันธ์ระหว่าง tRNA genes กับ retrotransposons จากฐานข้อมูลของ *Candida albicans* genome จากผลการทดลองที่ได้พบว่ามีเพียง *beta* element เท่านั้นที่มีการแทรกตำแหน่งอยู่ใกล้กับ tRNA genes การแทรกตำแหน่งอย่างเฉพาะเจาะจงของ *beta* LTRs นั้นอาจจะเกิดขึ้นจากกระบวนการอันซับซ้อนซึ่งสิ่งมีชีวิตใช้ในการควบคุมการเคลื่อนย้ายตำแหน่งของ LTR retrotransposons และเกิดวิวัฒนาการร่วมกับ host genome โดยส่งผล ให้การแทรกตำแหน่งของ *beta* element นั้นมีความสำคัญต่อการอยู่รอดของ host อีกด้วย ผู้วิจัยคาดว่าการมีอยู่ของ *beta* อาจจะมีส่วนช่วยในการควบคุมกระบวนการ transcription ของ tRNA genes ซึ่งจะต้องทำการทดลองต่อไปเพื่อ ทดสอบสมมติฐานดังกล่าว

ภาควิชาเทคโนโลยีชีวภาพ สำนักวิชาวิทยาศาสตร์ มหาวิทยาลัยแม่ฟ้าหลวง อำเภอเมือง จังหวัดเชียงราย 57100

Retrotransposons (or retroelements) are a class of mobile genetic elements capable of transposition *via* RNA intermediates (Boeke *et al*., 1985). In eukaryotic genomes, these retrotransposon elements are abundant and ubiquitous. For instance, the human LINE1 retrotransposons make up ~17% of the human genome (Kazazian and Moran, 1998) while more than 50% of the genome of maize (*Zea mays*) are identified as these elements (SanMiguel *et al*., 1996). All retroelements exhibit two unique features: i) ability to transpose within the organism's genome and ii) ability to replicate their copy number autonomously. Therefore, due to these properties, the retrotransposons are expected to play a key role in genomic rearrangements and thus yield an important consequence in genetic evolution.

LTR elements are the most numerous classes of retrotransposons. They generally contain two long terminal repeat sequences (LTRs), typically ranging from 250 – 600bp in length, flanking a protein-coding product with structural homology to retroviral *gag*- and *pol*-proteins (Boeke, 1989). However, the LTR retrotransposons do not have the *env* gene of the retroviruses. These LTR retrotransposons can be divided into two groups based upon the order of their protein-coding domains of the *pol*-gene unit and referred to as the *gypsy*-like and *copia*-like retrotransposons. Full-length LTR retrotransposons seem to be unstable and, for most cases, they are found in single, isolated form, the so-called "solo LTRs". The existence of the solo elements is believed to occur from recombination between their two LTRs with subsequent loss of the internal sequence. Both full-length retrotransposons and the solo LTRs are often found flanked by the short direct repeats (4 – 5bp), a characteristic footprint of retrotransposons (Roeder and Fink, 1980).

Retrotransposons can theoretically insert into many different sites in the organism's genome and, as such, are a common cause of genetic variation. Even worse, such a transposition may lead to a deleterious effect on host survival if some crucial genes were disrupted and thus nonfunctional. Surprisingly, the magnitude and successful dispersal of retrotransposons in the organism's genome seem not to affect the fitness of the host, however. Detailed analyses have shown that most retrotransposons exhibit a target site selectivity within the genome suggesting the non-random behaviour of the retroelement insertions (Zou and Voytas, 1997; Kim *et al*., 1998; Singleton and Levin, 2002). For example, all the Ty (Ty, *t*ransposon in *y*east) elements designated as Ty1 to Ty5 of the budding yeast *Saccharomyces cerevisiae*

have a strong bias for target site selection in the genome. Ty1, Ty2, Ty3, and Ty4 elements are found within 750bp upstream of tRNA genes or other genes transcribed by RNA polymerase III (Kim *et al*., 1998) whereas Ty5 targets the telomeres and mating type loci (Zou *et al*., 1996; Zou and Voytas, 1997). In addition, two retrotransposons, namely DRE and Tdd3, in a soil-living amoeba *Dictyostelium discoideum* also have the target site preference to the tRNA genes (Marschalek *et al*., 1992; Marschalek *et al*., 1993). The mechanisms that these elements choose to integrate at the specific regions, however, remain unanswered.

Recently, more than 30 distinct groups of retrotransposons have been identified and characterised in the human pathogenic yeast *Candida albicans* (Goodwin and Poulter, 2000). These LTR retrotransposon population found in *C. albicans* seem to differ considerably from Ty elements of *S. cerevisiae* in many aspects such as the number of distinct retrotransposon families and the copy number of retrotransposons. In addition, the majority of *C. albicans* LTR retrotransposons appear to be inactive, i.e., they are not able to move from one place to another in the genome. In this study, the relationship between LTRs and tRNA genes in *C. albicans* was examined. A wide-genome survey in *C. albicans* indicates that only the *beta* DNA element appears to preferentially integrate near the tRNA genes.

Experimental approach

Sequence data for *C. albicans* strain SC5314 was obtained from the Stanford DNA Sequencing and Technology Centre available at the following website: http://www-sequence.stanford.edu/group/candida. Sequencing of *C. albicans* was accomplished with the support of the NIDR and the Burroughs Wellcome Fund.

Goodwin and Poulter (2000) have recently identified thirty-four distinct families of the LTR retrotransposons from *C. albicans*. The complete set of these LTR sequences, deposited in GenBank, were retrieved according to their accession numbers and were submitted individually as a query in a BLASTN search (Altschul *et al*., 1997) of the Stanford *C. albicans* genome database. The resulting contigs derived from the search were then determined if they contained a tRNA gene using the tRNAscan-SE software version 1.21 available at the following website, http://www.genetics.wustl.edu/eddy/tRNAscan-SE (Lowe and Eddy, 1997). Please note that the results presented in this study were studied until September 2003.

The contigs containing LTRs and tRNA genes were given the reference numbers (see Table 1) that is identical to those assigned by the Stanford DNA Sequencing and Technology Centre. However, it should be noted that these ID numbers are subject to change when the *C. albicans* genome is complete.

Results and Discussion

The genome of *C. albicans* is presently being sequenced by the Stanford DNA Sequencing and Technology Centre. The data, obtained from the website http://www-sequence.stanford.edu/group/candida, were used to study if there is a specific pattern of the integration of the LTRs (i.e., near the tRNA genes) in *C. albicans*. In this study, the *C. albicans* genome was screened by performing sequential BLAST searches of individual LTR sequences against the database. By using the BLASTN programme (Altschul *et al*., 1997), a group of contigs that were varied in numbers depending on the LTR template used were retrieved. The presence of the tRNA genes within these contigs was subsequently confirmed by use of the tRNAscan-SE (Lowe and Eddy, 1997).

Among 34 distinct LTR retrotransposons investigated, only eight families were found associated with the tRNA genes (see Table 1). Considering that *C. albicans* is closely related to *S. cerevisiae* (Lloyd and Sharp, 1992), the results obtained were quite surprising since only a minority group of *C. albicans* LTRs was tRNA gene-related. In *S. cerevisiae*, all Ty1 – Ty4 elements (especially Ty1 and Ty3) preferentially insert within 750bp upstream of the tRNA genes (Kim *et al*., 1998). Besides, for most *C. albicans* tRNA gene-associated LTRs, the insertion bias

Table 1. The tRNA gene-related LTR retrotransposons in *Candida albicans*

LTR element[a]	Contig number[b]	tRNA gene(s)[c]	Distance[e]
alpha (10)	6-2474	Pro-tRNA^{TGG} (i)	808
beta (10)	6-2165	Asp-tRNA^{GTC}	36
	6-2041	Asp-tRNA^{GTC}	35
	6-2409	Val-tRNA^{AAC}	12
	6-2419	Glu-tRNA^{TTC} (i)	13
	6-2302	Arg-tRNA^{TCT}	14
	6-1799	Thr-tRNA^{AGT}	12
	6-1986	Gly-tRNA^{GCC}	32
	6-2469	Ala-tRNA^{AGC}	13
zeta (19)	6-2510	Phe-tRNA^{GAA} (i)[d]	—[f]
	6-2510	Val-tRNA^{TAC} (i)	669
	6-2254	His-tRNA^{GTG} (i)	617
upsilon (6)	6-2510	Phe-tRNA^{GAA} (i)[d]	499
psi (30)	6-2418	Ser-tRNA^{AGA}	78
kahu (17)	6-1837	Pro-tRNA^{TGG} (i)	53
rho (13)	6-2413	Leu-tRNA^{CAA} (i)	671
phi (14)	6-1695	Leu-tRNA^{TAA} (i)	251

[a] Of thirty-four distinct LTR retrotransposons being identified in *C. albicans*, only those associated with the tRNA genes are presented. The copy numbers of the LTRs, which are estimated from the *C. albicans* genomic database (Goodwin and Poulter, 2000), are also shown in parentheses.
[b] Reference numbers of the contigs used are identical to those assigned by the Stanford DNA Sequencing and Technology Centre.
[c] (i) indicates the presence of intron sequences.
[d] The Phe-tRNA^{GAA} gene identified from *zeta* and *upsilon* from the contig 6-2510 was, in fact, identical according to the sequence analysis (see text for details).
[e] Numbers of nucleotides are illustrated to show a distance between the tRNA genes and the LTR elements. The distance includes the short direct repeats (if any).
[f] The 3'-region of the Phe-tRNA^{GAA} gene is shown to be the initial part of the *zeta* element.

for tRNA genes was very low considering from the tRNA gene copy number(s) in relation to the copy numbers of the LTR templates. The target bias for these tRNA gene-related LTRs were as follows: 10% of *alpha* (1/10 insertions), 80% of *beta* (8/10 insertions), 15.79% of *zeta* (3/19 insertions), 3.33% of *psi* (1/30 insertions), 5.88% of *kahu* (1/17 insertions), 7.69% of *rho* (1/13 insertions), 16.67% of *upsilon* (1/6 insertions), and 7.14% of *phi* (1/14 insertions) (see Figure 1). Of these relationships, only the *beta* LTR exhibited a strong bias to integrate near the tRNA genes (80%, see Figure 1). In contrast, other *C. albicans* tRNA gene-related LTRs did not show a preference for inserting at this particular site since only one tRNA gene could be retrieved for each family of these LTRs (except *zeta* in which three tRNA genes were identified).

The genomic organisation of these LTRs was then subject to further analysis regarding their location with respect to the tRNA genes. For most cases, the tRNA gene-associated LTRs (except *upsilon*) were found upstream of the tRNA genes. However, there were differences in nucle-

otide distance between the LTRs and the tRNA genes which varied considerably ranging from 12bp of *beta* to 808bp of *alpha* (see Table 1). In the case of Phe-tRNA^{GAA} gene (contig 6-2510) identified from *zeta*, its sequence was found overlapping in which the 3'-region of the Phe-tRNA^{GAA} gene was used as the initial site of the *zeta* DNA element. Furthermore, two Phe-tRNA^{GAA} genes yielded from *zeta* and *upsilon* were, in fact, the same gene based on their structural gene (see Figure 2).

In the budding yeast *S. cerevisiae*, a close association between Ty elements and tRNA genes has been well established (Chalker and Sandmeyer, 1992; Ji *et al*., 1993; Kim *et al*., 1998). However, such a relationship is still unclear in *C. albicans*. According to the results presented here, it seemed that most *C. albicans* LTRs did not have the target site preference (i.e., near the tRNA genes).

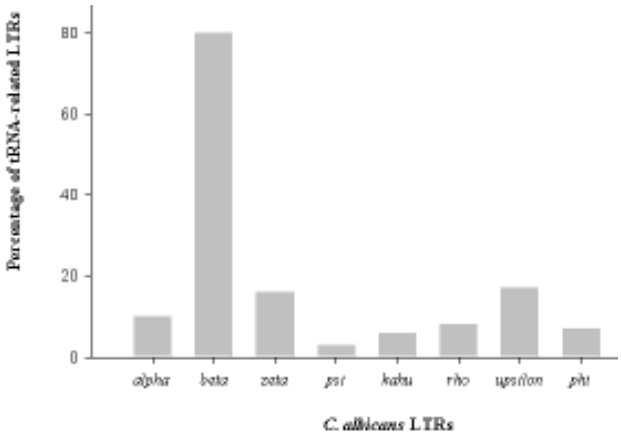

Figure 1. Inconsistent distribution of the tRNA gene-associated LTRs in *C. albicans*. Only *beta* DNA element appeared to be a hot spot harbouring the tRNA genes (see text for details).

Besides, most tRNA gene-related LTRs appeared to be less specific to the integration position in terms of their copy numbers (in which a few tRNA genes could be identified) and the nucleotide distance. The only element that was apparent in target bias is *beta* as previously described (Perreau *et al*., 1997; Chukeatirote, 2002). This element, like Ty3, was always found immediately upstream of the tRNA genes within a range of 12 – 36bp. Although, this preliminary study revealed that most LTRs in *C. albicans* did not select the tRNA genes as a specific integration, further study should be undertaken to determine if there would be any other specific sites for LTRs' integration. Due to the unfinished state of the *C. albicans* genome project, distribution of the *C. albicans* LTRs remained unclear. Undoubtedly, the organisation of the tRNA genes and LTR retrotransposons will be better understood with the complete nucleotide sequence of the *C. albicans* genome.

Retrotransposons are abundant within eukaryotic genomes. Considering the potential threat that transposition may have caused (i.e., by disrupting the essential genes), there must be a mechanism used by an organism's genome to minimise these deleterious mutations. Several studies show that the retrotransposon's integration is not random and usually displays a strong bias for specific regions (Zou *et al*., 1996; Winckler *et al*., 1998; Kim *et al*., 1998). To date, it has been pointed out that the tRNA genes are likely to be the primary determination, albeit not universal, for retrotransposon's insertion. The mechanism of how the LTRs select the tRNA genes as their preference is unknown, however.

Until recently, it was believed that transcription of the tRNA genes was solely under control of DNA sequences internal to the transcribed

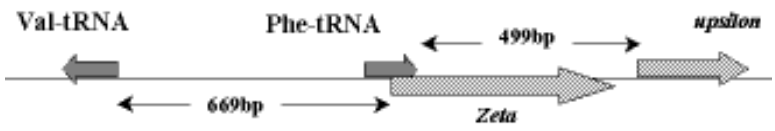

Figure 2. Structural gene of the contig 6-2510. There was only one Phe-tRNA^{GAA} gene. As shown, the Phe-tRNA^{GAA} gene could be retrieved from both *zeta* and *upsilon* elements.

region, the so-called "A and B boxes". However, there is growing evidence that external DNA sequences, i.e., upstream and downstream sequences, were required for full transcriptional activity (Sprague, 1995; Huang and Maraia, 2001). The upstream regions of the tRNA genes, in particular, are required for RNA pol III by recruiting the formation of pol III transcription factor TFIIIB. TFIIIB, a multiprotein complex, plays a major role in RNA pol III-mediated transcription initiation by binding with an upstream AT-rich region (Kassavetis *et al*., 1995; Grove *et al*., 1999). The *beta* DNA element found adjacent upstream of the tRNA genes in *C*. *albicans* is AT-rich (more than 65%) suggesting its possible role in tRNA gene expression. Such mechanisms are likely to develop during the process of genomic evolution allowing retrotransposons to be maintained within an organism's genome.

Acknowledgements

Sequence data for *C*. *albicans* was retrieved from the Stanford DNA Sequencing and Technology Centre website at http://www-sequence.stanford.edu/group/candida. Sequencing of *C*. *albicans* was accomplished with the support of the NIDR and the Burroughs Wellcome Fund. This work was also supported by an NSTDA grant (CO-B-06-2E-67-401).

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25: 3389-402.

Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. 1985. Ty elements transpose through an RNA intermediate. Cell, 40: 491-500.

Boeke, J.D. 1989. Transposable elements in *Saccharomyces cerevisiae*. **In**: Mobile DNA (Ed. by D.E. Berg and M.M. Howe), pp 335-374. Washington, DC: ASM Press.

Chalker, D.L. and Sandmeyer, S.B. 1992. Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. 6: 117-128.

Chukeatirote, E. 2002. Identification of transfer RNA (tRNA) genes from *Candida albicans* genome using *beta* element: a bioinformatics approach. Chiang Mai J. Sci., 29: 173-181.

Goodwin, T.J.D. and Poulter, R.T.M. 2000. Multiple LTR-retrotransposon families in the asexual yeast *Candida albicans*. Genome Res., 10: 174-191.

Grove, A., Kassavetis, G.A., Johnson, T.E. and Geiduschek, E.P. 1999. The RNA polymerase III-recruiting factor TFIIIB induces a DNA bend between the TATA box and the transcriptional start site. J. Mol. Biol., 285: 1429-1440.

Huang, Y. and Maraia, R.J. 2001. Comparison of the RNA polymerase III transcription machinery in *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* and human. Nucleic Acids Res., 29: 2675-2690.

Ji, H., Moore, D.P., Blomberg, M.A., Braiterman, L.T., Voytas, D.F., Natsoulis, G. and Boeke, J.D. 1993. Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. Cell, 73: 1007-1018.

Kassavetis, G.A., Nguyen, S.T., Kobayashi, R., Kumar, A., Geiduschek, E.P. and Pisano, M. 1995. Cloning, expression and function of TFC5, the gene encoding the B'' component of *Saccharomyces cerevisiae* RNA polymerase III transcription factor TFIIIB. Proc. Natl. Acad. Sci. USA, 92: 9786-9790.

Kazazian, H.H. Jr. and Moran, J.V. 1998. The impact of L1 retrotransposons on the human genome. Nat. Genet., 19: 19-24.

Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A. and Voytas, D.F. 1998. Transposable elements and genome organisation: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res., 8: 464-478.

Lloyd, A.T. and Sharp, P.M. 1992. Evolution of codon usage patterns: the extent and nature of diver-

gence between *Candida albicans* and *Saccharomyces cerevisiae*. Nucleic Acids Res., 20: 5289-5295.

Lowe, T.M. and Eddy, S.R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res., 25: 955-964.

Marschalek, R., Hofmann, J., Schumann, G., Gosseringer, R. and Dingerman, T. 1992. Structure of DRE, a retrotransposable element which integrates with position specificity upstream of *Dictyostelium discoideum* tRNA genes. Mol. Cell. Biol., 12: 229-239.

Marschalek, R., Hofmann, J., Schumann, G., Bach, M. and Dingerman, T. 1993. Different organisation of the transfer RNA-gene-associated repetitive element, DRE, in NC4-derived strains and in other wild-type *Dictyostelium discoideum* strains. Eur. J. Biochem., 217: 627-631.

Perreau, V.M., Santos, M.A.S. and Tuite, M.F. 1997. *beta*, a novel repetitive DNA element associated with tRNA genes in the pathogenic yeast *Candida albicans*. Mol. Microbiol., 25: 229-236.

Roeder, G.S. and Fink, G.R. 1980. DNA rearrangements associated with a transposable element in yeast. Cell, 21: 239-249.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z. and Bennetzen, J.L. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science, 274: 765-768.

Singleton, T.L. and Levin, H.L. 2002. A long terminal retrotransposon of fission yeast has strong preferences for specific sites of insertion. Eukaryot. Cell, 1: 44-55.

Sprague, K. 1995. Transcription of eukaryotic tRNA genes. **In**: tRNA: Structure, Biosynthesis, and Function (Ed. by D. Soll and U.L. RajBhandary), pp 31-50. Washington DC: ASM Press.

Winckler, T., Tschepke, C., de Hostos, E.L., Jendretzke, A. and Dingermann, T. 1998. Tdd-3, a tRNA gene-associated poly(A) retrotransposon from *Dictyostelium discoiderm*. Mol. Gen. Genet., 257: 655-661.

Zou, S., Ke, N., Kim, J.M. and Voytas, D.F. 1996. The *Saccharomyces* retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev., 10: 634-645.

Zou, S. and Voytas, D.F. 1997. Silent chromatin determines target preference of the retrotransposon Ty5. Proc. Natl. Acad. Sci. USA, 94: 7412-7416.