論 文 の 内 容 の 要 旨

論文題目

Studies on molecular mechanisms underlying retrotransposition of site-specific non-LTR retrotransposons in R1 clade

（標的特異的なR1クレードnon-LTRレトロトランスポゾンの転移機構の研究）

氏　　名　　那日蘇

INTRODUCTION:

Non-long terminal repeat (non-LTR) retrotransposons, also known as long interspersed nuclear elements (LINEs), are the most abundant mobile elements in many organisms. LINEs comprise ~21% of the human genome and are the only active transposable elements that influence the human genome through their involvement in genome evolution, genome mutation, and disease etiology. However, detailed processes for the retrotransposition of LINEs are still largely unknown. Previous studies have revealed that LINE moves using a unique system called Target Primed Reverse Transcription (TPRT), which is peculiar to LINE (Eickbush, 1993). After transcription, translation and the ribonucleoprotein (RNP) formation of LINE, the RNP is transferred into nuclei and starts TPRT on the target site of chromosomes (Fig. 1). In the initial step of TPRT, the endonuclease domain (EN) encoded in LINE nicks one strand of the target DNA and creates a 3′-hydroxyl end, which is used as a primer for reverse transcription. Thereafter, the LINE mRNA is reverse transcribed by the RT domain into cDNA at the target site. Although most non-LTR retrotransposons, including the human L1 element, are randomly inserted throughout the host genome, some elements (site-specific elements) are inserted into specific sites of repetitive genomic sequences, such as telomeric repeats, ribosomal DNA (rDNA) and microsatellites (Kojima & Fujiwara, 2004). Different to random integration, the site-specific retrotransposition into multiple copied genes can avoid the damage to the host genome. On the basis of structural and phylogenetic features, non-

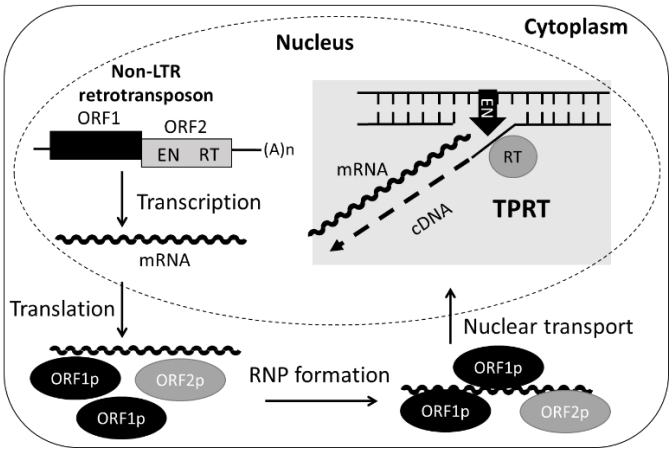

Fig.1. Life cycle and Target Primed Reverse Transcription (TPRT) of non-LTR retrotransposon.

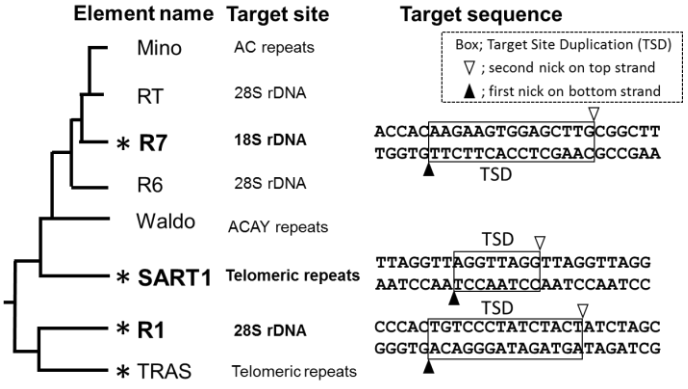

Fig.2. Site-specific non-LTR retrotransposons in R1-clade.* ; element which has the retrotransposition assay system established before.

LTR retrotransposons are classified into at least 16 clades. In R1-clade, it is reported that many site specific elements which target different chromosomal sites are known (Fig. 2).

In this study, I focus on site-specific elements in R1-clade, because their retrotransposition assays have been basically established before (Fujiwara, 2015) (Fig. 2). Since the R1-clade elements have the distinct DNA targets, comparative studies of their systems will answer the remaining main questions for retrotransposition of LINE; (1) how LINE recognizes the target DNA (Chapter 1), (2) how the LINE mRNA is reverse transcribed at the correct site (Chapter 2) and (3) how open reading frame (ORF) proteins in RNP of LINE access to the target site (Chapter 3). R1-clade elements encode two ORFs, ORF1 and ORF2 which includes EN and RT domains. In addition, most R1-clade elements usually end with a poly(A) tail and have target-site duplication (TSD) at both ends, which is caused by the repair process of sequence gaps between the first nick on bottom strand and the second nick on top strand of the target DNA (Fig. 2).

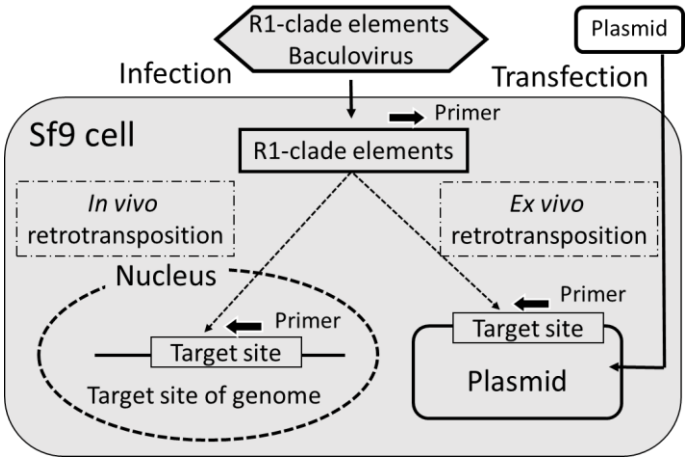

Fig.3. *In vivo* and *Ex vivo* retrotransposition assay in Sf9 cell.

By comparing the data among different site-specific R1-clade elements, I tried to draw the overall picture for each process of the retrotransposition. For comparison, I selected three site-specific elements, *Bombyx mori* SART1 (SART1Bm) which targets $(TTAGG)_n$ telomeric repeats , R7 of *Anopheles gambiae* (R7Ag) which targets 18S rDNA and *Bombyx mori* R1 (R1Bm) which targets 28S rDNA (Fig. 2). Since some data for R1Bm have been already reported (Anzai et al. 2005, Maita et al. 2007), I here mainly used SART1Bm and R7Ag to analyze the molecular mechanism underlying respective processes described above.

RESULTS AND DISCUSSION:

1. How is the target site DNA recognized by site-specific R1-clade elements?

Previous studies have shown that the primary determinant of sequence-specific integration of R1-clade elements is the EN domain (Anzai et al. 2001). However, how EN recognizes the target DNA has not been understood well. To analyze the specific recognition of target by EN efficiently, I here used a new *ex vivo* assay system (Fig. 3). In this system, instead of the chromosomal target of each R1-clade element in the genome, an exogenous target sequence on the plasmid is transfected into *Spodoptera frugiperda* 9 (Sf9) cells and used for the

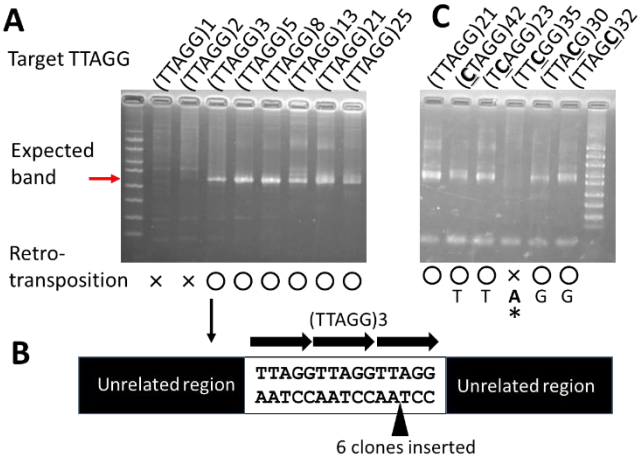

Fig.4. Telomeric repeats are recognized strictly by SART1Bm EN.

retrotransposition assay. This enables to change the target sequence easily. After infection of the recombinant baculovirus for each R1-clade element, the retrotransposition event was analyzed by PCR to amplify the 3′-junction region between the target DNA and integrated R1-clade element (Fig. 3).Using the above method, I first tested how SART1Bm recognizes telomeric repeats $(TTAGG)_n$ and found that at least three repeats were required for retrotransposition of the element (Fig. 4A). In the $(TTAGG)_3$ target, however, the retrotransposition of SART1Bm occurred only in the third (TTAGG) tract (6 clones identified) but not in the first and second (TTAGG) tract (Fig. 4B). Combining with other results, I speculate that 12bp upstream sequence and 3bp downstream sequence of the first nicking site is important for the recognition of SART1Bm EN. Furthermore, experiments using target plasmids having point-mutations (N to C replacement) on each TTAGG tract showed that only of A to C mutant abolished the retrotransposition activity, although other changes did not affect (Fig. 4C). This result suggests that the A residue of TTAGG is essential for recognition of the SART1Bm EN domain. I conclude that SART1Bm EN recognizes less than 15bp around the initial cleavage site, which has an analogy to the previous report of another telomeric repeat specific R1-clade element TRAS1 (Anzai et al. 2001).

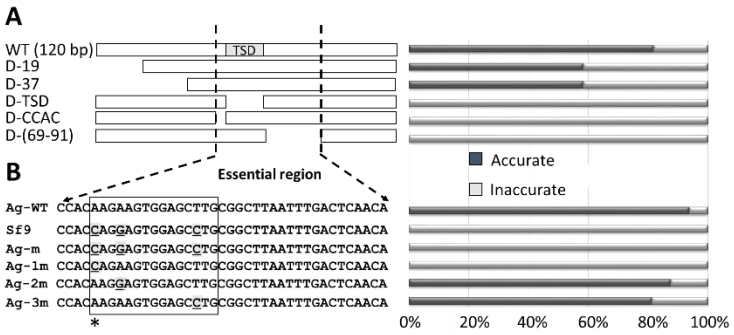

Fig.5. The first nucleotide A in TSD is critical for the accurate insertion of R7Ag.

For comparison, I also tested R7Ag using the same assay system, and found that of 120 bp of 18S rDNA target site, the central region of 39bp (essential region), was necessary for the R7Ag accurate insertion (Fig. 5A). I noticed that R7Ag did not integrate into the target site of *Spodoptera* genome accurately, and that there are 3 nucleotides differences within R7Ag TSD in 18SrDNA between *Anopheles* and *Spodoptera*. I studied the importance of these nucleotides, and found that the first nucleotide in TSD (A in *Ag*, C in *Sf*) is critical for the accurate insertion of R7Ag (Fig. 5B). The previous report using purified R1Bm EN also showed that the first nucleotide in TSD is essential for introducing a specific nick (Maita et al., 2007). Even though structural variations of EN and target sequences among R1-clade elements, telomere specific elements (TRAS and SART) and rDNA specific elements (R1 and R7) show the similar features for the recognition of target sequences, respectively.

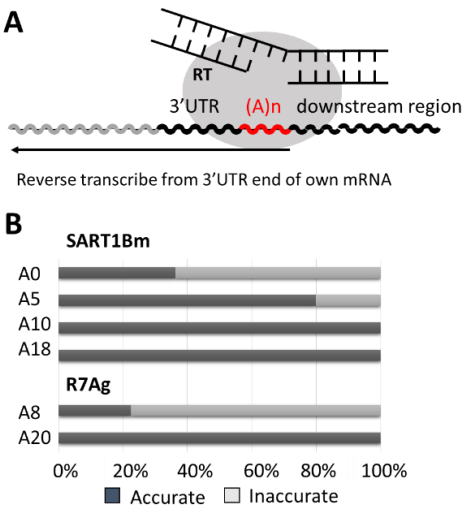

Fig.6. Long poly(A) tail is critical for accurate reverse transcription of SART1Bm and R7Ag.

2. Long poly(A) tail is critical for accurate reverse transcription of SART1Bm and R7Ag.

Non-LTR retrotransposons are often transcribed into downstream regions, but even such a read-though mRNA is usually reverse transcribed from the accurate 3′-end of own template mRNA. However, the mechanism underlying the accurate reverse transcription is still unclear. Notably, most LINEs including R1-clade

elements end with a poly(A) tail, which is suggested to be involved in this mechanism. To clarify this hypothesis, I generated several constructs for SART1Bm and R7Ag having various lengths of poly(A) tail, and conducted *in vivo* retrotransposition assay. In this systems, it is noteworthy that R1-clade elements are also read-thorough the vector region (Fig. 6A). In the host genome, the average poly(A) length for SART1Bm and R7Ag is estimated to be 20bp and 13.5bp, respectively. The results of *in vivo* assay showed that reduction of poly(A) length in constructs increased the inaccurate reverse transcription of mRNA from other regions in both elements (Fig. 6B). In this case, the reverse transcription occurred sometimes from poly(A) tail of the extreme end of read-through product, not from the poly(A) tail of 3′-end (data not shown). In both elements of SART1Bm and R7Ag, however, longer poly(A) tail similar to the genomic copy of the element showed reverse transcription from the accurate site (Fig. 6B). I speculate that the RT domain of most LINE recognizes both poly(A) tail and specific structures within 3′ UTR of own mRNA and starts the accurate reverse transcription.

3. <u>Subcellular localization of ORFp suggests the access of R1-clade elements to each target site.</u>

During retrotransposition, site-specific elements such as SART1Bm and R7Ag are assumed to approach respective target regions in nucleus, telomeric region (TR) which includes telomeric repeats and nucleolar organizer (NO) which includes rDNA cluster, respectively (Fig. 7B).

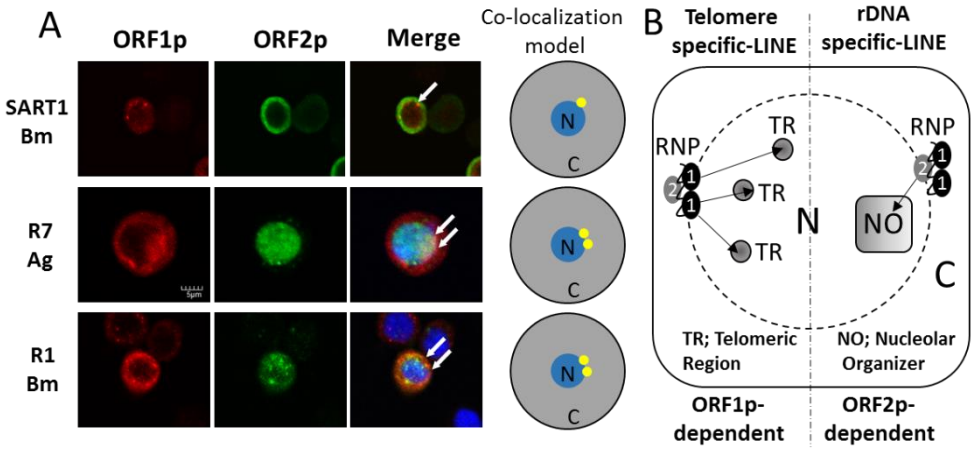

Fig.7. Localization and access model of ORF proteins of R1-clade elements
N; Nucleus, C; Cytoplasm. White arrows; co-localization signals.

However, there is a few evidences to support this idea. To clarify the above hypothesis, I examined the subcellular organization of ORF1 protein (ORF1p) and ORF2p of SART1Bm, R7Ag and R1Bm, respectively by immunofluorescent staining (Fig. 7A). In SART1Bm, when both proteins were co-expressed, ORF1p was localized in nucleus but ORF2p was localized predominantly in cytoplasm. Co-localization signals of ORF1p and ORF2p were observed weakly in peripheral region of nuclear membrane and as some dotted patterns within nucleus. Although further evidences are necessary, the dotted regions are predicted to be telomeric regions. In R7Ag and R1Bm, in contrast, ORF1p was localized predominantly in cytoplasm but ORF2p was localized in nucleus. The ORF2p signals corresponded partly with the localization signals for nucleolar marker fibrillarin (data not shown). Furthermore, ORF1p and ORF2p of R7Ag and R1Bm were co-localized largely in the nuclear periphery and to a lesser extent within the nucleus. These results suggest that telomeric repeat-specific element SART1Bm accesses the telomeric region in an ORF1p-dependent manner, but that rDNA-specific elements R7Ag and R1Bm access nucleolar region in an ORF2p-dependent manner (Fig. 7B).