博士論文（要約）

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)

平成 29 年 2 月　博士（生命科学）

東京大学大学院・新領域創成科学研究科

先端生命科学専攻・遺伝システム革新学分野

那日蘇

CONTENTS

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

Abbreviations (in alphabetical order)

AcNPV, *autographa californica* nuclear polyhedrosis virus

$(A)_n$, $poly(A)_n$

APE, apurinic/apyrimidinic endonuclease

Bac-to-Bac, bacteria to baculovirus

DR, direct repeat

EN, endonuclease

Gag, group-specific antigen (capsid proteins)

HA, hemagglutinin influenza virus epitope

His, histidine-tag

INT, integrase domain

ITR, inverted terminal repeat

LINE, long interspersed nuclear element

LTR, long terminal repeat

MOI, multiplicity of infection

NO, nucleolar organizer

ORF, open reading frame

ORF1p, open reading frame 1 protein

ORF2p, open reading frame 2 protein

PCR, polymerase chain reaction

Pol, polymerase

rDNA, ribosomal DNA

RLE, restriction enzyme-like endonuclease

RT, reverse transcriptase

SART1, inversion of TRAS1 (telomeric repeat associated sequence 1)

Sf9 Cell, *Spodoptera frugiperda* 9 cell

TE, transposable element

TPRT, target primed reverse transcription

TR, telomeric region

TRAS1, telomeric repeat associated sequence 1

TSD, target site duplication

UTR, untranslated region

ZK, zinc knuckle domain

General introduction

Transposons, also called transposable elements (TEs) or jumping genes, are segments of DNA that can move from one position to another on chromosomes inside a single cell. Almost all living things from bacteria to plants, animals, and even archaea are reported to harbor transposons (Levin and Moran, 2011). In human cells, for example, about half of the genome is occupied with various types of TEs. For a long time, it has been supposed that TEs are involved in various biological processes and phenomena such as hybrid dysgenesis, somatic mutations, heterochromatin formation, genome formation/rearrangement, gene expression, gene silencing and so on (Huang et al., 2012; Burt and Triverse, 2008).

Based on the transposition mechanism, transposons are classified into two major groups: Class II elements, or DNA transposons, which move by a ‘‘cut and paste’’ manner (Fig. 0-1A) and Class I elements, or retrotransposons, which move by a ‘‘copy and paste’’ manner (Fig. 0-1B). Retrotransposons are further subdivided into two groups according to whether or not they contain long terminal repeat (LTR) at both ends: LTR retrotransposons and non-LTR retrotransposons (Fig. 0-2). In addition, non-LTR retrotransposons are sub-classified as either long interspersed nuclear elements (LINEs) or short interspersed nuclear elements (SINEs) depending on whether they encode proteins to mediate their own mobility (autonomous elements) or rely upon proteins encoded by other elements (nonautonomous elements). SINEs are nonautonomous elements that do not encode proteins and as a consequence require LINEs for their propagation; thus non-LTR retrotransposons are also known as LINEs (Goodier and Kazazian, 2008). In this study, non-LTR retrotransposons will be designated as LINEs. It is noteworthy that the proportions and types of transposons in the genome vary from one

organism to another. For example, in baker's yeast, 100% of transposons are LTR retrotransposons, compared to 91.4 % of DNA transposons in nematode worms (Kazazian, 2011).

DNA transposons are commonly used for genetic engineering as vectors for delivering recombinant genes into target sites. Examples of such transposons include Sleeping Beauty (Ivics et al., 1997) and piggyBac (Ding et al., 2005), and their transposition mechanisms have been studied well in general. The retrotransposition process of LTR retrotransposon has been also understood largely, because they behave like retroviruses, such as human immunodeficiency virus (HIV) (Roth, 2000). LTR retrotransposons reverse transcribe their own mRNA into double stranded cDNA in the cytoplasm, which then moves into the nucleus and inserts into the target site (Havecker et al., 2004). Importantly, in both DNA transposons and LTR retrotransposons, the transposed DNA is integrated into the host genome by a recombinase activity in the final step of transposition, which is critically different to non-LTR retrotransposon, as described below.

Non-LTR retrotransposons or LINEs, are the most abundant mobile elements in many organisms. In particular, LINEs comprise ~about 21% of the human genome and are the only active TE that influence the human genome through their involvement in genome evolution, genome mutation, and disease etiology (Beck et al., 2011). However, detailed processes for the retrotransposition mechanism of LINEs are still largely unknown, compared to DNA transposons and LTR retrotransposons. This is probably because they have a very unique and different transposition system with that of LTR retrotransposons and retroviruses, and they constitute many groups having several different structures.

According to a structural feature, LINEs are categorized into two groups: the early-branched group with a single open reading frame (ORF) and a recently-branched group,

with two ORFs (ORF1 and ORF2) (Malik et al., 1999; Kojima and Fujiwara, 2003). There are two essential domains for retrotransposition of LINE. One is an endonuclease (EN) domain that cuts the target site DNA (Feng et al., 1996). Another is a reverse transcriptase (RT) domain that is responsible for reverse transcription of the RNA template (Malik et al., 1999). Both domains are encoded in the single ORF of early-branched element or in the ORF2 of recently-branched element, respectively (Fig. 0-3). The ORF1 of recently-branched element is involved in multimerization of the ORF1 and ORF2 proteins (Matsumoto et al., 2006), nucleic acid chaperone activity (Dawson et al., 1997), RNA binding activity (Hohjoh and Singer, 1996; Matsumoto et al., 2006), and formation of the ribonucleoprotein (RNP) complex (Dawson et al., 1997; Kulpa and Moran, 2005).Thus, ORF1 is indispensable for the retrotransposition of LINE, although the functional role of ORF1 is less understood than that of ORF2.

Persistent studies of an early-branched group element R2 by Eickbush's group have revealed a unique system called target primed reverse transcription (TPRT) mechanism, which is peculiar to LINE (Fig. 0-4) (Luan et al., 1993). During the TPRT process in the last step of retrotransposition of LINE, the EN domain nicks the bottom strand of the target DNA, and the RT domain uses the 3'-hydroxyl end of the nicked DNA as a primer for reverse transcription of the mRNA template into cDNA. It is noteworthy, therefore, that the target sequence of a LINE is not only used as the insertion site but also serves as a primer to initiate reverse transcription. However, how the target site is recognized, how the reverse transcription is initiated, and how these processes are controlled are still unclear. At the final step of TPRT, it is suggested that the top strand in the target site is cleaved and complementary DNA is synthesized, although further studies are also required to confirm this step.

Compared to R2 element, however, the retrotransposition mechanism of recently-branched element, has not been studied so much. After transcription of the element from the host genome, its mRNA encoding both ORFs is exported to the cytoplasm and translated into two proteins, ORF1p and ORF2p (Fig. 0-3). The resulting proteins associate with their own mRNAs in the cytoplasm to form an RNP complex and subsequently translocate to the nucleus (Moran et al., 1996; Matsumoto et al., 2006). Since they have two ORFs, which is different from R2, they should have a specialized system to regulate the translation of two proteins, the formation of the RNP, and the translocation of the RNP into the target site. However, the detailed mechanisms of respective processes are not fully studied.

There are two structural hallmarks commonly observed among many LINEs, which are involved in or caused by retrotransposition events. One hallmark is short direct repeat sequences called target site duplication (TSD) shown in both ends of the element. TSD is caused by the repair process of sequence gaps between the first nick on the bottom strand and the second cleavage on the top strand of the target DNA (Fig. 0-4) (Moran et al., 1996). Another hallmark is poly(A) tail at the 3' end of many LINEs. It is noteworthy that this poly(A) tail is not added after transcription as like in cellular mRNA, but originated from the genomic copy of the element. Thus, the poly(A) tail in the 3' end is thought to be involved in the initiation step of reverse transcription, but there is no detailed study to certify the idea.

Although most LINEs, including the human L1 element, are randomly inserted throughout the host genome, some elements (site-specific elements) are inserted into specific sites of repetitive genome sequences (Fujiwara, 2015), such as ribosomal DNA (rDNA) (Kojima and Fujiwara, 2003), telomeric repeats (Fujiwara et al., 2005), and

microsatellites (Busseau et al., 2001). In contrast to random integration which sometimes disrupts the essential gene of the host, site-specific retrotransposition into multiple copied genes can avoid damage to the host genome, which may be a kind of symbiotic strategy. Because site-specific LINEs copy themselves into a defined target DNA sequence, their insertion sites can be easily detected by PCR which makes them as a good model to study the retrotransposition mechanism. Most of early-branched clades (5 clades) include site-specific LINEs. In contrast, among at least 20 recently-branched clades of LINEs, it is known that only two clades, R1 and Tx1, include site specific elements (Kojima and Fujiwara, 2004). Especially, the R1 clade is reported to have many site-specific elements targeting different chromosomal sites (Fig. 0-5), which makes it interesting subject to understand the evolutionary history of site-specific retrotransposition.

Some R1-clade elements have distinct target specificities. The Mino and Waldo insert into short repeat sequences of $(AC)_n$ and $(ACAY)_n$; R1, R6 and RT insert into specific sites in 28S rDNA; R7 inserts into 18S rDNA; and TRAS and SART insert into different sites within $(TTAGG)_n$ telomeric repeats (Fig. 0-5). The phylogenetic relationship of the R1-clade elements revealed that they diverged from the same ancestral element and that their site-specificity might have changed during evolution (Kojima and Fujiwara, 2003; Fujiwara, 2015). Importantly, our group has established *in vivo* retrotransposition assay for several R1-clade elements (Fig. 0-5, shown in asterisks), using a baculovirus infection system and PCR detection of site-specific integration, which enables to analyze the detailed mechanisms of retrotransposition and to clarify the evolutionary process of target site specificity. Actually, in many recently-branched LINEs, it is usually difficult to analyze the detailed retrotransposition mechanism, because they rarely and randomly transpose into the genome, so that *de novo* insertion can be hardly detected. This is one

reason why I focus on the R1-clade elements to answer several questions for LINEs as described above.

Using the above method, our group has revealed some molecular processes of R1-clade elements retrotransposition: The primary determinant of site-specific integration in R1 clade-elements is the EN domain (Anzai et al., 2001; Takahashi and Fujiwara, 2002; Maita et al., 2007; Osanai-Futahashi and Fujiwara, 2011). Site-specificity also relies on base pairing between the read-through (downstream region) mRNA product of LINEs and target DNA sequence at the cleaved site (Osanai et al., 2004; Anzai et al., 2005). Some telomere-specific LINEs seem to localize at the telomeric region (TR) of the nucleus (Matsumoto et al., 2004). Even with these achievements, however, several questions remain to be answered. What nucleotides of target DNA are recognized by the EN? Is there a consensus rule for the target-site recognition of the EN among R1-clade elements? How does the reverse transcription initiate? Is there some difference of the translocation process of the RNP complex between R1-clade elements which target the different chromosomal sites such as telomere or rDNA? Further studies answer these questions and elucidate the detailed molecular mechanism of LINE retrotransposition not only for R1-clade elements but also for other elements.

Since the R1-clade elements have distinct DNA targets, comparative studies of their retrotransposition mechanisms will answer the remaining main questions for retrotransposition of LINEs: (1) how LINEs recognize the target DNA (Chapter I), (2) how LINE mRNA is reverse transcribed at the correct site (Chapter II), and (3) how open reading frame (ORF) proteins in the RNP of LINE access the target site (Chapter III).

In this study, I compared the data between different site-specific R1-clade elements and tried to draw the overall picture for each process of retrotransposition. I selected three

site-specific elements for comparison: *Bombyx mori* SART1 (SART1Bm), which targets $(TTAGG)_n$ telomeric repeats; *Anopheles gambiae* R7 (R7Ag), which targets 18S rDNA; and *Bombyx mori* R1 (R1Bm), which targets 28S rDNA (Fig. 0-5, shown in bold). 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 mechanisms underlying the processes described above. In Chapter I, I investigate the target site recognition mechanism and report my finding that SART1Bm and R7Ag specifically require their exact host target site for accurate insertion. In Chapter II, I investigate the role of the poly(A) tail in retrotransposition and report that the long poly(A) tail at the end of LINEs is involved in the accurate initiation of reverse transcription in both SART1Bm and R7Ag elements. In Chapter III, I investigate subcellular localization of ORFs and report that the telomere-specific element, SART1Bm and the rDNA-specific elements R7Ag and R1Bm approach target sites in an ORF1p-dependent and ORF2p-dependent manner, respectively.

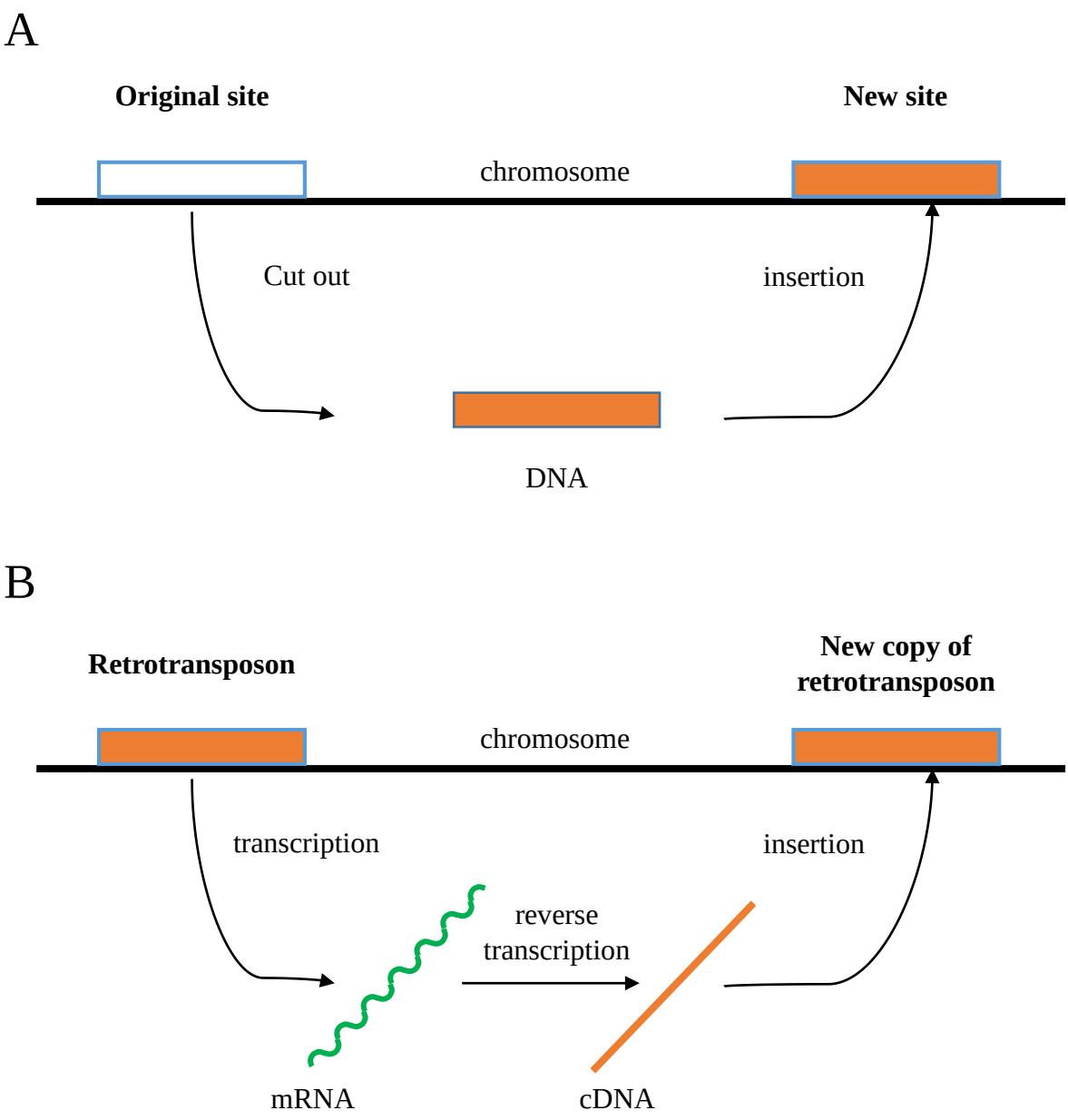

Fig. 0-1. Transposons are classified into two classes according to the transposition mechanism. (A) DNA transposons move by DNA intermediate ‘‘cut and paste’’ transposition mechanism in which DNA transposons are removed from original site and inserted into a new target site. **(B)** Retrotransposons amplify themselves through RNA intermediate ‘‘copy and paste’’ transposition mechanism, a process termed retrotransposition in which retrotransposons are transcribed from the genome and then reverse transcribed into cDNA that integrates back to a new target site.

Retrotransposons

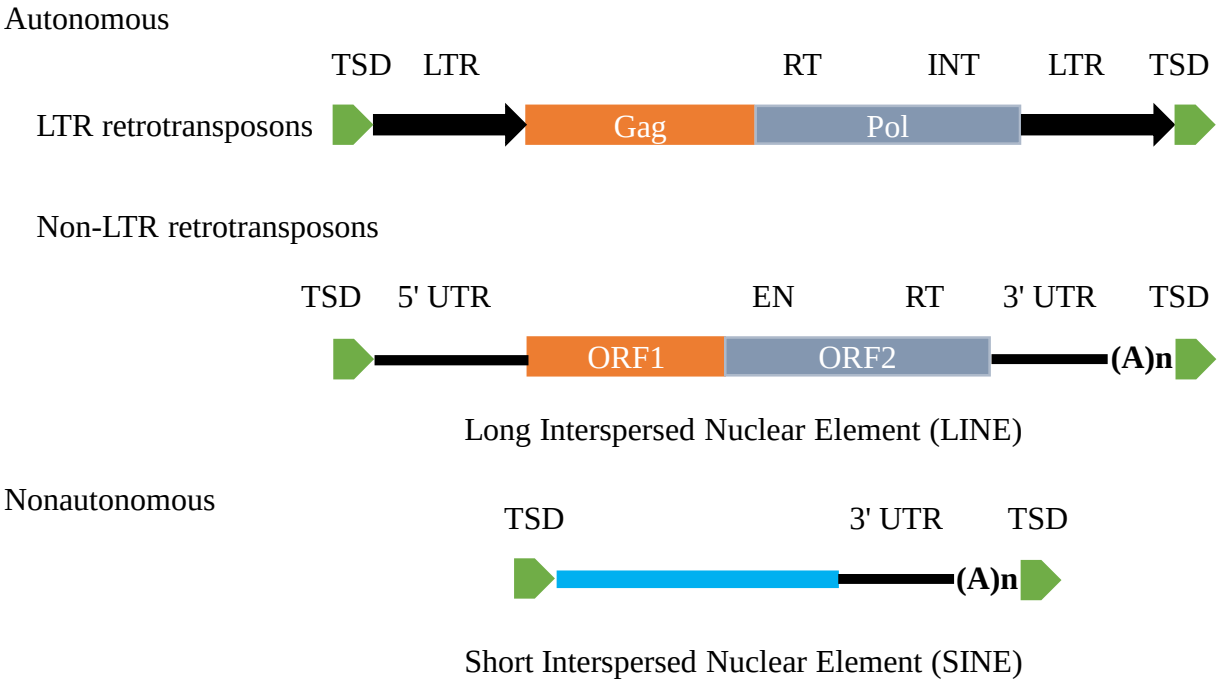

Fig. 0-2. Classification and structure of retrotransposons. LTR retrotransposons contain two long terminal repeats (LTRs; gray arrows) and encode Gag and Pol proteins. Pol contains reverse transcriptase (RT) and integrase (INT) activities, all of which are crucial for retrotransposition. LTR retrotransposon integrases create staggered cuts at the target sites, resulting in target site duplication (TSD) (green arrowheads). Non-LTR retrotransposons lack LTR and encode one or two open reading frames (ORFs) which contain two essential catalytic domains, endonuclease (EN) and reverse transcriptase (RT). Non-LTR retrotransposons are sub-classified into two major groups of long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). LINEs and SINEs end with poly(A) tail and are flanked by TSD. SINEs are nonautonomous retrotransposons which do not encode any proteins, and rely on LINE-encoded proteins for transposition.

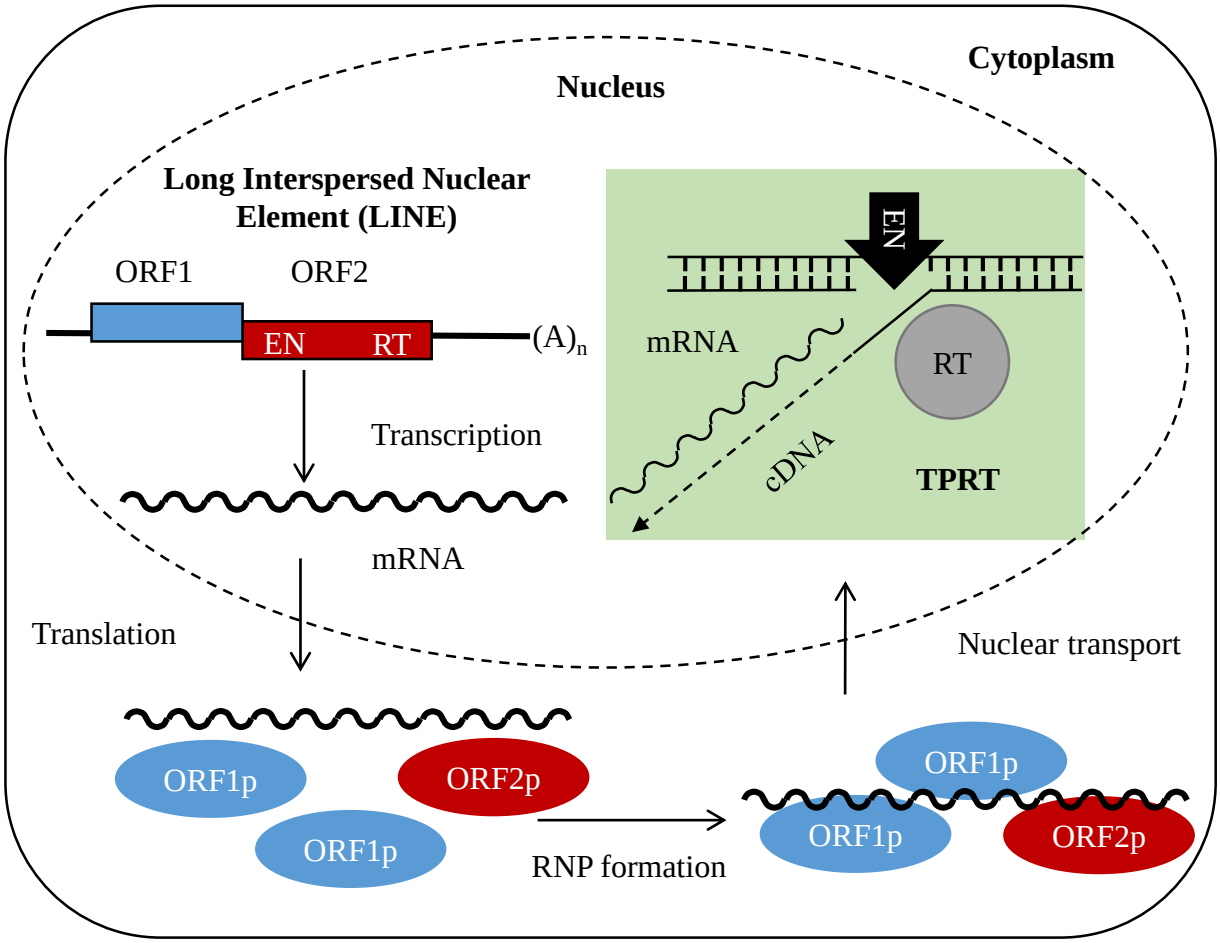

Fig. 0-3. Life cycle of recently-branched LINE. Recently-branched LINE encodes two ORFs followed by 3' UTR ends with poly(A) tail. ORF1 encodes domains involved in binding own mRNA and interacting with ORF2. ORF2 encodes two main domains of EN domain which cuts the target site and RT domain which reverse transcribes own mRNA into cDNA. The retrotransposition process occurs as follows: after transcription from the host genome, the LINE mRNA is exported to the cytoplasm and is translated into two proteins, ORF1p (blue oval) and ORF2p (red oval). The resulting proteins associate with their mRNA in the cytoplasm to form a ribonucleoprotein (RNP) complex and subsequently the RNP complex is translocated to the nucleus. In the last step, LINE inserts into genome through target primed reverse transcription (TPRT) mechanism.

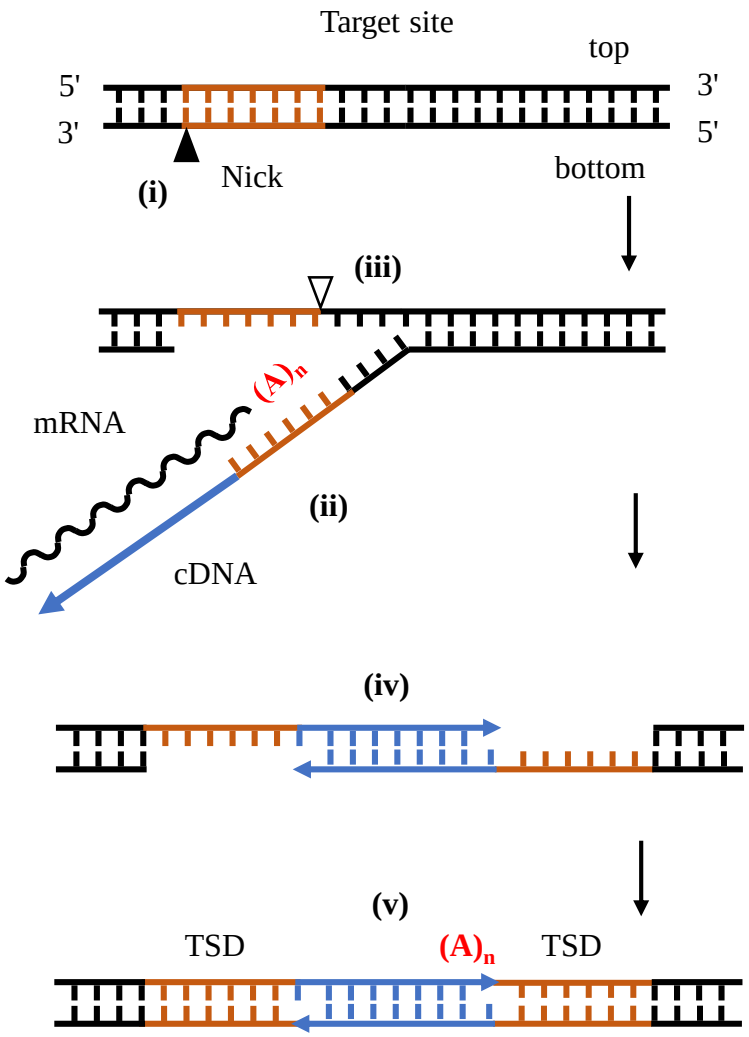

Fig. 0-4. Target primed reverse transcription (TPRT) is a unique mechanism for LINEs. (i) The EN domain nicks the bottom strand of genomic DNA to expose target site. **(ii)** The liberated target site is then used as a primer to reverse transcribe own mRNA into cDNA by the RT domain. **(iii)** The EN domain cleaves the top strand of genomic DNA to produce a staggered break. **(iv)** Complementary DNA strand of LINE is synthesized and genomic DNA ends are repaired. **(v)** A new copy of LINE ends with poly(A) tail and is flanked by target site duplication (TSD).

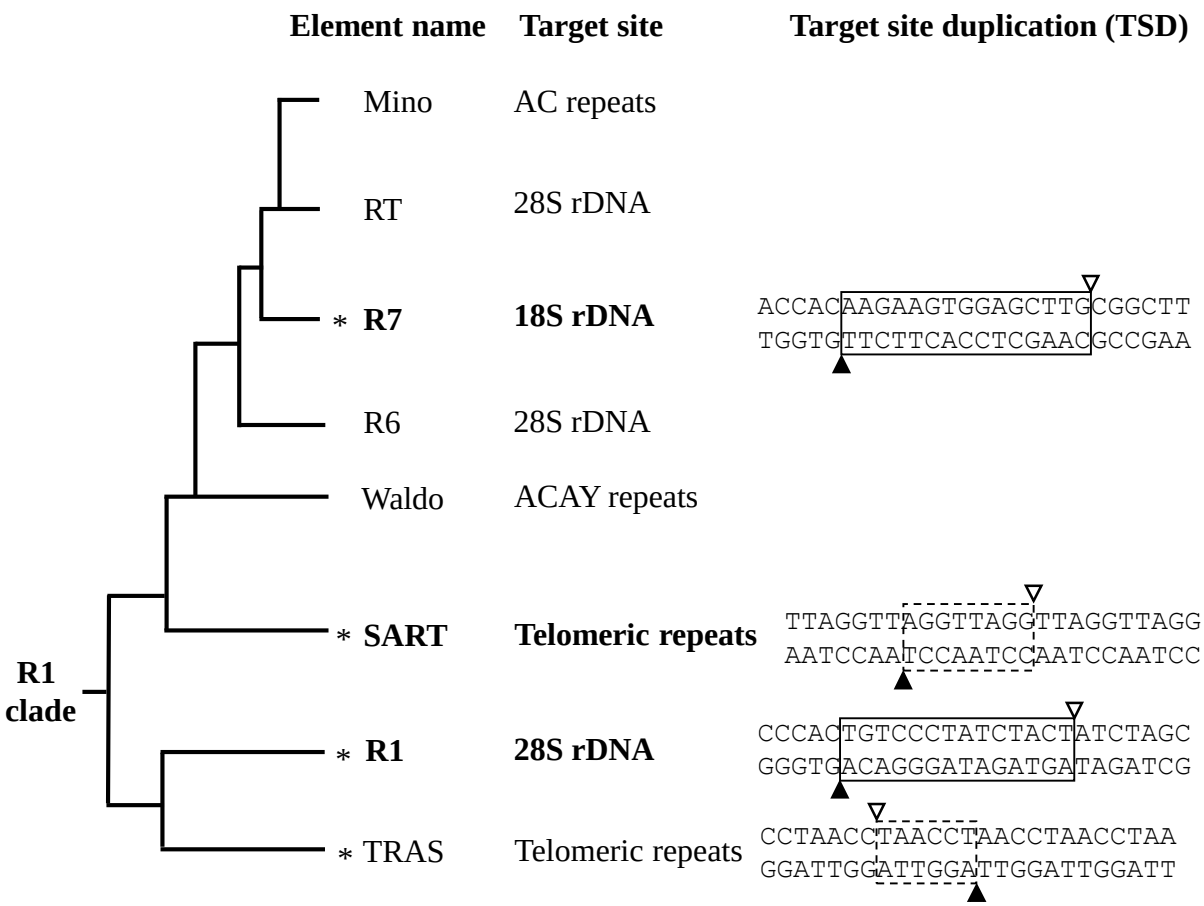

Fig. 0-5. Site-specific LINEs in R1 clade have distinct targets. The phylogeny is revised from data in Fujiwara, (2015). Asterisks (*) indicate R1-clade elements which have the *in vivo* retrotransposition assays established before. The elements used in this study are shown in bold. Target and insertion sequences of each element are shown on the right. R7 and R1 elements target 18S rDNA and 28S rDNA, respectively. Target site duplication (TSD) is indicated by box. SART and TRAS elements target the same $(TTAGG/CCATT)_n$ telomeric repeats with opposite orientations. Broken boxes indicate putative TSD because they target the tandem repeats TTAGG and the exact cleavage sites have not been identified. Solid arrowhead indicates first nick on the bottom strand. Open arrowhead indicates second cleavage on the top strand. In the TPRT process, the bottom strand is firstly nicked and then the top strand is cleaved.

References

Anzai T, Osanai M, Hamada M, Fujiwara H. 2005. Functional roles of 3'-terminal structures of template RNA during in vivo retrotransposition of non-LTR retrotransposon, R1Bm. Nucleic Acids Res. **33**:1993–2002.

Anzai T, Takahashi H, Fujiwara H. 2001. Sequence-specific recognition and cleavage of telomeric repeat (TTAGG)n by endonuclease of non-long terminal repeat retrotransposon TRAS1. Mol Cell Biol. **21**:100–108.

Beck CR, Garcia-Perez JL, Badge RM, Moran JV. 2011. LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet. **12**:187–215.

Burt A, Trivers R. 2008. Genes in Conflict. Harvard University Press, Garden Street, Cambridge.

Busseau I, Berezikov E, Bucheton A. 2001. Identification of Waldo-A and Waldo-B, two closely related non-LTR retrotransposons in Drosophila. Mol Biol Evol. **18**:196–205.

Dawson A, Hartswood E, Paterson T, Finnegan DJ. 1997. A LINE-like transposable element in Drosophila, the I factor, encodes a protein with properties similar to those of retroviral nucleocapsids. EMBO J. **16**:4448-4455.

Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T. 2005. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell. **122**:473–483.

Feng Q, Moran JV, Kazazian HH Jr, Boeke JD.1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell. **87**:905-916.

Fujiwara H. 2015. Site-specific non-LTR retrotransposons. Microbiol Spectr. **3**:MDNA3-0001-2014.

Fujiwara H, Osanai M, Matsumoto T, Kojima KK. 2005. Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, Bombyx mori. Chromosom Res. **13**:455–467.

Goodier JL, Kazazian HH Jr. 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell. **135**:23–35.

Havecker ER, Gao X, Voytas DF. 2004. The diversity of LTR retrotransposons. Genome Biol. **5**:225.

Hohjoh H, Singer MF. 1996. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA.

EMBO J. **15**:630–639.

Huang CR, Burns KH, Boeke JD. 2012. Active transposition in genomes. Annu Rev Genet. **46**:651-675.

Ivics Z, Hackett PB, Plasterk RH, Izsvák Z. 1997. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell. **91**:501–510.

Kazazian HH Jr. 2011. Mobile DNA, 1st ed. FT Press, Upper Saddle River, New Jersey.

Kojima KK, Fujiwara H. 2003. Evolution of target specificity in R1 clade non-LTR retrotransposons. Mol Biol Evol. **20**:351–361.

Kojima KK, Fujiwara H. 2004. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. Mol Biol Evol. **21**:207-217.

Kulpa DA, Moran JV. 2005. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. Hum Mol Genet. **14**:3237–3248.

Levin HL, Moran JV. 2011. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet. **12**:615–627.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. **72**:595–605.

Maita N, Aoyagi H, Osanai M, Shirakawa M, Fujiwara H. 2007. Characterization of the sequence specificity of the R1Bm endonuclease domain by structural and biochemical studies. Nucleic Acids Res. **35**:3918–3927.

Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol. **16**:793–805.

Matsumoto T, Hamada M, Osanai M, Fujiwara H. 2006. Essential domains for ribonucleoprotein complex formation required for retrotransposition of telomere-specific non-long terminal repeat retrotransposon SART1. Mol Cell Biol. **26**:5168–5179.

Matsumoto T, Takahashi H, Fujiwara H. 2004. Targeted nuclear import of open reading frame 1 protein is required for in vivo retrotransposition of a telomere-specific non-long terminal repeat retrotransposon, SART1. Mol Cell Biol.

24:105–122.

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr. 1996. High frequency retrotransposition in cultured mammalian cells. Cell. **87**:917–927.

Osanai-Futahashi M, Fujiwara H. 2011. Coevolution of telomeric repeats and telomeric repeat-specific non-LTR retrotransposons in insects. Mol Biol Evol. **28**:2983–2986.

Osanai M, Takahashi H, Kojima KK, Hamada M, Fujiwara H. 2004. Essential motifs in the 3' untranslated region required for retrotransposition and the precise start of reverse transcription in non-long-terminal-repeat retrotransposon. Mol Cell Biol. **24**:7902–7913.

Roth JF. 2000. The yeast Ty virus-like particles. Yeast. **16**:785–795.

Takahashi H, Fujiwara H. 2002. Transplantation of target site specificity by swapping the endonuclease domains of two LINEs. EMBO J. **21**:408–417.

Chapter III

Subcellular localization mechanisms of telomere-specific and rDNA-specific non-LTR retrotransposons (LINEs)

Abstract

During retrotransposition, site-specific elements such as SART1Bm and R7Ag are assumed to approach their respective target regions in the nucleus, telomeric region (TR) which includes telomeric repeats, and nucleolar organizer (NO) which includes rDNA cluster, respectively. However, there is little evidence to support this idea. To investigate this hypothesis, I examined the subcellular organization of ORF1 protein (ORF1p) and ORF2p of SART1Bm, R7Ag, and R1Bm by immunofluorescent staining. In SART1Bm, ORF1p localized in nucleus, but ORF2p localized predominantly in cytoplasm. Co-localization signals of ORF1p and ORF2p of SART1Bm were observed weakly in the periphery of the nuclear membrane and as some dotted patterns within the nucleus. Although further evidence is needed, dotted co-localization of ORF proteins of SART1Bm in the nucleus are predicted to be TR. In contrast, in R7Ag and R1Bm, ORF1p localized predominantly in the cytoplasm but ORF2p localized in the nucleus. The ORF2p signals corresponded partly with signals for the nucleolar marker fibrillarin. Furthermore, ORF1p and ORF2p of R7Ag and R1Bm co-localized largely in the nuclear periphery and to a lesser extent within the nucleus. These results suggest that telomere-specific element SART1Bm accesses the TR in an ORF1p-dependent manner, but that rDNA-specific elements R7Ag and R1Bm access the NO in an ORF2p-dependent manner.

Introduction

Chapters I and II mainly focus on the TPRT mechanism. However, before TPRT initiation, the RNP complex of LINEs needs to move from the cytoplasm to their target regions in the nucleus. Thus, migration of the RNP complex into the nucleus is an essential step for retrotransposition. However, the mechanism by which the RNP of LINEs crosses the nuclear membrane remains unclear. Some previous studies showed that the entrance of the RNP complex into the nucleus occurs when the nuclear membrane breaks down and is cell cycle-dependent. It is reported cell division increase the retrotransposition efficiency of human L1 element (Shi et al., 2007; Xie et al., 2013). On the other hand, L1 ORF1p, ORF2p and mRNA mainly co-localized with stress granules in the cytoplasm and it is speculated that preventing the RNP complex from entering the nucleus may be one of the mechanisms that cells use to protect their genomes from parasitic invaders (Goodier et al., 2007; Doucet et al., 2010).

In contrast, another study showed that L1 retrotransposition occurs in non-dividing somatic cells, which suggests nuclear import of the L1 RNP (Kubo et al., 2006). In the silkworm SART1Bm, nuclear localization signals (NLSs) in the N-terminal region of ORF1 are found to be involved in RNP transport into the nucleus (Ishibashi, 2004; Matsumoto et al., 2004). Furthermore, two *Drosophila* telomere-specific LINEs, HeT-A and TART, are delivered to their telomeric target sites by the HeT-A gag protein, a finding that strongly supports a nuclear transport mechanism (Rashkova et al., 2002). Site-specific elements recognize repetitive sequences as precise locations in the genome for insertion which considered to be a symbiotic strategy, allowing insertion into the genome with very little damage to the host. Thus I hypothesized that site-specific elements actively transport their RNP complex from the cytoplasm to target sites in the nucleus.

In this study, to elucidate the process whereby RNPs access their targets, I examined the subcellular organization of ORF proteins of telomere-specific element SART1Bm and rDNA-specific elements R7Ag and R1Bm. Former biochemical and genetic studies have revealed that SART1Bm and L1 of human and mouse ORF2 are translated by an unconventional mechanism (Kojima et al, 2005; Alisch et al., 2006). LINE ORF2p is hypothesized to be either translated at lower levels or less stable than ORF1p, which might help explain why it is hard to detect. This problem has been setback for studying the subcellular localization of ORF2p of some LINEs to date. Here, I used the HA and 3xFLAG epitope tagging systems to visualize ORF1p and ORF2p proteins, respectively. Using this strategy, I could easily assay ORF2p subcellular localization in three contexts, either alone, when co-expressed with ORF1p, or when co-expressed with the nucleolar marker fibrillarin. I also established a novel retrotransposition assay system for SART1Bm termed “plasmid-based *in vivo* retrotransposition assay”. Results from this assay indirectly support that the RNPs of SART1Bm access the TR for retrotransposition.

Materials and Methods

Sf9 Cell culture.

See chapter I Materials and Methods section.

Constructs for the immunofluorescence assay.

R7AgORF1-pIZT/V5-HA/His-dEGFP was constructed as follows: a portion of R7AgORF1 was amplified by PCR from the R7Ag WT plasmid with the primers EcoRI-KOZAK-R71-F and NotI-R71-R. The resulting PCR product was subcloned between the *Eco*RI and *Not*I sites of pIZT/V5-HA/His-dEGFP. To construct R1BmORF1-pIZT/V5-HA/His-dEGFP, a portion of R1BmORF1 was amplified by PCR from the pAcGHLT-B R1ORFs+3′ UTR plasmid (accession number AB182560) (Anzai et al., 2005) using the primers SpeI-KOZAK-R11-F and EcoRI-R11-R-2. The resulting PCR product was subcloned between the *Spe*I and *Eco*RI sites of pIZT/V5-HA/His-dEGFP. To construct SART1BmORF1-pIZT/V5-HA/His-dEGFP, a portion of SART1BmORF1 was amplified by PCR from the SART1WT-pAcGHLTB plasmid (Takahashi and Fujiwara, 2002) using the primers EcoRI-KOZAK-S1-S880-Ser and NotI-S1-A3016(FL). The resulting PCR product was subcloned between the *Eco*RI and *Not*I sites of pIZT/V5-HA/His-dEGFP.

3xFLAG-R7AgORF2-pIZT/V5-His-dEGFP was constructed as follows: a portion of the R7AgORF2/3′UTR was amplified by PCR from the R7Ag WT plasmid with the primers SpeI-R7O2S and EcoRI-R7O2A+3′ UTR. The PCR product was then subcloned between the *Spe*I and *Eco*RI sites of 3xFLAG-pIZT/V5-His-dEGFP. To construct 3xFLAG-R1BmORF2-pIZT/V5-His-dEGFP, a portion of the R1BmORF2/3′ UTR was amplified by PCR from the pAcGHLT-B -R1ORFs/3′ UTR plasmid (accession number AB182560) (Anzai et al., 2005) with the primers BamHI-R1O2S and R1 A5136 (NotI).

The resulting PCR product was subcloned between the *Bam*HI and *Not*I sites of 3XFLAG-pIZT/V5-His-dEGFP. To construct 3xFLAG-SART1BmORF2-pIZT/V5-His-dEGFP, a portion of the SART1BmORF2/3′ UTR was amplified by PCR from the SART1WT-pAcGHLTB plasmid (Takahashi and Fujiwara, 2002) with the primers S1ORF2-BamHI-S3018 and SART1 A6704 NotI-vert. The resulting PCR product was subcloned between the *Bam*HI and *Not*I sites of 3XFLAG-pIZT/V5-His-dEGFP.

Immunofluorescence assay.

Sf9 cells were grown on 12-well plates (Iwaki, Fukushima, Japan) and were transfected with 2 μg of plasmid DNA in the presence of 12 μl of TransFast Reagent (Promega).The transfected cells were harvested at 72 h after transfection, allowed to adhere to an EZ SLIDE 8 well glass slide (Millicell; Millipore, Billerica, MA, USA), and then fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 15 min, followed by ice-cold methanol for 5 min. Next, cells were blocked by incubating for 30 min at 37°C in 3% BSA + PBS-Tween 20 (T). The cells were then incubated with two primary antibodies alone or in a mixture: anti-HA rabbit polyclonal antibody (A190-108A, 1:2,000; Bethyl Laboratories, Montgomery, TX, USA) and anti-FLAG M2 monoclonal antibody (F1804, 1:400; Sigma-Aldrich, St. Louis, MO, USA) in 1% BSA in PBST for 1 h at room temperature. Cells were washed three times with 1x PBS (5 min per wash) and subsequently incubated in the dark for 1 h at room temperature with a mixture of the following fluorescently-labeled secondary antibodies in 1% BSA: Alexa Fluor® 488-labeled anti-mouse IgG (1:500; Abcam) and Alexa Fluor® 555 anti-rabbit IgG (1:1000; Abcam). For fibrillarin detection, an anti-fibrillarin nucleolar marker rabbit polyclonal antibody (ab5821, 1:200; Abcam) was used as a primary antibody, and Alexa Fluor® 555

anti-rabbit IgG (1:500, Abcam) was used as the secondary antibody. Finally, the slides were washed with 1x PBS and mounted with Vectashield mounting medium containing 4', 6-diamidino-2-phenylindole (DAPI; Vector Laboratories, Inc., Burlingame, CA, USA). Protein localization was analyzed using a FluoView™ FV1000 confocal microscope (Olympus Corporation, Tokyo, Japan), and images were taken using FV10-ASW 1.6 software (Olympus Corporation).

Western blotting.

Fibrillarin expression in Sf9 cells was detected by western blotting as follows: Sf9 cell extracts were subjected to 10% SDS-polyacrylamide gel electrophoresis. HeLa cell and *Drosophila melanogaster* embryo (Dm embryo 0–18 h) extracts were used as positive controls. Electrophoresed proteins were transferred to a polyvinylidene difluoride membrane (#66543; Pall Corporation, Port Washington, NY, USA) with the Trans-Blot Turbo Blotting System (#170-4155; Bio-Rad) at 0.1 A and 25 V for 60 min. The membrane was subsequently blocked in 5% skim milk in Tris-buffered saline + Tween-20 (TBS-T) overnight at 4°C, washed three times with TBS-T, and incubated overnight at 4°C with a primary antibody against fibrillarin (ab5821; Abcam, Cambridge, UK) or α-tubulin (CLT9002; Cedarlane Laboratories, Burlington, ON, Canada) for 1 h, diluted to 1:10,000 or 1:100,000 in 3% bovine serum albumin (BSA)/Can Get Signal 1 (NKB-101; Toyobo). Following another wash with TBS-T, the membrane was incubated for 1 h with a secondary antibody (anti-mouse IgG NA931 at 1:10,000 or anti-rabbit IgG NA934 at 1:10,000; GE Healthcare, Little Chalfont, UK) diluted in 3% BSA/Can Get Signal 2 (NKB-301; Toyobo). Finally, protein band detection was performed using ImmunoStar LD (290-69904; Wako Pure Chemical Industries, Ltd., Osaka, Japan), and the blot was

imaged on an ImageQuant™ LAS 4000 system (GE Healthcare).

Plasmid-based *in vivo* retrotransposition assay.

Sf9 cells were grown on 12-well plates (Iwaki) and were transfected with 2 μg of plasmid DNA in the presence of 12 μl of TransFast Reagent (Promega). The transfected cells were harvested at 72 h after transfection and the genomic DNA was extracted with Gentra Puregene® Kits (QIAGEN, Valencia, CA, USA). PCR assays were conducted using *Ex-Taq* with 1 μg of Sf9 DNA and the primers S1S6131 and CCTAA6. The reaction mixture was denatured at 96°C for 2 min, followed by 35 cycles of 96°C for 30 s, 62°C for 30 s, and 72°C for 1 min. One microliter of each mixture was subjected to 2% agarose electrophoresis in Tris-acetate-EDTA buffer and visualized by ethidium bromide staining. PCR products were directly cloned into the pGEM-T Easy vector (Promega). The cloned products were sequenced with a BigDye Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) on ABI3130 xl and 3500/3500xl Genetic Analyzers (Applied Biosystems). Sequence analysis was performed using the Vector NTI Advance 10 system (Invitrogen).

Results

Part 1: Subcellular localization of ORF proteins (ORF1p and ORF2p) of telomere-specific LINE SART1Bm.

1-1. SART1Bm ORF2p co-localizes with ORF1p in the nucleus.

To detect SART1Bm ORF proteins localization signals via immunofluorescence, I generated two constructs, in which ORF1p was tagged with HA and ORF2p was tagged with 3xFLAG (Fig. III-1).These plasmids were transfected transiently into Sf9 cells, and the subcellular localization of each expressed protein was detected using a confocal microscope after labeling with antibodies against HA and 3xFLAG. When SART1Bm ORF1p was expressed alone in Sf9 cells, the SART1Bm ORF1p signals were observed as a punctate (dotted) pattern in the nuclei (Fig. III-2-c, Table III-2). Previous study showed that the SART1Bm ORF1p dotted pattern in the nucleus partially co-localizes with the target site of telomeres (Ishibashi, 2008). To test whether ORF2p also localizes to the nucleus, SART1Bm ORF2p was expressed alone in Sf9 cells, and the corresponding signals were observed broadly throughout the cytoplasm (Fig. III-2-f, Table III-2). This result indicates that ORF2p does not enter the nucleus by itself. Next, co-transfection experiment was conducted to test whether ORF2p can co-localize with ORF1p in the nucleus. The results showed that co-localization signals of ORF1p and ORF2p were predominantly observed weakly in the periphery of the nuclear membrane and as some dotted patterns within the nucleus (Fig. III-2-i, Table III-2). This observation indicates that most of the ORF2p does not enter the nucleus, even co-expressed with ORF1p. Although future experiments are needed to determine whether mRNA co-localizes with ORF proteins, I speculate that SART1Bm ORF1p, ORF2p, and mRNA co-localize near to the nuclear membrane followed by the entry of some of these ORF proteins into the

nucleus to yield the dotted patterns seen within the nucleus (Fig. III-2-i, Table III-2).

Part 2: Subcellular localization of ORF proteins (ORF1p and ORF2p) of rDNA-specific LINE R7Ag and R1Bm.

2-1. Subcellular localization of R7Ag and R1Bm ORF proteins showed two types of co-localization.

During retrotransposition, rDNA-specific elements (R elements) are assumed to translocate to the NO, which includes the rDNA cluster, although there is no direct evidence to support this idea to date. To clarify this possibility and the mechanism by which proteins of R elements gain access to the target site, I next attempted to identify the subcellular localization of R7Ag ORF proteins and of the closely related element R1Bm, which targets 28S rDNA. To detect ORF proteins localization signals via immunofluorescence, I generated plasmid constructs for two elements, in which ORF1p was tagged with HA and ORF2p was tagged with 3xFLAG (Fig. III-4A).When R7Ag ORF1p was expressed alone in Sf9 cells, the corresponding signals were observed broadly throughout the cytoplasm (Fig. III-4B-b, Table III-3). Similar cellular localization signals were observed for R1Bm ORF1p (Fig. III-4B-e, Table III-3). In contrast, when ORF2p was expressed alone in Sf9 cells, I observed two types of localization patterns for both R7Ag and R1Bm (Fig. III-5, Table III-3). In type I, the localization signals of both R7Ag (Fig. III-5-b) and R1Bm (Fig. III-5-h) were observed broadly in the cytoplasm. In type II, the R7Ag ORF2p signals were observed as a punctate (dotted) pattern in the nuclei (Fig. III-5-e). Although weak R1Bm ORF2p signals were observed in the cytoplasm, some puncta were also observed in the nuclei (Fig. III-5-k). Next, I co-expressed ORF1p and ORF2p simultaneously to monitor the interaction of the two proteins in cells. In this

experiment, I also observed two types of co-localized signals for both R-elements (Fig. III-6, Table III-4). In type I, some signals for ORF1p (localized broadly in cytoplasm, Fig. III-6-a and-g) and ORF2p (localized mainly in nuclei and weakly in cytoplasm, Fig. III-6-b and-h) co-localized in peripheral regions just outside of the nuclear membrane (Fig. III-6-c and-i). Notably, 59.3% of cells expressing R7Ag proteins and 40% of cells expressing R1Bm proteins corresponded to this type (Table III-4). In type II, fewer ORF1p (localized mainly in cytoplasm with mild nuclear localization, Fig. III-6-d and-j) and ORF2p signals (localized in nuclei, Fig. III-6-e and-k) co-localized in nucleus with punctate pattern (Fig. III-6-f and-l), although the signals for R1Bm were less noticeable. Specifically, 12.5% of cells expressing R7Ag proteins and 15% of cells expressing R1Bm proteins corresponded to this type (Table III-4). The remaining cells expressing both R-elements exhibited no clear co-localization signals for ORF1p and ORF2p (28% for R7Ag and 45% for R1Bm, Table III-4). These results indicate that although most ORF proteins localize separately, some proportion of the ORF proteins of both R-elements co-localize in the cytoplasm and accumulate in the peripheral region of the nuclear membrane. The above results suggest that a small portion of the accumulated proteins of two R elements may translocate into nuclei in an ORF2p-dependent manner.

2-2. R7Ag and R1Bm ORF2p co-localize with the nucleolar marker protein fibrillarin

From the above data, R7Ag and R1Bm ORF2p (or co-localized signals of ORF1p and ORF2p) exhibited a punctate nuclear pattern (Figs. III-5 and III-6). Because the rDNA targets of R elements reside in the nucleolus, these punctuate ORF2p signals might indicate the nucleolar location. To test this possibility, I next used fibrillarin, a nucleolar

marker protein, to confirm the nuclear position of ORF2p. The antibody used herein is known to react with fibrillarin from mouse, human, *Xenopus laevis* and *Drosophila melanogaster* according to the manufacture's manual (see Materials and Methods). Although Xing et al. (Xing et al., 2011) demonstrated that this anti-fibrillarin antibody also reacts with fibrillarin from Sf9 cells, I first confirmed the specificity of the antibody in Sf9 cells via western blotting analysis of whole cell lysates from HeLa cells and *Drosophila melanogaster* embryos. The antibody yielded a 34 kDa band that represents authentic alpha-fibrillarin (Fig. III-7A). The Sf9 cell lysate also yielded a similar band at roughly 34 kDa, suggesting that the antibody recognizes *S. frugiperda* fibrillarin (Fig. III-7A). In Sf9 cells, the antibody detected fibrillarin proteins as punctate signals in nuclear regions (Fig. III-7B-c and -g). When R7Ag and R1Bm ORF2ps were expressed in Sf9 cells, I observed a few immunofluorescence co-localization signals of ORF2p with fibrillarin signals (Fig. III-7B -d and -h, arrowheads [yellow signals] in merged panel), suggesting that retrotransposable units of the two R elements gain access to the nucleolus through a mechanism involving ORF2p.

Despite having different retrotransposition target sites, R7Ag and R1Bm ORF proteins exhibited similar intracellular localization patterns: cytoplasmic localization of ORF1p, accumulation of ORF proteins in a peripheral region of the nuclear membrane, punctate nuclear localization of ORF2p, and localization of a small portion of ORF2p in the nucleolus (Figs. III-6 and III-7). These results indicate that both rDNA-specific elements share similar mechanisms of ORF proteins interaction, accumulation and access to the rDNA cluster target.

Discussion

Subcellular localization of rDNA-specific LINEs.

In this study, I found that the ORF2p of both R7Ag and R1Bm exhibits a punctate localization pattern in the nucleus; additionally, these proteins occasionally co-localize with the nucleolar component fibrillarin (Figs. III-6 and III-7). This observation indicates that different rDNA-specific elements share a common means of accessing the nucleolus. Notably, for both of the R elements, some of the ORF1p and ORF2p signals co-localized in the nucleus (Figs. III-5 and III-6), suggesting that ORF1p moved into the target site in an ORF2p-dependent manner. Recently Watanabe observed that R2Bm and R2Ol, other types of rDNA-specific elements that encode a single ORF and target 28S rDNA, also translocated to the nucleolus (Watanabe, 2013). Together with the above data, I speculate that rDNA-specific elements share transport mechanisms by which they access the nucleolus before initiating sequence-specific digestion of the bottom strands of rDNA targets. To clarify these rDNA-specific elements transportation mechanisms, I attempted to predict the potential nucleolar localization signal (NoLS) in ORF2p using the NoD server (Scott et al., 2011) and identified two NoLSs in R7Ag but found no signals in R1Bm (data not shown). The ORF2 of R1Bm may contain a non-canonical NoLS, or the two nuclear localization signals (NLSs) predicted in ORF2p by the cNLS Mapper (data not shown) might act as a NoLS because the NoLSs and NLSs are both rich in basic amino acids and often combine or overlap (Earley et al., 2015). Combining the lack of nuclear export signal (NES) predicted in both rDNA-specific elements (data not shown) with the above notion, it is implied that the subcellular localization of ORF2p may change from Type I (cytoplasm) to Type II (nuclei), but not from Type II to Type I. Further and comparative studies using various rDNA-specific elements will clarify the above

possibility and more detailed processes for the RNP localization.This is the first study to indicate that the ORF2p of LINEs is involved in the process by which the RNP complex accesses its target.

Proposed model for target accessing mechanism in R1-clade elements.

At the nuclear periphery, I observed the partial co-localization of both ORF1p and ORF2p in SART1Bm (Fig. III-2-i), R7Ag (Fig. III-6-c) and R1Bm (Fig. III-6-i), suggesting that these proteins associate with some cytoplasmic substructures. Stress granules (SGs) are cytoplasmic aggregates of stalled translational preinitiation complexes that are induced by a range of stress conditions. Processing bodies (P-bodys) are dynamic cytoplasmic compartments containing high concentrations of molecules involved mRNA translation repression, storage, and degradation (Anderson and Kedersha, 2006). Stress granules and processing bodies share RNA and protein components and can physically associate with one another (Kedersha et al., 2005). The ORF proteins and mRNA of L1 have been found to localize in the cytoplasmic foci often associated with stress granules (Doucet et al., 2010). This process is considered a type of host cell defense system by which the retrotransposition activity of L1 elements is controlled. In contrast, another report described that the yeast retrotransposons Ty3 and Ty1 are localized in the P-body, the components of which are involved in RNP complex assembly (Beliakova-Bethell et al., 2006; Checkley et al., 2010). Considering these observations, I speculate that ORF proteins co-localize in the P-body or stress granules, resulting in co-localization in the nuclear periphery.

Based on the above results and discussion, I proposed a model by which the R1-clade of site-specific LINEs access their targets (Fig. III-8). The ORF1p (blue oval) and ORF2p

(red oval) bind to their mRNA targets (waved black line) and form the RNP complex in the cytoplasm. The majority of RNP complexes accumulate in the periphery of nuclei associated with stress granules or P-bodies. Next, a portion of the accumulated RNP complex moves into the nucleus. During this process, the telomere-specific element accesses the TR in an ORF1p-dependent manner (Fig. III-8A). In contrast, the rDNA-specific elements R7Ag and R1Bm access the target site of the NO in an ORF2p-dependent manner (Fig. III-8B). Notably, only a very small fraction of the RNP can integrate into the target, which avoids critical damage to the host cells. Early-branched LINEs encode only one ORFp, and access of its RNP to the target site should depend on this protein. The recently-branched LINEs evolved from early-branched LINEs by acquiring an APE domain as well as ORF1 (Malik et al., 1999). These observations suggest that recently-branched LINEs evolved to access target sites in different manners of an ORF1p-dependent and ORF2p-dependent, respectively when diverged from early-branched LINEs.

TABLE III-1. Primers used in this study

Name	Sequence (5' to 3')
Plasmid construction for subcellular localization	
EcoRI-KOZAK-R71-F	AGAATTCGCCACCATGGATAAGCAACTGAGAGGAAGGAC
NotI-R71-R	AAAAAAAAGCGGCCGCAGTTCGAGGGTTGTCCACCGC
SpeI-KOZAK-R11-F	AACTAGTGCCACCATGTCGGAGGAGGAGAGGGAGC
EcoRI-R11-R-2	AAAAAAAAGAATTCCCATATCCATACTCGACCTGATTTAGGAAG
KpnI-KOZAK-3xFLAG-B-S	CGCCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCG ATTACAAGGACGA TGACGACAAGG
KpnI-KOZAK-3xFLAG-B-AS	GATCCCTTATCGTCATCGTCCTTGTAATCGATGTCATGATCTTTATAATCACCGT CATGGTCTTTGTAG TCCATGGTGGCGGTAC
SpeI-R7O2S	TTACTAGTATGGAAGTGCTACAGATCAA
EcoRI-R7O2A+3'UTR	AAGAATTCTTTTTTTTAATACTTAAGGATT
BamHI-R1O2S	AAAAAGGATCCATGGATATTAGGCCCCGAC
R1 A5136(NotI)	AAAAAGCGGCCGCTTCCCACCACCTCCCATGGTCCCACCAACCTTGC
EcoRI-KOZAK-S1-S880-Ser	AGAATTCGCCACCATGTCCAGTTATAAAGAAGAATTACCC
NotI-S1-A3016(FL)	AAAAAAAAGCGGCCGCCTTCGTCGTCCATTGGTGTCGC
S1ORF2-BamHI-S3018	TAGGATCCATGACCAGCAGCCCTTATCAT
SART1 A6704 NotI-vert	AAAAAAAAAAGCGGCCGCTCAAGGCAGCTGAGCAGG
Plasmid-based *in vivo* retrotransposition assay	
S1S6131	AGAAAGAGAGTGCGACCCAAACTCAGTT
CCTAA6	CCTAA CCTAA CCTAA CCTAA CCTAA CCTAA

TABLE III-2. Localization pattern of SART1Bm element proteins

Construct	n^t	Cytoplasmic	Nuclei
		n^c (%)	n^c (%)
SART1Bm ORF1	13	0	13(100%)
SART1Bm ORF2	15	15 (100%)	0
SART1Bm Co-expression	10	ORF1p and ORF2p co-localized predominantly in the cytoplasm with some dotted pattern in the nuclei.	

n^t: Total number of independent transfections observed

n^c (%): No. of counted cell (% of cell)

TABLE III-3. Localization pattern of R7Ag and R1Bm element proteins

Construct	n^t	Type I (Cytoplasmic)	Type II (Nuclei)
		n^c (%)	n^c (%)
R7Ag ORF1	25	25 (100%)	0
R1Bm ORF1	20	20 (100%)	0
R7Ag ORF2	17	8 (47.0%)	9 (53%)
R1Bm ORF2	22	9 (40.9%)	13 (59.1%)

n^t: Total number of independent transfections observed

n^c (%): No. of counted cell (% of cell)

R7Ag ORF2p Type I: Cytoplasmic broad localization.

R7Ag ORF2p Type II: Nucleus localization with some dotted pattern.

R1Bm ORF2p Type I: Cytoplasmic broad localization.

R1Bm ORF2p Type II: Cytoplasmic broad localization with some nucleus dotted pattern.

TABLE III-4. RNP localization pattern of R7Ag and R1Bm elements

Construct	n^t	Type I (Cytoplasmic)	Type II (Nuclei)	Non-overlap
		n^c (%):	n^c (%)	n^c (%):
R7Ag Co-expression	32	19 (59.3%)	4 (12.5%)	9 (28.2%)
R1Bm Co-expression	20	8 (40%)	3 (15.0%)	9 (45%)

n^t: Total number of independent transfections observed

n^c (%): No. of counted cell (% of cell)

Type I: Cytoplasmic dotted localization.

Type II: Nucleus dotted localization.

Non-overlap: An overlap of two proteins in fluorescence signal is not sufficient so that the overall degree of co-localization is not visually apparent.

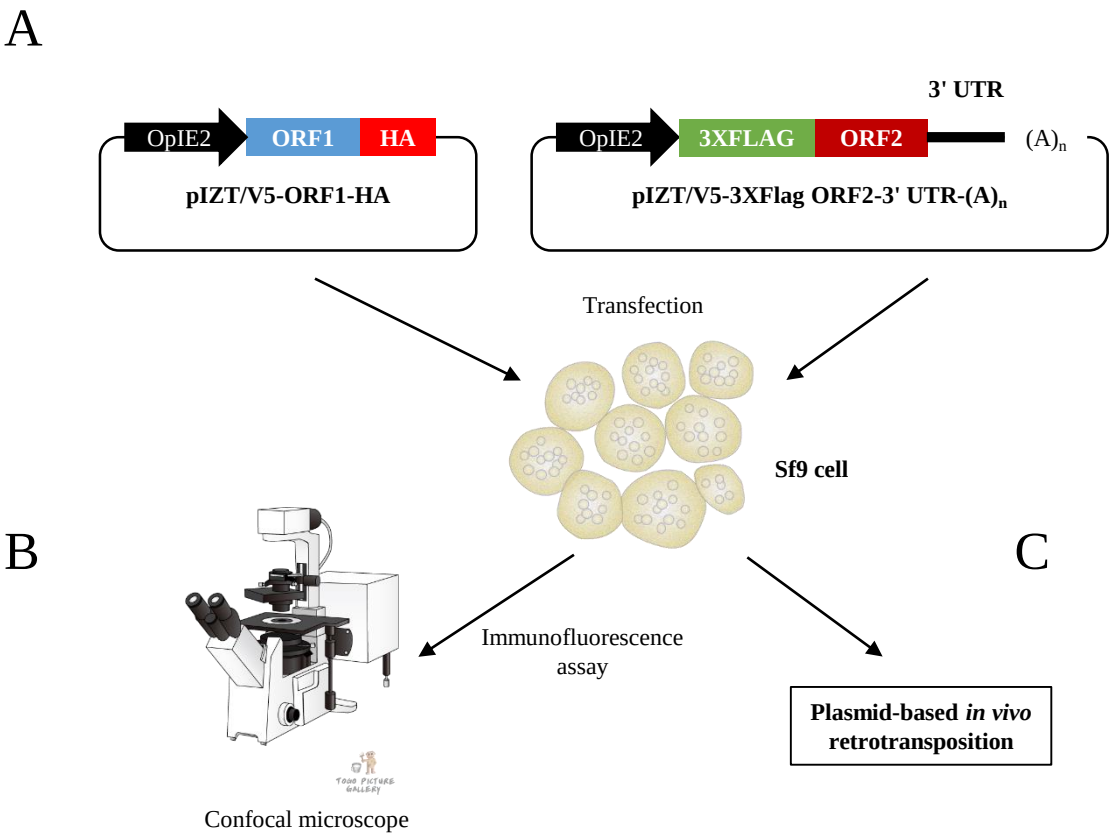

Fig. III-1. A schematic overview of a strategy for investigation of SART1Bm ORF proteins localization in transiently transfected Sf9 cells. (A) A diagram of SART1Bm plasmids used in this study. ORF1p was tagged with HA and ORF2p was tagged with 3xFLAG. Both ORFps expressed under the OpIE2 promoter. **(B)** Immunofluorescence assay was conducted on Sf9 cells 72 hours post-transfection and cells were observed by confocal microscope. **(C)** Plasmid-based *in vivo* retrotransposition was performed 72 hours post-transfection with extraction of Sf9 cell genome and detection of SART1Bm retrotransposition by PCR.

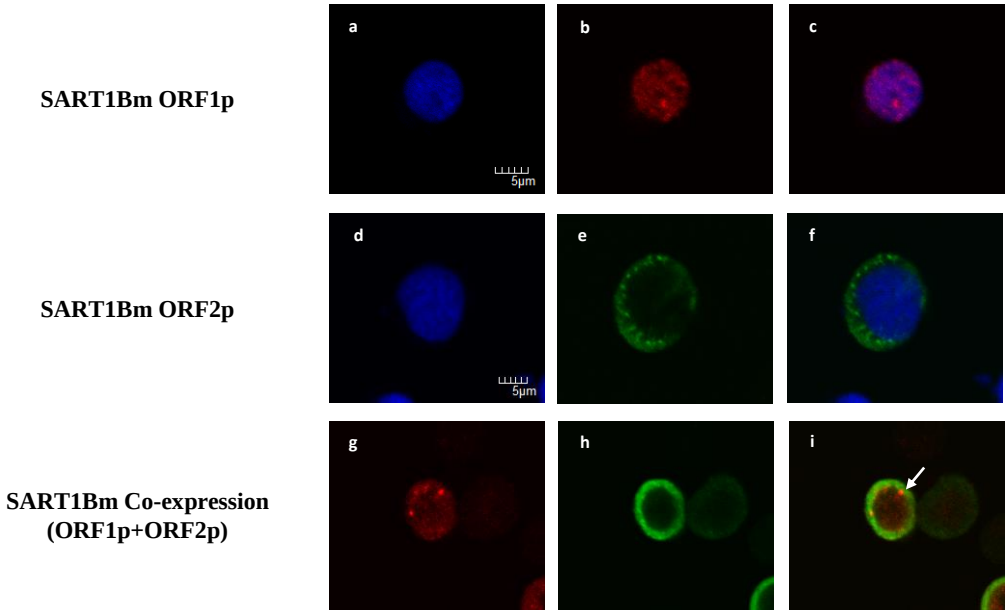

Fig. III-2. Subcellular localization of SART1Bm ORF proteins. Representative images of SART1Bm ORF1p (b and g) (red images) and SART1Bm ORF2p (e and h) (green images) in transfected cells are shown. DAPI (blue) was used to stain nuclear DNA (blue) (a and d). A merged image is shown at the far right (c, f and i). Scale bar; 5 mm.

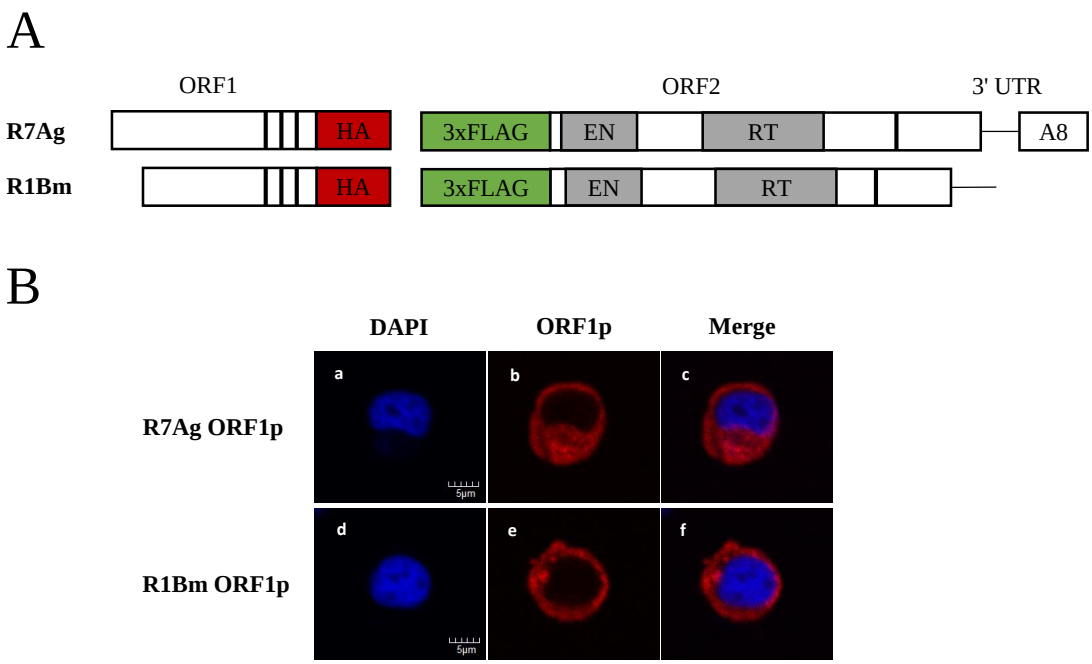

Fig. III-4. Subcellular localization of R7Ag and R1Bm ORF1p in transiently transfected Sf9 cells. (A) R7Ag and R1Bm constructs for the subcellular localization analysis. ORF1p is fused with an HA tag, and ORF2p is fused with a 3xFLAG tag. **(B)** Subcellular localization of the R7Ag and R1Bm ORF1p. Immunofluorescence of ORF1p in Sf9 cells was analyzed at 72 h post-transfection. Representative images of R7Ag (b) and R1Bm (e) ORF1p (red images) in transfected cells are shown. DAPI (blue) was used to stain nuclear DNA (a and d). A merged image is shown at the far right (c and f). Scale bar; 5 μm.

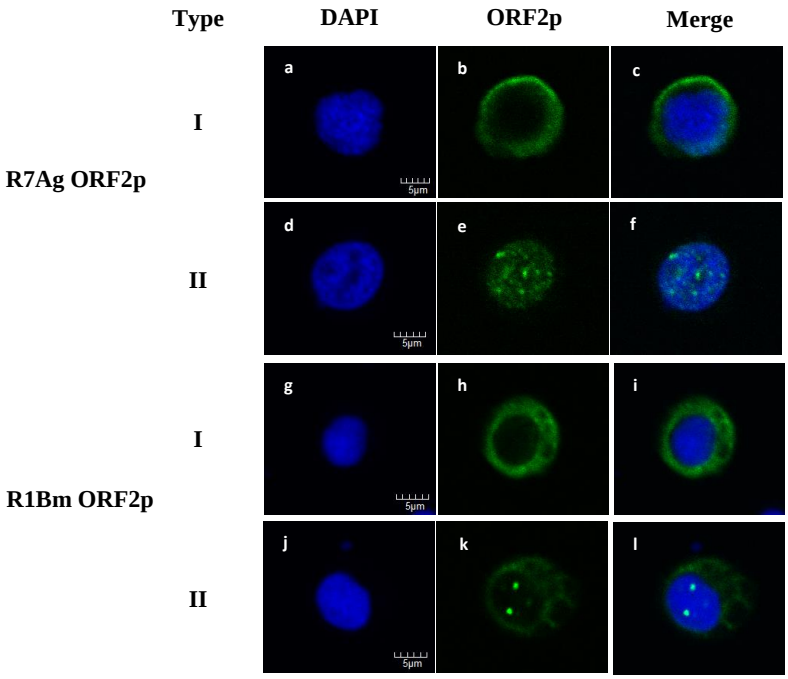

Fig. III-5. Subcellular localization of the R7Ag and R1Bm ORF2p. Representative images of R7Ag (b and e) and R1Bm (h and k) ORF2p (green) in transfected cells are shown. DAPI (blue) was used to stain nuclear DNA (a, d, g and j). A merged image is shown at the far right (c, f, i and l). Scale bar; 5 μm. Two types of ORF2p localization pattern were observed. Type I, cytoplasmic localization (b and h); Type II, nucleus localization with dotted signals (e and k). In Type II of R1Bm, weak cytoplasmic signals were also observed (k and l).

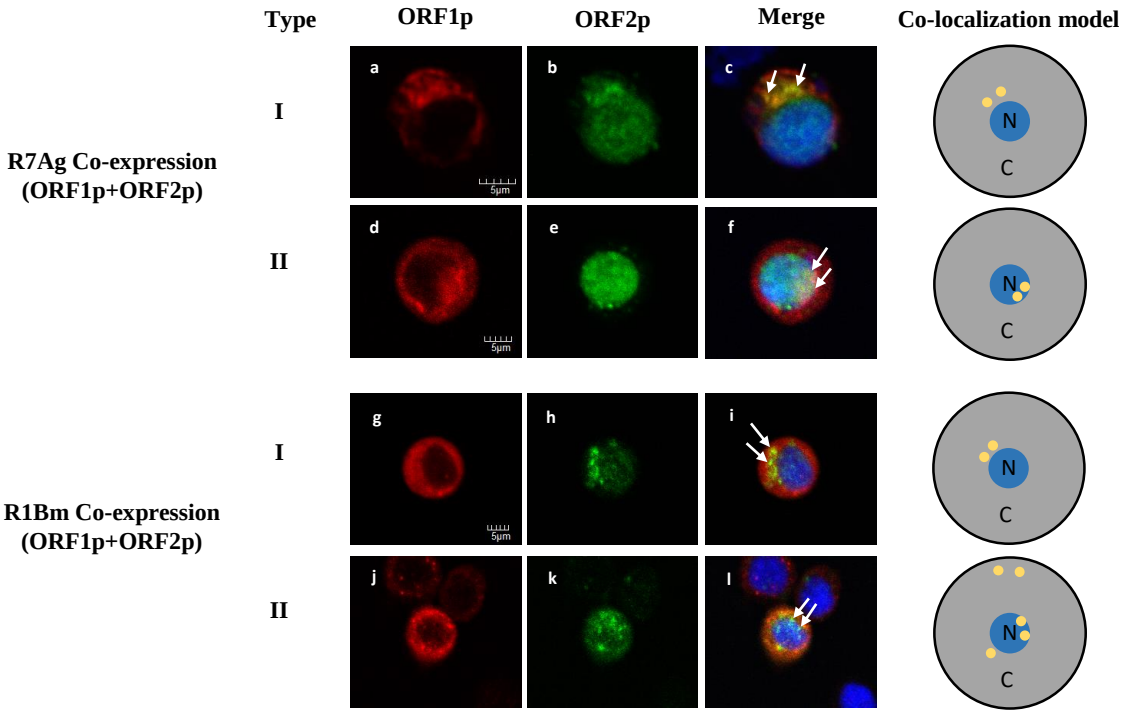

Fig. III-6. Co-localization of R7Ag and R1Bm ORF2p with own ORF1p.

Co-transfection of ORF1p and ORF2p constructs simultaneously yielded two type images of localization patterns both in R7Ag and R1Bm. A merged image is shown at right and co-localization of ORF1p (red) and ORF2p (green) is shown as yellow signals. In type I, ORF1p was localized in cytoplasm (a and g) and ORF2p was localized predominantly in nuclei but some in cytoplasm (b and h). A few signals for co-localization of ORF1p and ORF2p, were observed at peripheral region outside the nuclear membrane (c and i, arrowheads). In type II, ORF1p was localized predominantly in cytoplasm but some in nuclei (d and j) and ORF2p was localized in nuclei (e and k). Co-localization signals in type II were observed in nuclei (f and l, arrowheads). The co-localization model is shown in the right end. Yellow dot, co-localization signal; C, cytoplasm; N, nucleus.

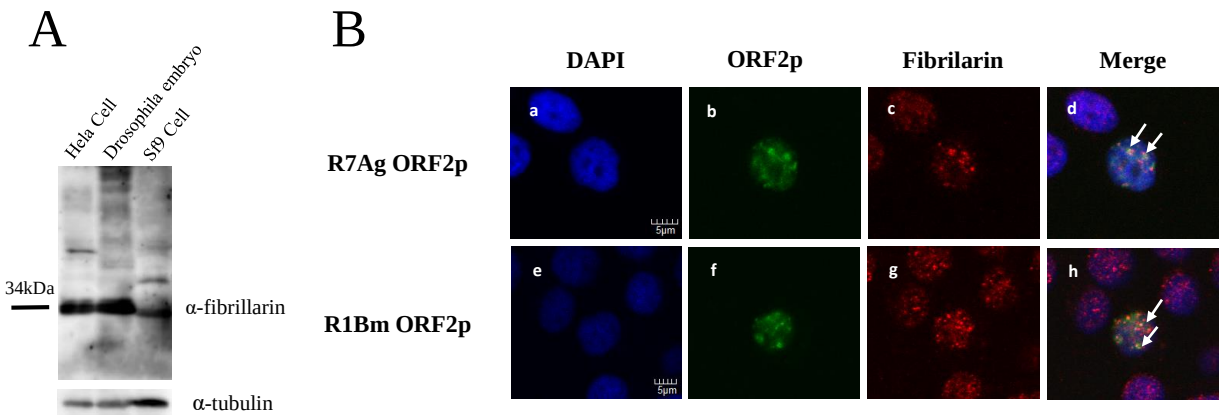

Fig. III-7. Co-localization of R7Ag and R1Bm ORF2p the nucleolar marker protein fibrillarin. (A) A rabbit polyclonal anti-fibrillarin antibody specifically detects an endogenous fibrillarin protein in extracts from HeLa cell, *Drosophila* embryo and Sf9 cells. α-tubulin is shown as a gel loading control. **(B)** Immunofluorescence analysis of Sf9 cells transfected with R7Ag and R1Bm ORF2 constructs. Images of cells stained with antibodies against the endogenous nucleolus component fibrillarin (red) are shown. R7Ag and R1Bm ORF2p (green) are indicated. DAPI (blue) was used to stain nuclear DNA. A merged image is shown in the rightmost columns. Co-localization of ORF2p and fibrillarin is shown by arrows (yellow).

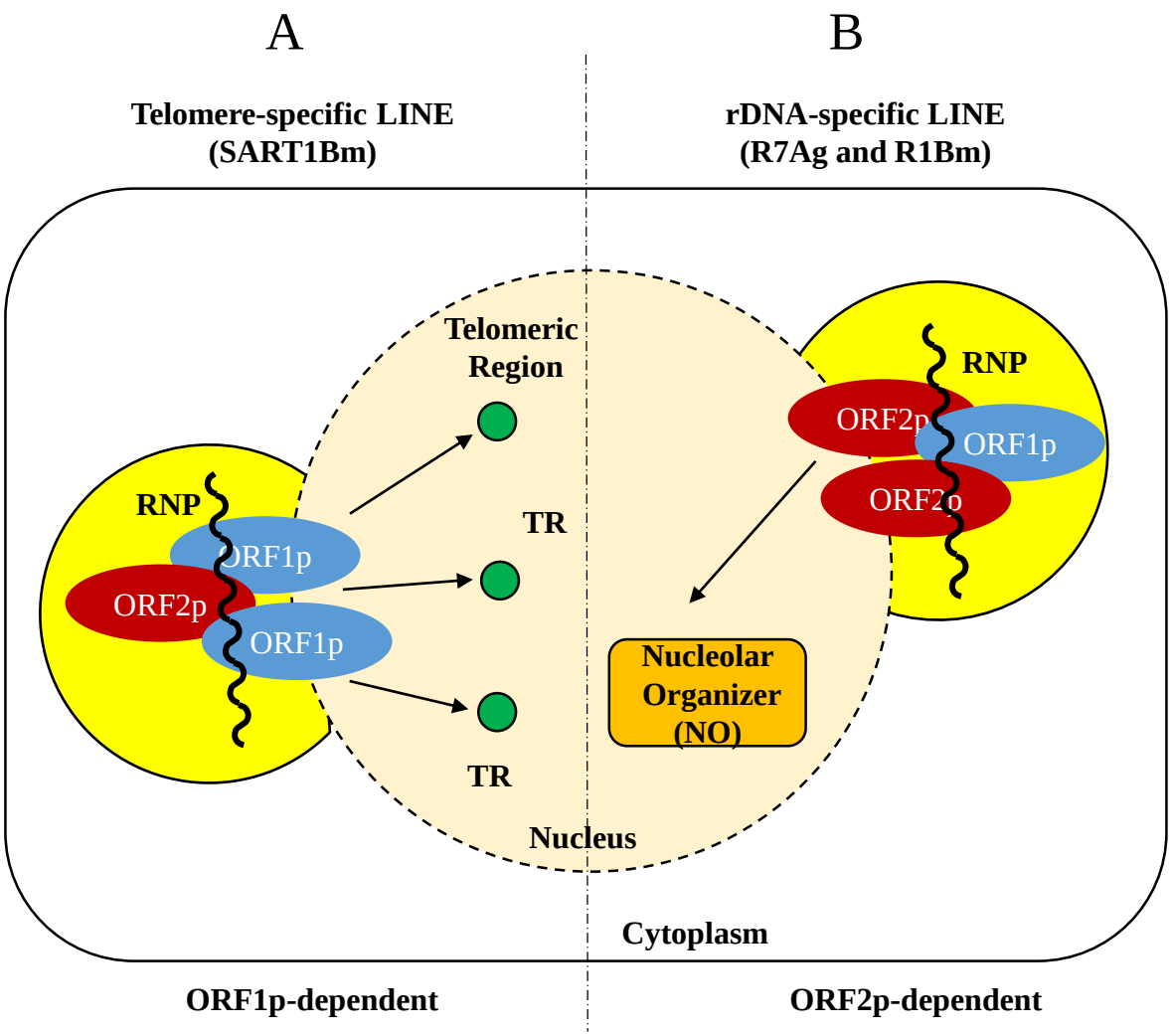

Fig. III-8. Proposed model for subcellular localization of ORFps for telomere-specific and rDNA-specific elements. ORF1p, blue oval; ORF2p, red oval; mRNA of R elements, waved black line; TR, telomeric region; NO, nucleolar organizer; RNP, ribonucleoprotein complex. The ORF1p and ORF2p bind to their mRNA and form the RNP complex in the cytoplasm. The majority of the RNP complexes accumulate in the periphery of nuclei which may associate with stress granules or P-bodies (yellow circle). Next, a portion of the accumulated RNP complex moves into the nucleus. During this process, **(A)** telomere-specific element accesses the TR in an ORF1p-dependent manner. **(B)** rDNA-specific elements R7Ag and R1Bm access the target site of the NO in an ORF2p-dependent manner.

References

Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH, Moran JV. 2006. Unconventional translation of mammalian LINE-1 retrotransposons. Genes Dev. **20**:210-224

Anderson P, Kedersha N. 2006. RNA granules. J Cell Biol. **172**:803-808.

Beliakova-Bethell N, Beckham C, Giddings TH, Winey M, Parker R, Sandmeyer S. 2006. Virus-like particles of the Ty3 retrotransposon assemble in association with P-body components. RNA. **12**:94–101.

Checkley MA, Nagashima K, Lockett SJ, Nyswaner KM, Garfinkel DJ. 2010. P-body components are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles. Mol Cell Biol. **30**:382–398.

Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran JV, Gilbert N. 2010. Characterization of LINE-1 ribonucleoprotein particles. PLoS Genet. **6**. pii: e1001150.

Earley LF, Kawano Y, Adachi K, Sun X-X, Dai M-S, Nakai H. 2015. Identification and Characterization of Nuclear and Nucleolar Localization Signals in the Adeno-Associated Virus Serotype 2 Assembly-Activating Protein. J Virol. **89**:3038–3048.

Goodier JL, Zhang L, Vetter MR, Kazazian HH Jr. 2007. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. Mol Cell Biol. **27**:6469-6483.

Ishibashi, M. 2004. Analysis for region required for telomere localization of telomeric repeat-specific LINE, SART1. Master Thesis, University of Tokyo.

Kedersha N, Stoecklin G, Ayodele M, Yacono P, Lykke-Andersen J, Fritzler MJ, Scheuner D, Kaufman RJ, Golan DE, Anderson P. 2005. Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. J Cell Biol. **169**:871-884.

Kojima KK, Matsumoto T, Fujiwara H. 2005. Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon SART1. Mol Cell Biol. **25**:7675-7686.

Kubo S, Seleme MDC, Soifer HS, Perez JLG, Moran J V, Kazazian HH Jr, Kasahara N. 2006. L1

retrotransposition in nondividing and primary human somatic cells. Proc Natl Acad Sci USA. **103**:8036–8041

Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol. **16**:793–805.

Matsumoto T. 2006. Functional analysis of ORF1 protein of telomere-specific non-LTR retrotransposon, SART1. Ph. D. Dissertation, University of Tokyo.

Matsumoto T, Hamada M, Osanai M, Fujiwara H. 2006. Essential domains for ribonucleoprotein complex formation required for retrotransposition of telomere-specific non-long terminal repeat retrotransposon SART1. Mol Cell Biol. **26**:5168–5179.

Matsumoto T, Takahashi H, Fujiwara H. 2004. Targeted nuclear import of open reading frame 1 protein is required for in vivo retrotransposition of a telomere-specific non-long terminal repeat retrotransposon, SART1. Mol Cell Biol. **24**:105–122.

Rashkova S, Karam SE, Kellum R, Pardue M. 2002. Gag proteins of the two Drosophila telomeric retrotransposons are targeted to chromosome ends. J Cell Biol. **159**:397–402.

Scott MS, Troshin P V, Barton GJ. 2011. NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins. BMC Bioinformatics. **12**:317.

Shi X, Seluanov A, Gorbunova V. 2007. Cell divisions are required for L1 retrotransposition. Mol Cell Biol. **27**:1264–1270.

Theilmann DA, Stewart S. 1992. Molecular analysis of the trans-activating IE-2 gene of Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus. Virology. **187**:84-96.

Watanabe H. 2013. Subcellular localization mechanism of site-specific LINE. Master Thesis, University of Tokyo.

Xie Y, Mates L, Ivics Z, Izsvák Z, Martin SL, An W. 2013. Cell division promotes efficient retrotransposition in a stable L1 reporter cell line. Mob DNA. **4**:10.

Xing Y, Shi Z. 2011. Nucleocapsid protein VP15 of White spot syndrome virus colocalizes with the nucleolar proteins nucleolin and fibrillarin. Can J Microbiol. **57**:759-764.

General conclusion

In this study, to elucidate how R1-clade elements specifically insert into their target sites, I compared two telomere-specific elements (SART1Bm and TRAS1Bm) and two rDNA-specific elements (R7Ag and R1Bm) with respect to three parameters: the precision of DNA target site recognition (Chapter I), the accurate initiation of reverse transcription (Chapter II), and the subcellular localization of ORFps (Chapter III). There were two important differences between telomere-specific and rDNA-specific elements. First, telomere-specific elements flexibly recognize relatively short sequence, while rDNA-specific LINEs recognize longer target sites around TSD (Chapter I). Second, telomere-specific elements approach their target sites in an ORF1p-dependent manner, while rDNA-specific elements in an ORF2p-dependent manner, respectively (Chapter III). In addition, one common feature is that a long poly(A) tail is necessary for the accurate initiation of reverse transcription in both telomere-specific and rDNA-specific elements, a novel mechanism found in LINEs that may apply to other R1-clade elements (Chapter II).

Based on these observation, I speculate that when R1-clade elements diverged from a common ancestral element, the ORFps of each element gained access to repetitive sequences with a high copy number to determine site-specificity; telomere and rDNA are selected as ideal targets. The EN domain then evolved to obtain sequence-specificity and diversified its target sites in telomeric or rDNA repeats. Thus, in R1-clade elements, their target selection is not simply determined by the EN domain but by every step of the LINE life cycle, ensuring their specificity to the target site.

On the other hand, in this study, I developed two novel retrotransposition assays, the *trans-in vivo* retrotransposition assay and the plasmid-based *in vivo* retrotransposition

assay, which are more convenient than the previous baculovirus-based system (Table GC-1). Especially, the plasmid-based *in vivo* retrotransposition assay should allow us to gain a greater understanding of R1-clade elements retrotransposition mechanisms in the future.

TABLE GC-1. Comparison of retrotransposition assays

Assay / Features	*In vivo* retrotransposition			*Ex vivo* retrotransposition
	Baculovirus-based *in vivo* retrotransposition	*Trans-in vivo* retrotransposition	Plasmid-based *in vivo* retrotransposition	
Vector type	Baculovirus	Baculovirus (Driver) Plasmid (Reporter)	Plasmid	Baculovirus
Target	Genome	Genome	Genome	Plasmid
References	Takahashi and Fujiwara, 2002	Novel in this study	Novel in this study	Osanai-Futahashi and Fujiwara, 2011

References

Osanai-Futahashi M, Fujiwara H. 2011. Coevolution of telomeric repeats and telomeric repeat-specific non-LTR retrotransposons in insects. Mol Biol Evol **28**:2983–2986.

Takahashi H, Fujiwara H. 2002. Transplantation of target site specificity by swapping the endonuclease domains of two LINEs. EMBO J **21**:408–417.

ACKNOWLEDGEMENTS

At first, I would like to thank my supervisor Haruhiko Fujiwara who has offered me valuable suggestions in the academic studies. He has spent much time on this dissertation. Without his patient instruction, insightful criticism and expert guidance, the completion of this dissertation would not have been possible.

Second, I thank post-doc Azusa Kuroki, who taught me not just principle of science but also methods of experiment one by one just like a baby-sitter. Without her sincere help I could not enter into the science field.

Next, I would like to thank associate professor Kojima, Tetsuya, post-doc Akira, Ishizuka and other lab members who always provided me good suggestions and kind encouragements.

At last but not the least, I would like to thank my family members who have been assisting, supporting and caring for me all of my life. Especially I owe my deepest gratitude to my love Surina and my daughter Cheelus. Without their unyielding supports, I could not complete this dissertation.