Coevolution of Telomeric Repeats and Telomeric Repeat–Specific Non-LTR Retrotransposons in Insects

Mizuko Osanai-Futahashi†,1 and Haruhiko Fujiwara*,1

†Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba, Japan
‡Present address: Transgenic Silkworm Research Center, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan
*Corresponding author: E-mail: haruh@k.u-tokyo.ac.jp.

Abstract

In the telomeres of the silkworm Bombyx mori, telomeric repeat–specific non-long terminal repeat (LTR) retrotransposon SARTBm1 is accumulated in the TTAGG telomeric repeats. Here, we identify novel telomeric repeat–specific non-LTR retrotransposons, SARTTc family, from the red flour beetle Tribolium castaneum in the unconventional TCAGG telomeric repeats. To compare the sequence specificity of SARTBm1 and SARTTc1, we developed a comparable ex vivo retrotransposition assay. Both SARTBm1 and SARTTc1 preferred the telomeric sequence of their hosts, suggesting that the target specificity of these retrotransposons coevolved with their host’s telomeric repeats. Swapping experiment indicated that the endonuclease domain is involved in recognizing the target sequence. Moreover, SARTBm1 proteins could retrotranspose 3’ untranslated region (UTR) sequence of SARTTc1 as well as their own 3’ UTR, whereas SARTTc1 proteins could only retrotranspose their own 3’ UTRs. These results provide insights to the mechanism and divergence of sequence specificity and 3’ UTR recognition in non-LTR retrotransposons.

Key words: non-LTR retrotransposons, telomeric repeats, sequence specificity, coevolution, telomere.

Non-long terminal repeat (LTR) retrotransposons are mobile elements that are widespread across the genome of eukaryotes. Although most of them insert randomly throughout the genome, few of them insert exclusively into specific genomic sites, such as ribosomal DNA and telomeric repeats, in a sequence-specific manner (Xiong and Eickbush 1988a, 1988b; Kojima and Fujiwara 2003, 2004; Fujiwara et al. 2005). Sequence specificity in these retrotransposons is considered as a survival strategy to avoid destroying their hosts by inserting into essential genes. In the silkworm Bombyx mori, there are telomeric repeat–specific retrotransposons, such as SART1 (SARTBm1) and TRAS1 in the TTAGG telomeric repeats (Fujiwara et al. 2005). Because telomerase activities could not be detected in all investigated tissues of B. mori, including testis (Sasaki and Fujiwara 2000), it is hypothesized that SARTBm1 and TRAS1 support the telomere elongation by retrotransposition, gene conversion, and recombination (Fujiwara et al. 2005). To date, telomeric repeat–specific retrotransposons have been only reported in Lepidoptera.

Recently, we found that telomeres of red flour beetle Tribolium castaneum consist of TCAGG repeats (Osanai et al. 2006), instead of TTAGG repeats found in many other insect and arthropod species (Okazaki et al. 1993; Kojima et al. 2002; Fujiwara et al. 2005) (fig. 1A). Here, we identify seven types of non-LTR retrotransposons inserted into TCAGG telomeric repeats of T. castaneum (fig. 1B, supplementary materials and methods, Supplementary Material online). The orientations of all these retrotransposons in telomeric repeats were the same as SARTBm1 (reverse orientation with TRAS). The domain structures and junction site sequences of these retrotransposons were all quite similar to those of SARTBm1 (supplementary fig. S1 and table S1, Supplementary Material online). Phylogenetic analyses using amino acid sequences of reverse transcriptase (RT) domain revealed that the all these retrotransposons of T. castaneum belong to R1 clade and were phylogenetically closer to SARTBm1 than TRAS1 (fig. 1B, supplementary fig. S2, Supplementary Material online). Thus, we named these telomeric repeat–specific non-LTR retrotransposons of T. castaneum SARTTc (SARTTc1–SARTTc7). These are first examples of telomeric repeat–specific retrotransposons outside Lepidoptera.

SARTBm1 and SARTTc seem to be an ideal model to study how retrotransposition mechanisms of closely related non-LTR retrotransposons diverge. To investigate the sequence specificity of SARTTc (SARTTc1) and SARTBm1, we modified the previous ex vivo retrotransposition assay (Takahashi and Fujiwara 2002; Osanai et al. 2004; Matsumoto et al. 2006). Because the original insertion sequence of SARTTc is (TCAGG)n, we transfected Sf9 cells with plasmids containing TCAGG or TTAGG repeats as targets for retrotransposition (supplementary fig. S3, Supplementary Material online). Both plasmids have the same backbone, and only the telomeric repeat sequences are different. We infected these cells with the Autographa californica nuclear polyhedrosis virus construct which contains ORF1/ORF2/3’ untranslated region (UTR) portion of either SARTTc1 or SARTBm1. Retrotransposition was detected by polymerase chain reaction (PCR) amplifying the 3’ junction of retrotransposed SARTTc1/SARTBm1 elements to the telomeric repeats of the plasmids (supplementary fig. S3A, Supplementary Material online). We quantified the target sequence preferences by two approaches: 1) transfecting each plasmids into separate cells and detecting retrotransposition by quantitative PCR
FIG. 1. (A): Phylogenetic relationships of host species and their telomeric repeats. Dipterans do not have conventional telomeric repeats. (B) Neighbor Joining tree of RT domains among SARTTc family (○) and other R1 clade non-LTR retrotransposons. The numbers at each node are the bootstrap values (over 50% are shown). The host species and insertion sites of non-LTR retrotransposons are indicated in the right.

(supplementary fig. S3B, Supplementary Material online) and 2) transfecting the mixture of equal amounts of both plasmids into the same cells and detecting the retrotransposition by PCR, subcloning and sequencing the PCR products, and counting the number of clones for each plasmid (supplementary fig. S3C, Supplementary Material online). By both methods, SARTTc1 had target sequence preference to TCAGG repeats, and SARTBm1 had target sequence preference to TTAGG repeats (fig. 2, supplementary fig. S4, Supplementary Material online). The 3' UTR sequences were always followed by GG(TTAGG)n or GG(TCAGG)n as expected (supplementary table S3–S7, Supplementary Material online). Thus, the target sequence preference of SARTTc1 and SARTBm1 coincides with their host's telomeric repeat sequence.

From the wide distribution pattern of (TTAGG)n among arthropods, the telomeric repeat of the T. castaneum or its ancestor are likely to have converted from TTAGG to TCAGG. Notably, although both SARTTc1 and SARTBm1

FIG. 2. Target sequence preference analyses of SARTBm1 and SARTTc1 by ex vivo retrotransposition assay (simultaneous transfection method). Sf9 cells were transfected with the mixture of equal amounts of TCAGG29-pBSK and TTAGG25-pBSK and subsequently infected by baculovirus Autographa californica nuclear polyhedrosis virus containing SARTTc1 or SARTBm1 construct (shown on the left). To analyze the target preferences, first, retrotransposition into the plasmids was detected by PCR using SARTTc1 or SARTBm1 primers designed in the 3' region of the ORF2 (TcF6l or Bm-S6131, supplementary table S2) and a primer designed on the common region of the telomere plasmids (A878T, supplementary table S2). Next, PCR products were subcloned and sequenced, and the number of clones obtained for retrotransposition into the TCAGG (white) and TTAGG (black) were counted (right panel). Endonuclease swapping was conducted by trans complementation.
prefer their host telomeric repeat sequences, they could also retrotranspose into TTAGG and TCAGG repeats, respectively (fig. 2, supplementary fig. S4, Supplementary Material online). If the sequence specificity were so strict that it could not insert into other sequences at all, telomeric repeat–specific retrotransposons would not be able to accommodate to the alteration of the telomeric repeat. Thus, B. mori and T. castaneum may be a unique example of co-evolution of telomeric repeat–specific retrotransposons and telomeric repeats. Interestingly, in R1 clade, all sequence-specific non-LTR retrotransposons (not into telomeric repeats) except for R1 are found from Anopheles gambiae whose telomeric repeats are lost (fig. 1A). RT, R7, R6, Waldo (fig. 18) may be retrotransposons which have altered the sequence specificity from (TTAGG), to other sequences in order to adapt to the change of telomere structure of the host genome, as like SARTTc family.

Several reports have suggested the involvement of the endonuclease (EN) domain in sequence specificity in some non-LTR retrotransposons (Feng et al. 1998; Anzai et al. 2001; Maita et al. 2004, 2007), although sequence specificity of EN itself has not been determined in SARTBm1. To examine if the endonuclease domain plays a major role in the sequence specificity of SARTBm1 and SARTTc1, we attempted EN swapping experiments by trans complementation. The retrotransposition of EN deletion constructs (∆EN) could not be detected but were rescued with their native sequence preference by coexpressing with their own ENs (fig. 2, supplementary fig. S4, Supplementary Material online). Importantly, when SARTBm1EN was coexpressed with SARTTc1∆EN, the retrotransposition events into TTAGG repeats exceeded that into TCAGG repeats (fig. 2, bold, supplementary fig. S4, Supplementary Material online), suggesting that one major cause of the sequence specificity is derived from the endonuclease. When SARTBm1∆EN was coexpressed with SARTTc1EN, the retrotransposition was below the detectable level (supplementary fig. S4, Supplementary Material online). It is assumed that the residues at the putative DNA-binding surface of EN are important for sequence specificity (Maita et al. 2004, 2007). When the EN amino acid sequences were aligned, there were 41 positions which residues were only in TTAGG inserting elements, 4 positions which residues were only in TCAGG inserting elements, and 5 positions which different residues were conserved for TTAGG and TCAGG inserting elements (supplementary fig. S5, Supplementary Material online). On the SARTTc1EN and SARTBm1EN structure model based on TRAS1EN crystal structure (supplementary fig. S6, Supplementary Material online), 11 of these residues were located in the DNA-binding surface (supplementary fig. S5, Supplementary Material online), suggesting that these residues may be needed for the sequence specificity of SARTTc1EN/SARTBm1EN.

Previously, we found that SARTBm1 proteins recognize their 3’ UTR in retrotransposition (Osanai et al. 2004). Similar to SARTBm1, when the 180 bp of SARTTc1 3’ UTR was added to the enhanced green fluorescent protein (EGFP), the fusion construct could retrotranspose by SARTTc1 proteins (supplementary fig. S7, Supplementary Material online). When 1–60 bp of the 3’ UTR was deleted, retrotransposition failed. In this essential region, there was a stem–loop structure (supplementary fig. S7E, Supplementary Material online) similar to SARTBm1 3’ UTR (Osanai et al. 2004) (supplementary fig. S7F, Supplementary Material online). Moreover, in SARTTc1 3’ UTR, there were several (G)GU(U/C) and UAG sequences (supplementary fig. S7D, Supplementary Material online). This is also similar to SARTBm1 3’ UTR, which have GGUU and UAG sequences in the essential region, which could possibly anneal to target telomeric repeat (CCTAA), strand and function in insertion specificity into telomeric repeats (Osanai et al. 2004). Although the primary sequences of the 3’ UTR of SARTTc1 and SARTBm1 have low sequence similarity (supplementary fig. S8, Supplementary Material online), the similar characteristics such as stem–loop structure in the essential region and the existence of downstream (G)GU(U/C) sequence made us think the possibility that SARTBm1 proteins can recognize SARTTc1 3’ UTR and vice versa. To analyze this possibility, we coexpressed SARTTc1 proteins with EGFP-SARTBm1-3’ UTR construct and SARTBm1 proteins with EGFP-SARTTc1-3’ UTR construct. As shown in fig. 3, SARTBm1 proteins could retrotranspose SARTTc1 3’ UTR construct as well as their own 3’ UTR construct, suggesting that the SARTBm1 proteins recognize the higher order structure of the 3’ UTR RNA rather than the primary sequence in retrotransposition. Conversely, with SARTTc1 proteins, retrotransposition was only detected with their own 3’ UTR construct. Because the
stem–loop in the essential region of SARTBm1 is larger than that of SARTTc1 (Osanai et al. 2004) (supplementary fig. S7E, F, Supplementary Material online), one possible reason for the failure of the recognition of the SARTBm1 3′ UTR by SARTTc1 proteins is that the RNA recognition domain of SARTTc1 is smaller than SARTBm1. At present, there is very little information on how non-LTR retrotransposon proteins recognize their RNAs in retrotransposition. Extensive analysis of secondary structure of 3′ UTRs and protein recognition of RNA in SARTBm1, SARTTc1, and related elements should reveal the evolution and mechanisms of 3′ UTR recognition.

Supplementary Material
Supplementary tables S1–S7 and figures S1–S8 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
We thank Dr N. Maita for advice on structure modeling. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan and the Program for Promotion of Basic Research Activities for Innovative Biosciences to H.F.

References
Xiong Y, Eckbush TH. 1988a. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol Cell Biol. 8:114–123.