

A new family of site-specific retrotransposons, SART1, is inserted into telomeric repeats of the silkworm, *Bombyx mori*

Hidekazu Takahashi, Satoshi Okazaki and Haruhiko Fujiwara*

Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Received December 16, 1996; Revised and Accepted March 4, 1997

DDBJ/EMBL/GenBank accession no. D85594

ABSTRACT

The telomeres of the silkworm, *Bombyx mori*, consist of pentanucleotide repeats (TTAGG)_n. We previously characterized the non-LTR element TRAS1, which terminates with oligo (A) in a head to tail orientation at the exact position (between A and C) of the (CCTAA)_n repeats. Here we characterized another family of telomere-specific non-LTR retrotransposon named SART1. The SART1 family was inserted at another site of the (TTAGG)_n in a reverse orientation from that of TRAS1. The complete unit of SART1, 6.7 kb in length with a poly (A) stretch, contains two open reading frames encoding putative gag and pol products, overlapping by 54 bp in the –1 reading frame. Most of the 600 SART1 copies in the silkworm haploid genome are completely conserved in structure without 5' truncation. All SART1 sequences analyzed were inserted at the same position (between T and A) within the (TTAGG)_n repeats. Fluorescence *in situ* hybridization showed that many of the SART1 copies were localized in the chromosomal ends. A phylogenetic tree showed that the SART1, TRAS1 and two other site-specific elements, R1 and RT, which insert into 28S ribosomal RNA genes in insects, belong to the same group. Based on the orientation for the chromosomal insertion and structural similarities, these elements could be further classified into two subgroups, R1/TRAS1 and RT/SART1, suggesting that the target specificity of the two telomere-associated elements was changed independently.

INTRODUCTION

Retrotransposons are mobile DNA elements within the genome that transpose via RNA intermediates using their own-encoding reverse transcriptase. These elements have been classified into two groups according to whether or not they contain long terminal repeats at both ends (LTR type or retrovirus-like element) (non-LTR type or LINE-like element) (1–3). Although many retrotransposable elements are located nearly at random in the

genome, several non-LTR retrotransposons are located at specific sites on the chromosome. To date, site-specific elements have been found in several organisms, including yeast, insects, nematodes, amphibians and protists. Tx1 of *Xenopus laevis* inserts within another transposable element (4). CRE1, SLACS and CZAR are found in the spliced leader exons of *trypanosomes* (5–7). R1 and R2, which are located at specific sites of 28S rDNA in most insects, are site-specific retrotransposons and have been investigated in detail (8–10). R4 (*Ascaris lumbricoides*) inserts at a site midway between those of R1 and R2 (11). In the sibling mosquito species, RT1 (*Anopheles gambiae*) and RT2 (*A. arabiensis*) are inserted at the same position ~630 bp downstream of the R1 insertion site (12,13). It was shown that R2 element owes its site specificity to the sequence-specific endonuclease that it putatively encodes. The translation product of the R2Bm open reading frame (ORF) from *Bombyx mori* cuts the 28S gene at the correct insertion site and uses that site as primer for its reverse transcription (14,15). Moreover, some group II introns, which have a close evolutionary relationship to non-LTR retrotransposons, encode reverse transcriptases that make a double strand break on the target sequences and promote site-specific insertion (16,17). However, a general mechanism underlying the expression and integration of the site-specific non-LTR retrotransposons remains unclear.

The extreme ends of the silkworm chromosomes are composed of the simple telomeric repeat (TTAGG)_n, which is >6–8 kb in length (18). When studying the telomeric structure of the silkworm, we found >2000 copies of retrotransposable elements, which may be classified into several different classes, inserted into the telomeric short repeats (TTAGG)_n (Okazaki *et al.*, in preparation). We showed that a family of non-LTR retrotransposons, TRAS1 (telomeric repeat associated sequence 1) interrupts the (TTAGG)_n repeats at a highly specific site (18,19). On the other hand, retrotransposons of the principle telomere-associated family are inserted at another site of (TTAGG)_n in a reverse orientation to the TRAS1 insertion. Thus, this new family of retrotransposons will be transcribed in the reverse direction to TRAS1. We called this family SART1, the name of which derives from the inversion of 'TRAS'. To try to understand the factors and structure involved in target specificity, we characterized SART1, because it seems to be a novel telomeric repeat-specific

*To whom correspondence should be addressed. Tel: +81 3 3812 2111; Fax: +81 3 3816 1965; Email: haruh@uts2.s.u-tokyo.ac.jp

retrotransposon with a different target sequence from TRAS1. We studied the structural features and genomic organization of SART1 and discuss the evolutionary origin of the target-specific non-LTR retrotransposons in insects.

MATERIALS AND METHODS

Bombyx mori strain

The *B. mori* strain, Kinshu × Showa, purchased from Kyodo-shiryo Co., Tokyo, Japan, was reared on an artificial diet (Yakult Co., Tokyo) in the laboratory.

Screening of SART1 clones from a genomic lambda phage library

We screened phage clones from an EMBL3 genomic DNA library constructed from *Sau3AI* partial digests of *Bombyx* fat body DNA (19). Clones containing (TTAGG)_n repeats were isolated with a ³²P-labeled (TTAGG)₅ oligonucleotide probe. After sequencing adjacent regions around the (TTAGG)_n from >20 positive clones, we found that several clones terminating with a poly(A) tail had the same nucleotide sequence that neighbors the (TTAGG)_n in the reverse orientation of the TRAS1 insertion. We named this putative retrotransposon family, SART1. From one phage clone containing SART1, we isolated a 1.8 kb *HindIII*–*SacI* DNA fragment from the 3′-terminal region near the poly(A) tail. We cloned this fragment into pBluescript SK+ (Stratagene) and named it p51S.

Cloning a complete unit of SART1 and sequencing

To identify a clone containing a complete unit of the SART1 element, we screened the EMBL phage library again with the ³²P-labeled 1.8 kb insert of the p51S as a probe. Among 20 000 plaques, we isolated 20 clones with intense signals and found eight clones containing two (TTAGG)_n separated repeats, which may correspond to the terminal junction in both ends of SART1. We named one of the eight clones, which includes a full-length SART1 element as determined by genomic Southern hybridization, BS103. We digested BS103 with *Bam*HI and *Hind*III and obtained from the insert region, 3.5 kb *Bam*HI/*Bam*HI, 6.7 kb *Hind*III/*Hind*III and 6.7 kb *Bam*HI/*Bam*HI digests, which were subcloned into pBluescript SK+. Because we found that BS103 included three SART1 elements arrayed in tandem, these subclones were sequenced to determine the complete structure of SART1. Sequencing was performed by dideoxynucleotide termination (20) using a Thermo Sequenase core sequencing kit (Amersham) and an automatic DNA sequencer SQ5500 (Hitachi). We used DNASIS-Mac version 3.0 (Hitachi Software) for sequence analysis.

Genomic Southern hybridization

Genomic DNA was prepared from the silk glands of fifth instar larvae as described previously (19). DNAs digested with restriction enzymes were electrophoresed on 0.9% agarose gels and blotted onto nitrocellulose membranes (BA85; Schleicher and Schuell) in 20× SSC (3 M NaCl, 0.3 M sodium citrate) by capillary transfer (21). Hybridization was performed in 0.9 M NaCl, 90 mM Tris–HCl (pH 7.9), 6 mM EDTA, 0.5% SDS, 2.0% skimmed milk. The (TTAGG)₅ oligonucleotide was labeled with [³²P]ATP by T4 polynucleotide kinase (Toyobo). DNA fragments

were labeled with [³²P]dCTP by random priming using the BcaBEST DNA labeling kit (Takara). The blotted membranes were incubated at 50°C with the (TTAGG)₅ probe or at 65°C with the random primer labeled probe overnight. Thereafter, the membranes were washed in 4×, 2×, 1×, 0.5× and 0.1× SSC for 20 min each at the same temperature as the hybridization.

Fluorescence *in situ* hybridization (FISH)

Fluorescence *in situ* hybridization was performed as described using prometaphase chromosomes prepared from testes of fifth instar larvae (18,19). To obtain a labeled probe, we amplified a 1.4 kb 3′-terminal region of SART1 by PCR (22) using biotinylated dUTP (Bio-16-dUTP; Boehringer). The primer set used for PCR, 4629 (5′-CGGCACCTTGAAAATGTCCG-3′) and 6359 (5′-ACAACTGGACTATCGTGTGCG-3′), is shown in Figure 2A.

PCR amplification of SART1–(TTAGG)_n junction sites

We found that about half of the genomic copies of the SART1 element formed tandem arrays of SART1–(TTAGG)_n as a repeat unit in the silkworm genome. Thus, to identify the specific target site of the SART1 insertion, we amplified the adjacent regions of the (TTAGG)_n in the SART1–(TTAGG)_n–SART1 in the genome by PCR (22). The primers for this PCR, 6610 (5′-CGGAGTCCGACATAACCCGGTCCGA-3′) and 62 (5′-TGGAAGTCCAGCAAACTCCCCAC-3′) are shown in Figure 2B. The PCR reaction mixture (50 μl) contained 1× PCR buffer (500 mM KCl, 15 mM MgCl₂, 100 mM Tris–HCl, pH 8.3), 5 mM dNTP, 10 pmol each primer, 1.5 μg genomic DNA and 1 U *Taq* polymerase (Takara). PCR proceeded for 30 cycles under the following conditions: 94°C, 30 s; 64°C, 30 s and 73°C, 1 min. A denaturing step at 94°C for 3 min preceded the first cycle and the final cycle was followed by a further extension at 73°C for 10 min. The amplified DNA fragments were cloned into a plasmid vector with the TA Cloning Kit (Invitrogen).

Phylogenetic analysis

We aligned the amino acid sequences of reverse transcriptase domains of 18 retrotransposable elements by the neighbor joining method using the CLUSTAL V program (23,24). The region of the amino acid sequence in ORF2 of the SART1 element starts at 455 and ends at 783, as shown in Figure 2A. The monophyly of groups was assessed with 1000 bootstrap resamplings. The sources and NCBI sequence identification numbers of these sequences are as follows: I (*Drosophila melanogaster*), 85020; LIHs (human), 106903; LIMd (mouse), 130402; Tx1 (*Xenopus laevis*), 141475; R2Dm (*D.melanogaster*), 130551; R1Dm (*D.melanogaster*), 140023; R1Bm (*B.mori*), 84806; TRAS1 (*B.mori*), 940390; RT1 (*Anopheles gambiae*), 159617; RT2 (*A.arabiensis*), 159620; TART (*D.melanogaster*), 435415; jockey (*D.melanogaster*), 134083; F (*D.melanogaster*), 103353; Doc (*D.melanogaster*), 103221; T1Ag (*A.gambiae*), 103015. The sequences of ingi (*Trypanosoma brucei*) and R2Bm (*B.mori*) are from references 25 and 26 respectively.

Nucleotide sequence accession number

The nucleotide sequence data reported in this paper will appear in the GenBank, EMBL and DDBJ nucleotide sequence databases with the accession number D85594.

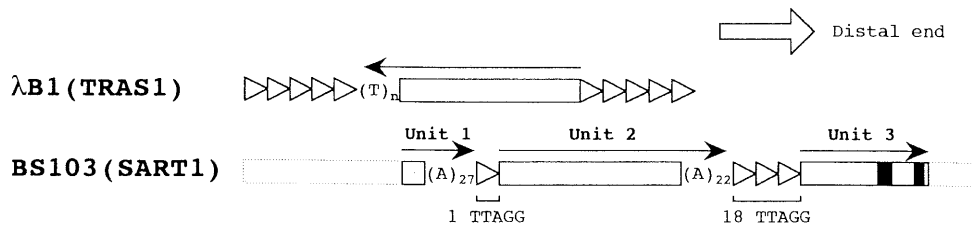


Figure 1. Comparison of schematic structure between BS103 (a SART1-containing clone) and TRAS1. BS103 contains three direct repeats of SART1 elements (units 1–3) separated by short TTAGG repeats. Open triangles represent the telomeric repeats. Open boxes indicate TRAS1 or SART1 elements with their poly(A) tails. Filled boxes show the regions in which most of the nucleotide substitutions of unit 3 compared with unit 2 were observed (see Discussion). Arrows indicate the directions of transcription of these elements. TRAS1 is transcribed from the distal to the centromere-proximal, whereas the SART1 elements is transcribed in reverse.

RESULTS

Isolation of a complete SART1 element

We first identified several clones containing SART1 from a lambda genomic library by screening with the (TTAGG)₅ probe (see Materials and Methods). The SART1 elements in these clones terminated with a poly(A) tail which is directly joined to the (TTAGG)_n repeat on the strand. Since the poly(A) tail of TRAS1 was joined to the (CCTAA)_n repeat, transcription of SART1 should be in the opposite direction to that of TRAS1 (Fig. 1). Using the 3'-terminal region of a SART1 element (the insert sequence of p51S) as a probe, we isolated 20 phage clones and identified eight that included a complete SART1 unit. The sequence of the SART1 element is highly conserved among many genomic copies, as described below. Because BS103 showed the same restriction profile as that of genomic Southern hybridization in the corresponding region, we further analyzed this clone. We found that BS103 consisted of three direct repeats of the SART1 sequence (Fig. 1). Structural analysis revealed that a 6.7 kb sequence sandwiched between two short (TTAGG)_n regions corresponds to a complete unit of the SART1. Thus, the BS103 insert consisted of a 3'-terminal region (unit 1), one complete unit (unit 2) and a 5'-half (unit 3) of the SART1 element, as described below.

Structural features of the SART1 element

Unit 2 of the SART1 element, terminating with poly(A), was 6704 bp in length and G+C rich (64%). The absence of long direct or inverted repeats at both ends indicates that the SART1 is a non-LTR retrotransposon.

An 867 bp 5' untranslated region (5'-UTR) precedes the first ORF. There was a T rich region (68%) of 122 bp in the middle of the 5'-UTR (data not shown). SART1 contained two long ORFs which occupy 80% of its total length (Fig. 2A). The first ORF (ORF1) was 2148 bp long (from 868 to 3015) and putatively encoded 712 amino acids, which is quite similar to the Gag-like protein of non-LTR retrotransposons. The first ATG codon found at the 13th base position is flanked by the translation start consensus sequences among eukaryotes (27). ORF1 is basic (isoelectric point 10.3), contains many proline residues (8%) and charged amino acids (29%), which are characteristic of Gag proteins (28). In addition, three putative zinc finger motifs (CCHC) were found near the carboxyl end of the ORF1 (Fig. 2A).

The second ORF (ORF2) is 3266 bp long (from 2952 to 6217) and putatively encodes 1067 amino acids. This amino acid sequence shows significant similarity to the Pol-like protein. The

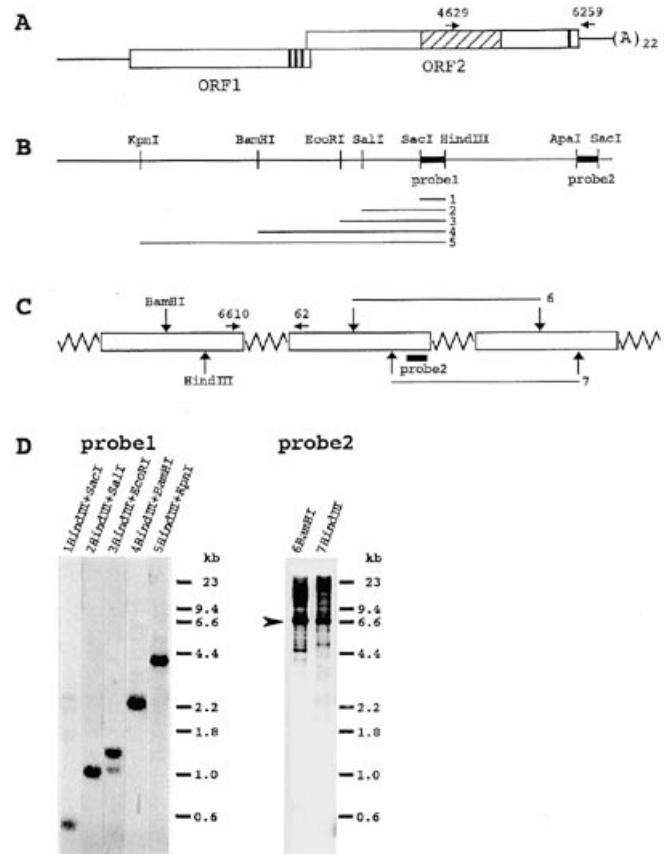


Figure 2. Conserved structure of SART1 elements based on genomic Southern hybridization. (A) Schematic structure of SART1 element. Open boxes show two ORFs of SART1 and vertical lines represent the cysteine-histidine motifs. Hatched region represents the reverse transcriptase domain, which was used to construct the phylogenetic tree in Figure 5. A PCR product used as a probe for FISH (Fig. 3) is shown as the region between two small arrows, 4629 and 6259. (B) Restriction map of SART1. Probes 1 and 2, used in the genomic Southern hybridization in (D), are shown as bold lines in the map. Five thin lines below the map represent the restriction fragments detected in the genomic hybridization. (C) Tandem array of SART1 elements (open boxes) separated by short telomeric repeats (jagged lines). Two small arrows, 6610 and 62, represent a PCR primer set to amplify the junction regions between SART1 elements (Fig. 4). Two lines, 6 and 7, correspond to a main band of 6.7 kb detected in (D) using probe 2. (D) Genomic Southern hybridization of the *Bombyx* DNA with the two probes described above. Numbers in each lane correspond to the restriction fragments illustrated in (B) and (C).

ORF2 overlaps with ORF1 by 54 nt in the -1 reading frame. Although in many retroviruses and retrotransposons, the two

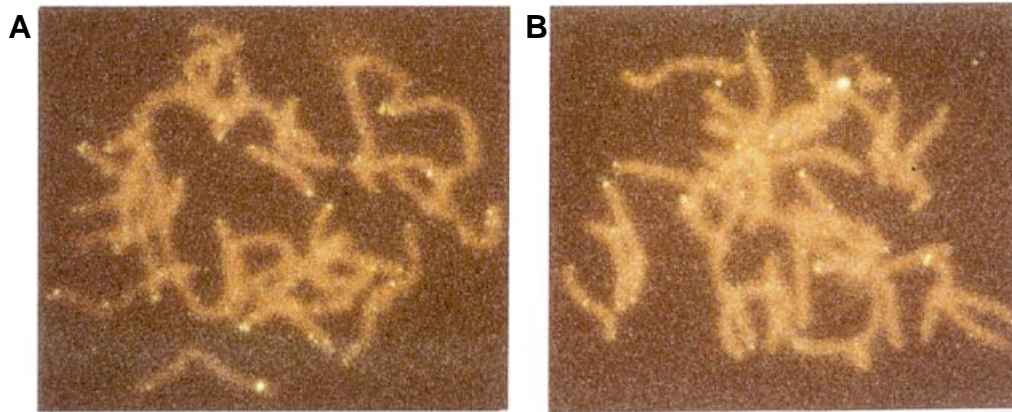


Figure 3. Chromosomal localization of SART1 elements. The probe region for this FISH is shown in (A). Chromosomes were prepared from the testes of day 2 of the 5th instar larvae of *B.mori*. Yellow signals indicate the positions of SART1 elements.

ORFs are translated as a fusion protein by ribosomal frameshifting (29,30), translation of the ORF2 in SART1 seems to start at the first ATG 66 bp from the beginning of the ORF. The N-terminal region of the putative translation product using this methionine codon, which is flanked by a consensus sequence for initiation (27), was homologous to the corresponding area of the TRAS1 and R1 ORF2 (data not shown). However, the upstream region of that N-terminal had no obvious homology to the corresponding area of the two retrotransposons.

In the ORF2, we identified a reverse transcriptase domain in the middle region and a cysteine-histidine motif near the C-terminus, as depicted in Figure 2A (1). The 886 bp 3'-UTR ends in a 27 bp poly(A) tail, although we could not find an evident polyadenylation signal (AATAAA).

Half of the genomic copies of SART1 are tandem arrays interrupted by short telomeric repeats

In three tandem SART1 units of the BS103 clone, there were 18 TTAGG repeats between units 1 and 2, but only one between units 2 and 3 (Fig. 1). The nucleotide sequence of the coding region in unit 3 is different at 23 positions from that of the corresponding area in unit 2 (data not shown). Of the 23 substitutions, 21 are synonymous and, thus, do not influence the amino acid sequence. This conservation of the protein encoding sequence provides indirect evidence that the element can still retrotranspose. To understand the distribution of SART1 elements within the *B.mori* genome, we performed genomic Southern hybridization using two DNA fragments from unit 2 shown in Figure 2B as probes. First, the genomic DNA was double-digested with *Hind*III and one of five other enzymes, blotted and hybridized with probe 1 (*Sac*I-*Hind*III fragment) (Fig. 2D, lanes 1-5). When the genomic DNA was digested with *Hind*III and *Sac*I, a 0.4 kb band, which has the same length as probe 1, was the most prominent (lane 1). This result indicates that these *Hind*III and *Sac*I sites are conserved in most SART1 copies within the genome. The results shown in lanes 2-5 and with the other 10 enzymes (data not shown) are also consistent with the restriction enzyme map for unit 2. By comparison with the signal intensity for a diluted series of DNA fragments of BS103 as a positive hybridization control (data not shown), we estimated the copy number of SART1 per haploid genome as 600. Thus, SART1 is highly conserved within

the *B.mori* genome and unit 2 in the BS103 is a representative of conserved SART1 copies.

Next, to examine whether SART1 elements form 'tandem arrays' in the genome, as in BS103, we digested the genomic DNA with *Hind*III or *Bam*HI which cuts once within unit 2 and hybridized it with probe 2. When SART1 copies are tandemly located within the genome, a major band will appear that corresponds to the complete length of the element, 6.7 kb (Fig. 2C). Figure 2D (lanes 6 and 7) shows a prominent 6.7 kb band, supporting this notion. As estimated based on the intensity of each band, at least half of the SART1 copies exist as tandem arrays in the genome. Many bands other than the 6.7 kb band were separated discretely. This indicates that most of these bands could result from insertion of a SART1 close to *Bam*HI and *Hind*III sites in flanking DNA, such as TRAS1 and other telomeric repeat-associated sequences.

We further investigated the localization of SART1 elements on the chromosomes by FISH. Most FITC signals were given by the ends of 15-20 chromosomes (Fig. 3). However, it is not yet determined whether these terminal signals are associated with specific chromosomes.

Junction sequences between SART1 and telomeric repeats revealed site-specific insertion of SART1

To investigate whether SART1 elements are actually associated with the telomeric repeats and at which position they insert, we identified 12 independent SART1 terminal regions by PCR (22). Because ~300 SART1 elements form tandem arrays as described above, we amplified the junction regions between two neighboring SART1 genomic copies, using a set of primer extending over the two copies (see Fig. 2C). The PCR products cloned into a plasmid vector (GP1-GP12) were sequenced and junction regions in SART1-(TTAGG)_n-SART1 are shown in Figure 4. J₁₂ and J₂₃ are the junction sequences between units 1 and 2 and between units 2 and 3 of the phage clone BS103 respectively. DNA sequences of the SART1 portion in these clones were essentially identical at least in the region shown in the figure. All clones contain from 1 to 20 TTAGG repeats. The length of the poly(A) tract varied from 13 to 61 nt. The junctions between the 3' termini of SART1 elements and the telomeric repeats were identical among all the clones in which poly(A) tails are connected with

	3' end	5' end
J ₁₂	TGGGATTCCCCATCGATACC (A) ₂₇	GG (TTAGG) ₁₈ TT CCGGGCCC--GGGACCTGGGCGGGCCC
J ₂₃	TGGGATTCCCCATCGATACC (A) ₂₂	GG (TTAGG) ₁ TT -----CCC--GGGACCTGGGCGGGCCC
GP1	TGGGATTCCCCATCGATACC (A) ₂₅	GG (TTAGG) ₁₉ TT -----CCC--GGGACCTGGGCGGGCCC
GP2	TGGGATTCCCCATCGATACC (A) ₁₈	GG (TTAGG) ₂₀ TT -----CCC--GGGACCTGGGCGGGCCC
GP3	TGGGATTCCCCATCGATACC (A) ₁₇	GG (TTAGG) ₁₉ TT -----CCC--GGGACCTGGGCGGGCCC
GP4	TGGGATTCCCCATCGATACC (A) ₁₃	GG (TTAGG) ₁₆ TT -----CCC--GGGACCTGGGCGGGCCC
GP5	TGGGATTCCCCATCGATACC (A) ₂₁ ^a	GG (TTAGG) ₁₇ TT -----CCC--GGGACCTGGGCGGGCCC
GP6	TGGGATTCCCCATCGATACC (A) ₂₂ ^b	GG (TTAGG) ₁₇ TT -----CCC--GGGACCTGGGCGGGCCC
GP7	TGGGATTCCCCATCGATACC (A) ₁₃	GG (TTAGG) ₁₆ TT -----CCC--GGGACCTGGGCGGGCCC
GP8	TGGGATTCCCCATCGATACC (A) ₂₀ ^c	GG (TTAGG) ₁₉ TT -----CCC--GGGACCTGGGCGGGCCC
GP9	TGGGATTCCCCATCGATACC (A) ₂₂	GG (TTAGG) ₄ TT -----CCCGAGGGACCTGGGCGGGCCC
GP10	TGGGATTCCCCATCGATACC (A) ₂₂	GG (TTAGG) ₄ ^d TT -----CCCGAGGGACCTGGGCGGGCCC
GP11	TGGGATTCCCCATCGATACC (A) ₂₉	GG (TTAGG) ₄ TT -----CCCGAGGGACCTGGGCGGGCCC
GP12	TGGGATTCCCCATCGATACC (A) ₆₁	GG (TTAGG) ₄ TT ^e -----CCCGAGGGACCTGGGCGGGCCC

Figure 4. Nucleotide sequences of SART1-(TTAGG)_n-SART1 junctions. J12 and J23 denote the junctions in the BS103 clone between units 1 and 2 and between units 2 and 3 respectively (Fig. 1). Other GP clones were obtained by PCR using a primer set shown in Figure 3C. The sequences underlined were disrupted by other nucleotides. The precise sequences are as follows: a, (A)₁₄T(A)₆; b, (A)₁₄T(A)₇; c, (A)₃CC(A)₁₇; d, (TTAGG)₂TTAAGTTAGG; e, (TTAGG)₄CTT. Dashes represent nucleotide deletions.

GG in a (TTAGG)_n unit. However, it is not clear whether the last nucleotide A of the poly(A) belongs to the SART1 element or to the telomeric repeat.

On the other hand, the 5'-ends of SART1 in these clones assume three different forms. The 5' termini of SART1 in GP9, 10, 11 and 12 had additional GA nucleotides compared with J23. In the corresponding region in J12, however, an additional CCGG was duplicated. Such 5' duplication is also found in R1 elements in *B.mori* (28). Except for the 5' duplication in J12 the telomeric repeats in the other clones terminated in the TT of a (TTAGG)_n repeat unit and the following CCC nucleotides which are presumed to be part of the SART1 element.

We also amplified a part of BS103 that had been subcloned into a plasmid as a control. In four sequenced clones we did not find additional GA nucleotides, CCGG duplication or any base substitutions, although the length of poly(A) tails and the numbers of the telomeric repeats were inaccurate (data not shown). This sequence homogeneity indicates that the sequence variety seen in Figure 4 is not due to PCR errors. These results indicate that SART1 is inserted between TT and AGG of the telomeric repeat unit in a highly specific manner.

Sequence comparison classified SART1 related site-specific elements into two groups

To understand the evolutionary origin of SART1, we constructed a phylogenetic tree among non-LTR retrotransposons by the neighbor joining method (24) based upon the amino acid sequences of the reverse transcriptase domain (Fig. 2A). As shown in Figure 5, SART1 falls into one group with R1, RT elements and TRAS1, all of which are site-specific retrotransposons (12,19,31). This relationship indicates that they derived from a common evolutionary origin and have changed the specificity for their insertion sites. These four types of retrotransposon further branched into two subgroups, R1/TRAS1 and RT/SART1. Although the bootstrap value shown in Figure 5 is not high enough to establish the subgroups, another phylogenetic tree based on sequences for the entire Gag region supported this classification (data not shown).

We compared the putative zinc finger motifs and their flanking regions between the two subgroups (data not shown). In general,

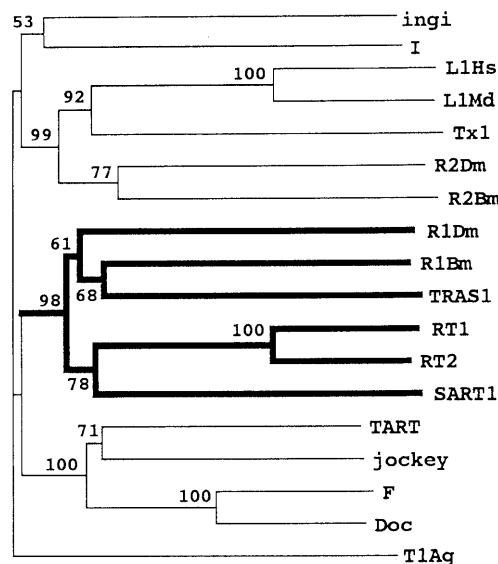


Figure 5. A Phylogenetic tree among non-LTR retrotransposable elements by neighbor joining method based on the amino acid sequence of the reverse transcriptase domain. The numbers at the nodes denote bootstrap values per 1000 replications. Only the values >50% are shown. The bold lines indicate the close evolutionary relationship among TRAS1, R1, RT and SART1 elements.

the amino acid sequences in these regions were conserved better within members of the same subgroup than between the two groups, supporting the above classification. The structural difference between the two subgroups, such as spacing Cys and His residues, is evident in the C-terminal region of the pol-ORF. The conserved cysteine-histidine motifs in the pol-ORF are HX₁₇CXCX₈₋₉HX₄C in the TRAS1 group and HX₁₃HX₁₇CX₂CX₁₃C in the SART1 group respectively. In retroviruses, an integrase domain containing an HX₃HX₂₂₋₃₂CX₂C motif follows a reverse transcriptase domain (32) and in the corresponding location, a CXCX₈₋₉HX₄C motif is generally conserved among divergent species of non-LTR retrotransposons (31). Recent studies of the site-specific integration of the group II intron also demonstrated the importance of this domain (33). The

suggestive of the site preference of SART1 insertion. Ty5 was found near the telomeric repeats of yeast chromosomes (39). This indicates SART1 and TRAS1 might prefer to insert into the telomeric repeats by recognizing proteins in the telomeric heterochromatin, as suggested for Ty5 insertion.

A phylogenetic tree based on the amino acid sequences of the reverse transcriptase domain and sequence comparisons in the zinc finger regions revealed that SART1 is more closely related to the RT elements of *Anopheles* than to any other retrotransposable elements (Fig. 5). The TRAS1 and R1 elements are closely related to the SART1/RT group but may be further categorized into another subgroup. The classification of these two subgroups based on structural similarities is not in accord with that judged by the chromosomal locus they interrupt. R1 and RT elements are inserted in the 28S rRNA gene, but TRAS1 and SART1 are inserted in the telomeric repeats (13,19,28). Besides having different target sequences, the relative orientation of the inserted elements to the target sequences is not consistent between two rDNA elements and between two telomeric repeat-associated elements (Fig. 6). The RT elements would be transcribed in the opposite direction to R1, R2 and R4 (Fig. 6A). Considering the structural similarities, R1/TRAS1 and RT/SART1 groups may have different preference for chromosomal polarity in integration. The direction of the insertion may be related to the regulation of transcription and the integration of non-LTR retrotransposons. The same direction of the R1 element relative to the rRNA gene enables this element to be transcribed in combination with 28S rRNA (28). In contrast, we believe the telomeric repeats cannot be transcribed. Thus, flanked by telomeric repeats, TRAS1 and SART1 should possess their own promoters for transcription. In 5'-UTR regions, we did not find any obvious homology between TRAS1 and SART1 except for CCCG at the extreme ends (data not shown). Another interesting possibility is that the telomeric short repeats work as promoters for the retrotransposons.

The *Drosophila* telomeres are free from short repetitive sequences. The retrotransposons, HeT-A and TART, transpose to reconstitute the ends that balance the terminal loss during DNA replication (40,41). As shown in the phylogenetic tree (Fig. 5), TRAS1 and SART1 are not closely related to HeT-A and TART. At present, we have no evidence for the role of these telomeric repeat-specific elements on telomere function. However, we consider them important in understanding how telomeres evolved in insects. These telomere-associated elements of the silkworm may be able to regulate the organization and length of the telomere, which influences telomeric heterochromatin or to work as a buffer and an alternative backup pathway as in the Y' element in yeast (42). They may also increase the number of telomeric repeats during retrotransposition by short terminal duplication of the target sequence at both ends of the element but not at the chromosomal ends. TRAS1 and SART1 might have spread in the silkworm strains through vertical transmission as implicated for R1 and R2 (43). Studies of telomeric repeat-specific retrotransposons in other species and on overall telomere structure would help in understanding the evolution of telomeres in insects.

ACKNOWLEDGEMENTS

We thank Miss Y. Nakazato for help with FISH. This work was supported by grants for helping the study of FISH. This work was supported by a grant to H.F. (Evolutionary Life Systems Research Group) from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- Eickbush, T.H. (1992) *New Biol.*, **4**, 430–440.
- Gabriel, A. and Boeke, J.D. (1993) In Skalka, A.M. and Goff, S.P. (eds), *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 275–328.
- Hutchison, C.A., Hardies, S.C., III, Loeb, D.D., Shehee, W.R. and Edgell, M.H. (1989) In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 594–617.
- Garrett, J.E., Knutzon, D.S. and Carroll, D. (1989) *Mol. Cell. Biol.*, **9**, 3018–3027.
- Aksoy, S., Williams, S., Chang, S. and Richards, F.F. (1990) *Nucleic Acids Res.*, **18**, 785–792.
- Gabriel, A., Yen, T.J., Schwartz, D.C., Smith, C.L., Boeke, J.D., Sollner-Webb, B. and Cleveland, D.W. (1990) *Mol. Cell. Biol.*, **10**, 615–624.
- Villanueva, M.S., Williams, S.P., Beard, C.B., Richards, F.F. and Aksoy, S. (1991) *Mol. Cell. Biol.*, **11**, 6139–6148.
- Eickbush, T.H. and Robins, B. (1985) *EMBO J.*, **4**, 2281–2285.
- Fujiwara, H., Ogura, T., Takada, N., Miyajima, N., Ishikawa, H. and Maekawa, H. (1984) *Nucleic Acids Res.*, **12**, 6861–6869.
- Jakubczak, J.L., Burke, W.D. and Eickbush, T.H. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3295–3299.
- Burke, W.D., Muller, F. and Eickbush, T.H. (1995) *Nucleic Acids Res.*, **23**, 4628–4634.
- Besansky, N.J., Paskewitz, S.M., Hamm, D.M. and Collins, F.H. (1992) *Mol. Cell. Biol.*, **12**, 5102–5110.
- Paskewitz, S.M. and Collins, F.H. (1989) *Nucleic Acids Res.*, **17**, 8125–8133.
- Xiong, Y. and Eickbush, T.H. (1988) *Cell*, **55**, 235–246.
- Luan, D.D., Korman, M.H., Jakubczak, J.L. and Eickbush, T.H. (1993) *Cell*, **72**, 595–605.
- Mori, M., Niemer, I. and Schmelzer, C. (1992) *Cell*, **70**, 803–810.
- Zimmerly, S., Guo, H., Perlman, P.S. and Lambowitz, A.M. (1995) *Cell*, **82**, 545–554.
- Okazaki, S., Tsuchida, K., Maekawa, H., Ishikawa, H. and Fujiwara, H. (1993) *Mol. Cell. Biol.*, **13**, 1424–1432.
- Okazaki, S., Ishikawa, H. and Fujiwara, H. (1995) *Mol. Cell. Biol.*, **15**, 4545–4552.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- Southern, E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
- Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) *Science*, **230**, 1350–1354.
- Higgins, D.G. (1994) *Methods Mol. Biol.*, **25**, 307–318.
- Saitou, N. and Nei, M. (1987) *Mol. Biol. Evol.*, **4**, 406–425.
- Kimmel, B.E., ole-Moi Yoi, O.K. and Young, J.R. (1987) *Mol. Cell. Biol.*, **7**, 1465–1475.
- Burke, W.D., Calalang, C.C. and Eickbush, T.H. (1987) *Mol. Cell. Biol.*, **7**, 2221–2230.
- Kozak, M. (1986) *Cell*, **44**, 283–292.
- Xiong, Y. and Eickbush, T.H. (1988) *Mol. Cell. Biol.*, **8**, 114–123.
- Jacks, T. and Varmus, H.E. (1985) *Science*, **230**, 1237–1242.
- Mellor, J., Fulton, S.M., Dobson, M. J., Wilson, W., Kingsman, S.M. and Kingsman, A.J. (1985) *Nature*, **313**, 243–246.
- Jakubczak, J.L., Xiong, Y. and Eickbush, T.H. (1990) *J. Mol. Biol.*, **212**, 37–52.
- Johnson, M.S., McClure, M.A., Feng, D.F., Gray, J. and Doolittle, R.F. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7648–7652.
- Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P.S. and Lambowitz, A.M. (1995) *Cell*, **83**, 529–538.
- Greider, C.W. (1995) In Blackburn, E.H. and Greider, C.W. (eds), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 35–68.
- Katz, R.A. and Skalka, A.M. (1990) *Annu. Rev. Genet.*, **24**, 409–445.
- Jakubczak, J.L., Zenni, M.K., Woodruff, R.C. and Eickbush, T.H. (1992) *Genetics*, **131**, 129–142.
- Zakian, V.A. (1995) *Science*, **270**, 1601–1607.
- Feng, Q., Moran, J.V., Kazazian, H.H., Jr and Boeke, J.D. (1996) *Cell*, **87**, 905–916.
- Zou, S., Ke, N., Kim, J.M. and Voytas, D.F. (1996) *Gene Dev.*, **10**, 634–645.
- Biessmann, H., Valgeirsdottir, K., Lofsky, A., Chin, C., Ginther, B., Levis, R.W. and Pardue, M.L. (1992) *Mol. Cell. Biol.*, **12**, 3910–3918.
- Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A. and Sheen, F.M. (1993) *Cell*, **75**, 1083–1093.
- Lundblad, V. and Blackburn, E.H. (1993) *Cell*, **73**, 347–360.
- Eickbush, D.G. and Eickbush, T.H. (1995) *Genetics*, **139**, 671–684.