MOLECULAR AND CELLULAR BIOLOGY, Aug. 1995, p. 4545–4552 Vol. 15, No. 8
0270-7306/95/$04.00+0

Structural Analysis of TRAS1, a Novel Family of Telomeric Repeat-Associated Retrotransposons in the Silkworm, *Bombyx mori*

SATOSHI OKAZAKI, HAJIME ISHIKAWA, AND HARUHIKO FUJIWARA*

Zoological Institute, Graduate School of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Received 17 February 1995/Returned for modification 30 March 1995/Accepted 15 May 1995

We characterized TRAS1, a retrotransposable element which was inserted into the telomeric repetitive sequence $(CCTAA)_n$ of the silkworm, *Bombyx mori*. The complete sequence of TRAS1, a stretch of 7.8 kb with a poly(A) tract at the 3 end, was determined. No long terminal repeat (LTR) was found at the termini of the element. TRAS1 contains *gag*- and *pol*-like open reading frames (ORFs) which are similar to those of non-LTR retrotransposons. The two ORFs overlap but are one nucleotide out of frame (+1 frameshift). Most of the approximately 250 copies of TRAS1 elements in the genome were highly conserved in the structure. Chromosomal in situ hybridization showed that TRAS1 elements are clustered at the telomeres of *Bombyx* chromosomes. A phylogenetic analysis using the amino acid sequence of the reverse transcriptase domain within the *pol*-like ORF revealed that TRAS1 falls into one lineage with R1, which is a family of non-LTR retrotransposons inserted into the same site within the 28S ribosomal DNA unit in most insects. TRAS1 may have been derived from R1 and changed the target specificity so that TRAS1 inserts into the telomeric repetitive sequence $(CCTAA)_n$. Southern hybridization and *Bal* 31 exonuclease analyses showed that TRAS1 elements are clustered proximal to the terminal long tract of $(CCTAA)_n$. TRAS1 is a novel family of non-LTR retrotransposons which are inserted into the telomeric repetitive sequences as target sites.

Telomeres are the ends of eukaryotic chromosomes that protect chromosomes from fusion, degradation, and incomplete replication. Telomeric repetitive sequences have been identified at the chromosomal ends of a number of organisms including protozoans, fungi, vertebrates, plants, and some insects (7, 39, 54). The telomeric repeat of those insects, $(CCTAA)_n$, was identified in the silkworm, *Bombyx mori*. Many other insects, including representatives from eight orders of insect species, contain related sequences in their genomes, suggesting that $(CCTAA)_n$ is a widespread telomeric sequence among insects (39).

However, there is no or undetectable amounts of $(CCTAA)_n$ in the genomes of *Drosophila melanogaster* and some other insects (39). Furthermore, short repetitive sequences have not been detected at telomeric regions of *Drosophila* chromosomes (27). Instead of a telomeric repeat, the *Drosophila* telomere is composed of families of clustered DNA elements designated HeT-A and TART (4, 28, 53). The DNA sequences of HeT-A and TART have suggested that these DNA elements are retrotransposable elements (3, 11, 28). Their transpositions to the ends of the chromosomal DNAs could compensate for the terminal nucleotide loss of the chromosomes (3). Thus, the presence of HeT-A and TART in *D. melanogaster* suggests that retrotransposons may act as substitutes for telomeric repeats which are synthesized by telomerase activity.

In our previous report, we have identified a telomeric repeat-associated sequence which has a poly(A) tract directly adjacent to the telomeric repeat $(CCTAA)_n$ (39). We designated this poly(A)-bearing DNA element TRAS1 (telomeric repeat-associated sequence 1). In this study, we report the detailed structure of the TRAS1 element. Structural analysis demonstrated that TRAS1 is a family of non-long terminal repeat (LTR) retrotransposons, which are well-known as LINE-like elements in eukaryotes (16, 19, 49).

MATERIALS AND METHODS

***B. mori* strain.** A *B. mori* strain, Kinshu × Showa, purchased from Kyodo-shiryo Co., Tokyo, Japan, was used.

DNA preparation. Genomic DNAs were prepared from posterior silk glands which were dissected from fifth-instar larvae. Isolated posterior silk glands were homogenized with a glass homogenizer in 20 mM Tris-HCl (pH 8.9)–50 mM EDTA–0.5% sodium dodecyl sulfate (SDS) at 4°C. Proteinase K was added to a final concentration of 0.5 mg/ml, and the homogenate was incubated at 50°C for 12 h. After two phenol extractions and two chloroform extractions, nucleic acids were precipitated with ethanol. Fibrous DNA was collected with a glass rod and was dissolved in 20 mM Tris-HCl (pH 7.5)–50 mM EDTA. After RNase A treatment, DNA was reextracted twice with phenol-chloroform, precipitated with ethanol, and dissolved in TE (10 mM Tris-HCl [pH 8.0], 1 mM EDTA).

Genomic Southern blots and hybridization. DNA samples digested with appropriate enzymes were electrophoresed on 0.9% agarose gels and blotted onto nitrocellulose filters (BA85; Schleicher & Schuell) in 20× SSC (1× SSC is 0.15 M NaCl plus 15 mM sodium citrate) by capillary transfer by the method of Southern (44). Hybridization was performed in 0.9 M NaCl–90 mM Tris-HCl (pH 7.9)–6 mM EDTA–0.5% SDS–2.0% skim milk. When a synthetic oligonucleotide $(TTAGG)_5$ was used as a probe, the probe was labeled by phosphorylation with [γ-^{32}P]ATP by bacteriophage T4 polynucleotide kinase (Toyobo). When DNA fragments from plasmid clones were used as probes, the radiolabeled probes were obtained by incorporation of [α-^{32}P]dCTP with random priming reaction, using the *Bca*BEST DNA labeling kit (Takara). Hybridization was performed either at 50°C [for $(TTAGG)_5$ probe] or 65°C (for random-primer-labeled probes). Filters were washed in 4× SSC and then in 2× SSC, 1× SSC, and 0.5× SSC (and 0.1× SSC when a random-primer-labeled probe was used). Each step was taken for 20 min at the same temperature as in the hybridization.

Screening for the genomic lambda phage library and subcloning. An EMBL3 genomic DNA library has been constructed from partial *Sau*3AI digests of the *Bombyx* genomic DNA (39). Clones containing $(CCTAA)_n$ repeats were isolated with the ^{32}P-labeled $(TTAGG)_5$ probe. Restriction fragments containing telomeric repeats were further subcloned into pBluescript (Stratagene). Hybridization conditions are described in the preceding paragraph.

DNA sequencing. Sequencing of cloned DNA was performed either directly or on deletion derivatives prepared with exonuclease III. Deletion series were generated by using a commercial kit, Kilo-Sequence Deletion Kit (Takara). Sequencing was carried out by the dideoxynucleotide chain termination method (43) with either the Sequenase kit (United States Biochemical) or *Bca*BEST sequence kit (Takara). The open reading frame (ORF) search and dot matrix

* Corresponding author. Phone: 81-3-3812-2111, ext. 4447. Fax: 81-3-3816-1965.

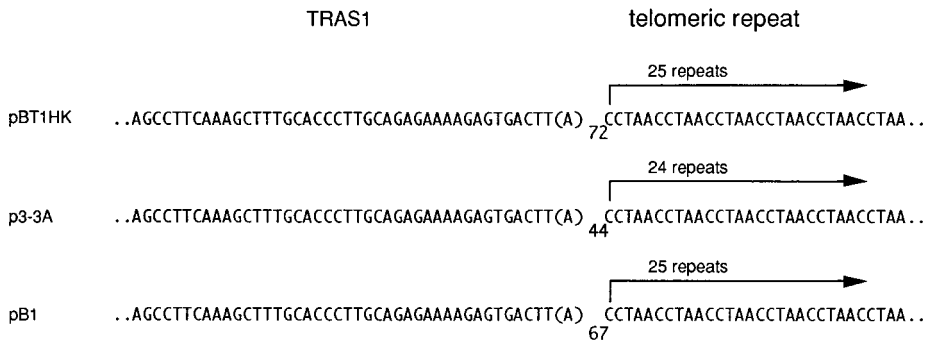

FIG. 1. Boundary sequences between the TRAS1 elements and telomeric repeats. The lambda clones containing TRAS1 elements were screened from the *Bombyx* genomic library, and the fragments which include the boundary between TRAS1 and telomeric repeat were further subcloned into plasmids and sequenced.

comparisons were performed with DNASIS software version 7.00 (Hitachi Software Engineering Inc.).

In situ hybridization of *Bombyx* chromosomes with the TRAS1 element. In situ hybridization was performed basically as described in our previous paper (39). Prometaphase chromosomes prepared from testes were denatured by immersion in 70% formamide in 2× SSC at 70°C for 3 min and immediately dehydrated through a cold ethanol series. To obtain a labeled probe, the corresponding region in the clone λB1 shown in Fig. 2A was amplified by PCR with biotinylated dUTP (Bio-16-dUTP; Boehringer). The primer pair used for PCR was 5′-CAAAGCGGCACTCCTCACAG-3′ and 5′-TTCTCTGCAAGGGTGCAAAG-3′. Prior to incubation, the DNA probe was denatured at 95°C for 5 min. The hybridization solution consisted of 0.2 μg of biotin-labeled probe per ml, 50% formamide, and 10% dextran sulfate in 2× SSC. The slide was incubated in a moist chamber at 40°C overnight, rinsed in three changes of 50% formamide in 2× SSC for 3 min per rinse at 40°C, and then rinsed three times in 2× SSC for 3 min per rinse at 40°C. To avoid high background levels, the slide was rinsed once with 50 μl of blocking buffer containing 3% bovine serum albumin, 0.1% Tween 20, and 10 mM $MgSO_4$ in 4× SSC and then incubated with 50 μl of 1/50 dilution of fluorescein isothiocyanate (FITC)-streptavidin (Vector Laboratories) in blocking buffer for 30 min at 37°C. The intensity of the biotin-linked fluorescence was amplified by adding a layer of FITC-antistreptavidin goat serum (Vector). The slides were observed under a fluorescence microscope after the slides were mounted with a fluorescence antifade solution including DNA counterstain (propidium iodide).

***Bal* 31 exonuclease analysis.** Fourteen micrograms of *Bombyx* DNA was dissolved in 75 μl of *Bal* 31 reaction buffer (20 mM Tris-HCl [pH 7.2], 600 mM NaCl, 12.5 mM $MgCl_2$, 1 mM EDTA), 5 U of *Bal* 31 exonuclease (Takara) was added, and the mixture was incubated at 30°C. After 0, 5, 15, 30, 45, or 60 min, 10 μl of the solution was recovered and extracted once by phenol-chloroform and precipitated with ethanol. These samples were blotted onto nitrocellulose filters and hybridized with either ^{32}P-labeled $(TTAGG)_5$ or the 0.7-kb *Eco*RV-*Hin*dIII fragment of TRAS1 in λB1 (the same probe used in the experiment shown in Fig. 7B). Hybridization conditions were as described above. After autoradiography, each signal intensity was quantified by scanning the film with a densitometer (CS-9000; Shimadzu Co.) and integrating the optical density.

Phylogenetic analysis. A conserved region of amino acid sequences of retrotransposons was aligned with the CLUSTAL V program (18). The analyzed region of TRAS1 consisted of positions 467 to 787 of ORF2. The sequence (total of 321 amino acids) was aligned with the corresponding regions of other non-LTR retrotransposons. The alignment is available upon request. In a multiple alignment of amino acid sequences, any positions which include gaps in the sequences were excluded from the analysis. A phylogenetic tree was inferred by the neighbor-joining method (42) with the CLUSTAL V program. A total of 256 positions of the alignment (without the gaps mentioned above) were used for the calculation. Monophyly of groups was assessed with 100 replications of bootstrap resampling (14).

The sources and NCBI sequence identification numbers of the sequences compared in the phylogenetic analysis are as follows: L1Hs (human, L1), 106903; L1Md (mouse, L1), 130402; Tx1 (*Xenopus laevis*), 141475; R2Dm (*D. melanogaster*, rRNA gene [rDNA] insertion), 130551; R1Dm (*D. melanogaster*, rDNA insertion), 140023; R1Bm (*B. mori*, rDNA insertion), 84806; TART (*D. melanogaster*, clustered at telomere), 435415; jockey (*D. melanogaster*), 134083; F (*D. melanogaster*), 103353; Doc (*D. melanogaster*), 103221; I (*D. melanogaster*), 85020; and T1Ag (mosquito [*Anopheles gambiae*]), 103015. The sequence of R2Bm (*B. mori*, rDNA insertion) was from reference 9.

Nucleotide sequence accession numbers. The nucleotide sequence data reported in this article for TRAS1 will appear in the GSDB, DDBJ, EMBL, and NCBI nucleotide sequence databases with accession number D38414.

RESULTS

Restriction enzyme mapping of major copies of TRAS1 elements. A TRAS1 element was originally selected from a *Bombyx* genomic DNA library as a clone containing the internal telomeric repeat (39). Southern analysis of whole *Bombyx* genomic DNA indicated that sequences within the TRAS1 element are moderately repeated within the *Bombyx* genome (see "TRAS1 is associated with the telomeric repeat"). We analyzed three clones containing TRAS1 elements isolated from the *Bombyx* genomic library. Sequence analysis revealed that all of the three clones also contain telomeric repeats directly adjacent to the poly(A) tracts (Fig. 1). One clone, λB1, was further analyzed. Figure 2A shows the structure of the TRAS1 element in λB1. The TRAS1 element is defined as the region between telomeric repeats, $(CCTAA)_n$. To determine whether the TRAS1 element in λB1 is representative of TRAS1 elements, Southern hybridizations were performed with parts of λB1 as the probes. The *Bombyx* genomic DNA was initially digested with an appropriate restriction enzyme (*Eco*RV for probe 1 and *Bam*HI for probe 2) and then with various enzymes. If the structure of each TRAS1 element in the genome is conserved, one prominent band would appear in each lane, reflecting the structure of the major copies of TRAS1 elements. In fact, prominent bands were generated in several lanes by hybridization (Fig. 2B and C). The position of each band was compared with the structure of λB1, and the corresponding restriction site in λB1 was indicated by numbering. When 0.4 kb of the *Eco*RV-*Hin*dIII fragment of λB1 (probe 1) was used as a probe, identical restriction fragments were detected by Southern hybridization of the genomic DNA digested with *Eco*RV plus *Hin*dIII (Fig. 2B, band 1). Bands 2, 3, 5, 6, and 7 are all consistent with the predicted structure of λB1. Only band 4 was not expected from the structure of λB1, because its position corresponded with the second *Sac*I recognition site from the *Eco*RV site of λB1 (Fig. 2A, position 4). In the *Eco*RV+*Sac*I lane in Fig. 2B, two dominant bands with similar signal intensities were detected. This result showed that about the half of TRAS1 copies lack the *Sac*I site at position 2 in Fig. 2A. The positions of the other restriction enzyme sites were identical within the area examined among different TRAS1 copies. Figure 2C shows the result of a similar analysis when another region in λB1 was used as a probe (probe 2). All of the prominent bands detected in each lane were consistent with the structure of λB1. In the lanes of *Sac*I-plus-*Bam*HI and *Bam*HI digestion, there were no prominent bands. This suggested that no more recognition sites for these enzymes are included in TRAS1 elements and that the various fragments

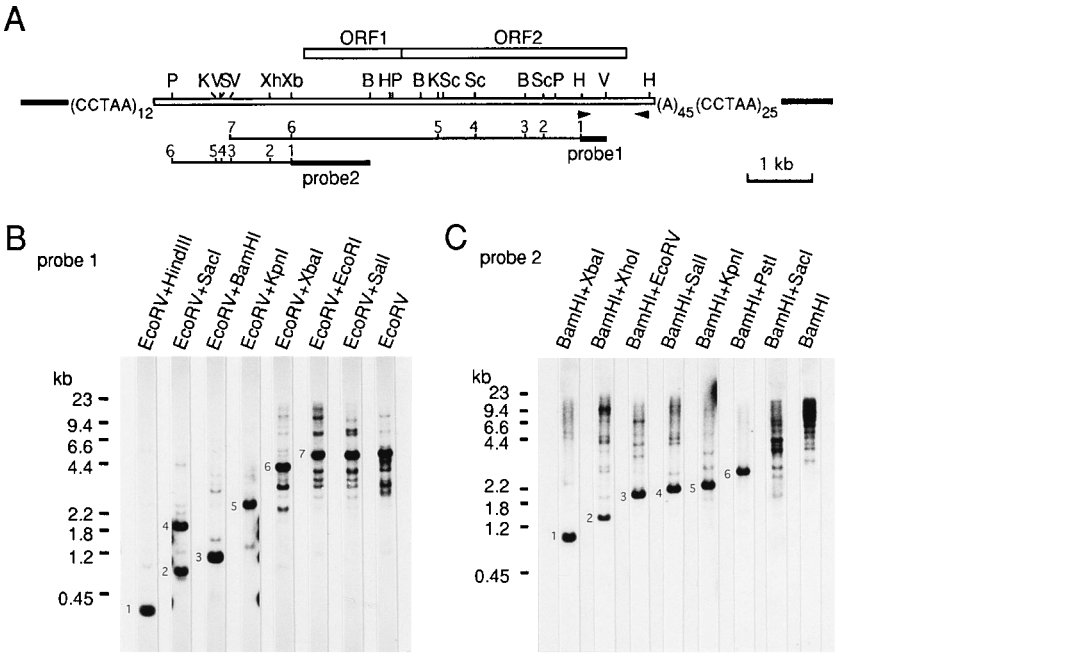

FIG. 2. Conservation of the structure of TRAS1 elements in the *Bombyx* genome. (A) Restriction map of a TRAS1 element within λB1. The locations of ORFs which were identified by sequence analysis are shown above the map. The regions of probes used for Southern hybridizations are shown below the map as solid boxes (probe 1 and probe 2). The positions of the numbers on the lines indicated beside the probe regions correspond to the restriction fragment length detected in the Southern hybridizations shown in panels B and C. A PCR product used as a probe for in situ hybridization shown in Fig. 6 is indicated as the region between the two arrowheads. Abbreviations for restriction sites used in the map: P, *Pst*I; K, *Kpn*I; V, *Eco*RV; S, *Sal*I; Xh, *Xho*I; Xb, *Xba*I; B, *Bam*HI; H, *Hin*dIII; Sc, *Sac*I. (B) Genomic Southern hybridization of *Bombyx* DNA with probe 1. For each lane, 2 μg of genomic DNA was doubly digested with *Eco*RV and another enzyme as indicated above the lane. The digested DNA was fractionated on a 0.9% agarose gel and transferred to a filter. The blot was hybridized with ^{32}P-labeled probe 1. Numbers at the left of each signal correspond to the positions in the TRAS1 clone in λB1 shown in panel A. (C) Genomic Southern hybridization of *Bombyx* DNA with probe 2. For each lane, genomic DNA was digested with *Bam*HI and another enzyme as indicated above the lane.

were generated by digestion of sequence outside the TRAS1 elements, where the sequences are not conserved. These results were predicted from the structure of λB1, because there are no recognition sites near the left end of the TRAS1 element in λB1. These experiments demonstrated that most copies of TRAS1 elements in the genome share a highly conserved structure and that λB1 contains a representative of the TRAS1 element.

Sequence analyses of TRAS1 suggest that TRAS1 is a non-LTR retrotransposon. The sequence analyses of the TRAS1 element in λB1 revealed that the TRAS1 element is 7,850 bp long followed by a 45-bp poly(A) sequence at the 3′ end of the element (Fig. 2A). The TRAS1 element contains two long ORFs. The first ORF (ORF1) is 1,419 bp long, putatively coding for 473 amino acids. The second ORF (ORF2) is 3,675 bp long, putatively coding for 1,225 amino acids. ORF1 and ORF2 showed significant similarity to the Gag- and Pol-like proteins of non-LTR retrotransposons, respectively. A putative reverse transcriptase (RT) domain was identified in the *pol*-like ORF (ORF2). There were no long direct or inverted repeat sequences at the termini of the element, indicating that TRAS1 is a non-LTR retrotransposon. ORF2 overlaps with ORF1 by 21 nucleotides in the +1 reading frame. In the overlapping region, we found a potential stem-loop structure (Fig. 3; see Discussion).

The 7.9-kb TRAS1 element is longer than most non-LTR retrotransposons, which are usually 5 to 6 kb. TRAS1 has a long noncoding region in the 5′ end of the element. No obvious ORF was found in the first 2.4 kb of TRAS1, which contains an oligo(A) (36-nucleotide) and $(CA)_n$ dinucleotide alternating sequence (82 nucleotides including incomplete repeats) (data not shown).

Amino acid sequence of TRAS1 compared with those of other retrotransposons. Non-LTR retroposons are a divergent group of retroelements throughout eukaryotes (37, 52). An oligo(A) or AT-rich tract is often found at the 3′ end. A phylogenetic study of the RT-like sequence of retroelements has shown that the non-LTR retrotransposons fall into one group, suggesting that they have a common evolutionary origin (37, 51, 52). To understand the phylogenetic relationship of TRAS1 with other elements, we compared the TRAS1 amino acid sequence to those of other non-LTR retrotransposons. Within the ORFs of retrotransposons, the RT domain in the *pol*-like ORF is the most conserved region (51). The amino acid sequences of RT domains of several non-LTR retrotransposons were aligned with that of TRAS1 (data not shown). The region analyzed consisted of positions 467 to 787 of the

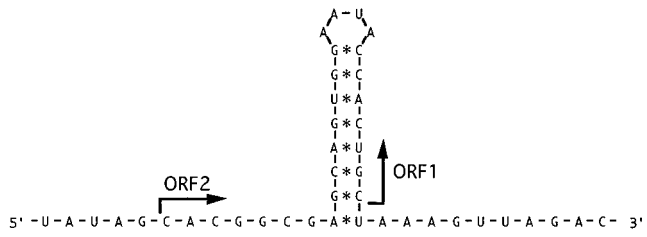

FIG. 3. A potential stem-loop structure in the overlapping region of ORF1 and ORF2.

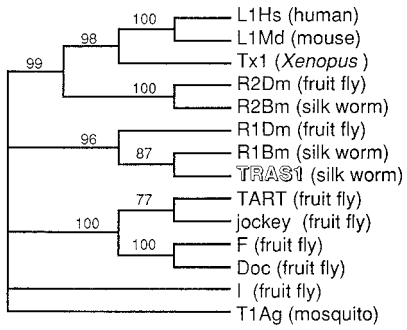

FIG. 4. A phylogeny of aligned amino acid sequences of RT domains of non-LTR retrotransposons calculated by the neighbor-joining method (42). The retrotransposons compared here are as follows: L1Hs (human, L1), L1Md (mouse, L1), Tx1 (*Xenopus laevis*), R2Dm (*D. melanogaster*, rDNA insertion), R2Bm (*B. mori*, rDNA insertion), R1Dm (*D. melanogaster*, rDNA insertion), R1Bm (*B. mori*, rDNA insertion), TART (*D. melanogaster*, clustered at telomere), jockey (*D. melanogaster*), F (*D. melanogaster*), Doc (*D. melanogaster*), I (*D. melanogaster*), and T1Ag (mosquito [*Anopheles gambiae*]). Bootstrap values based on 100 replications are shown above the branches. Only nodes found in >50% of the replicates are shown. The horizontal lengths of each branch have no significance.

TRAS1-ORF2 (from SFTV to DKLT) and includes the four conserved amino acid motifs identified by Poch et al. (40). A phylogenetic tree constructed from the alignment is shown in Fig. 4.

As shown in the tree, TRAS1 forms a subbranch with R1 retrotransposons. Within the compared region, TRAS1 had 35% identity to R1Bm (R1 of *B. mori*) and 31% to R1Dm (R1 of *D. melanogaster*). R1 is a family of retrotransposons identified in the rRNA genes of most insects (24, 25, 49). The monophyly of this subbranch was supported by a high bootstrap value (96%). On the other hand, TART, a retrotransposon identified in the *Drosophila* telomere, is rather similar to the jockey element (28), forming a subbranch with F and Doc elements. On the basis of the sequences of their RT domains, TRAS1 and TART are not closely related.

The overall sequence similarity between TRAS1 ORFs and R1 ORFs was visualized by dot matrix comparison (Fig. 5). Matched regions between TRAS1 and R1 are located in the upper region of the *pol*-like ORF, including the whole RT region, and around the cysteine-histidine motif (zinc finger domain) at the carboxyl-terminal end of the *pol*-like ORF. In the *gag*-like ORF, there are matched regions around the carboxyl-terminal region of the ORF.

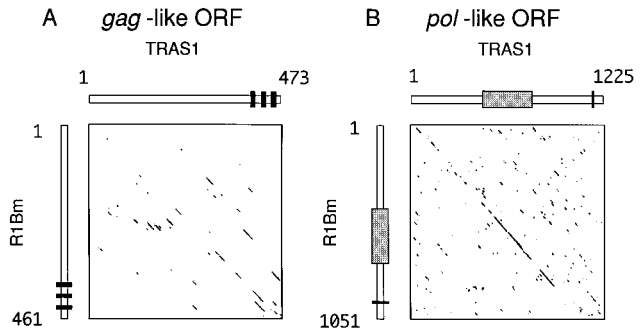

FIG. 5. Dot matrix comparisons of amino acid sequences between TRAS1 and R1Bm. The coincidence of 7 of 27 amino acids was taken as a criterion of homology. (A) The *gag*-like ORFs of TRAS1 and R1Bm are shown on the horizontal and vertical axes, respectively. The cysteine-histidine motifs within each ORF are indicated beside the matrix by the solid boxes. (B) The *pol*-like ORFs of TRAS1 and R1Bm are shown. RT regions and cysteine-histidine motifs are indicated beside the matrix by the shaded boxes and solid boxes, respectively.

TRAS1 is associated with the telomeric repeat. To know whether the TRAS1 elements are actually associated with telomeric repeats $(CCTAA)_n$, the *Bombyx* genomic DNA library was screened with TRAS1 and the telomeric repeat simultaneously. Plaque hybridization of a *Bombyx* genomic DNA library was done with two probes, $(TTAGG)_5$ and TRAS1 (probe 1 in Fig. 2A). A lambda phage library was screened with the two probes, and positive plaques were counted. A total of 255 plaques hybridized with probe $(TTAGG)_5$ but not with TRAS1, 20 plaques hybridized with $(TTAGG)_5$ and TRAS1, and 7 plaques hybridized with TRAS1 but not with $(TTAGG)_5$. Each phage clone carries approximately 10 kb of *Bombyx* genomic DNA. The total number of plaques screened was 5.5×10^3. From the above data, it is estimated that there are approximately 200 copies of TRAS1 associated with TTAGG repeats and 70 copies that are not. This result demonstrated that most TRAS1 elements are associated with the telomeric repeat in the *Bombyx* genome.

The localization of TRAS1 elements on the *B. mori* chromosomes was identified by fluorescent in situ hybridization using the biotin-labeled probe shown in Fig. 2A (Fig. 6). The fluorescence signals were seen mainly on the chromosomal tips at about 20 of 56 termini (28 chromosomes per haploid genome). There were some signals in the internal regions, but it remains to be determined whether these are specific internal signals. These results demonstrate that the TRAS1 elements are basically located at the telomeric regions of *Bombyx* chromosomes.

Subtelomeric location of TRAS1. In *B. mori*, there are several kilobases of a telomeric repeat $(CCTAA)_n$ at the terminal regions of the chromosomes (39). To investigate whether TRAS1 elements interrupt this terminal tract of the telomeric repeat, Southern hybridization and *Bal* 31 exonuclease digestion were performed.

To estimate the length of the telomeric repeat in the chromosomes, the *Bombyx* genomic DNA was treated with a restriction enzyme with 4-bp recognition sites. Since the tract of the telomeric repeat $(CCTAA)_n$ contains no recognition sequence for restriction enzymes, the telomeric repeat tract remains uncut. We digested the *Bombyx* DNA with the frequently cutting enzymes *Hin*fI, *Sau*3AI, *Hae*III, and *Hha*I, and Southern hybridization was performed with $(TTAGG)_5$ as a probe (Fig. 7A). Hybridization was observed for fragments ranging from 4 to more than 20 kb, indicating that there are relatively long, uninterrupted tracts of telomeric repeats.

Next, we estimated the length of the telomeric repeats flanking TRAS1 elements. When the genomic DNA is digested with the enzymes used above, TRAS1-flanked telomeric repeats will also be excised by digestion at both flanking regions of the telomeric repeats. One digestion site is within a TRAS1 element. The recognition sites of *Hin*fI, *Sau*3AI, *Hae*III, and *Hha*I are not located within the last 0.7-kb part of TRAS1 (Fig. 7C). We digested the *Bombyx* genomic DNA with each of these enzymes and performed Southern hybridization with an *Eco*RV-*Hin*dIII fragment of the 3′ part of TRAS1 as a probe (Fig. 7B). The size of the signal should reflect the length of the TRAS1-flanked telomeric repeat. In contrast to the result shown in Fig. 7A, the signals indicated shorter fragments (about 1 kb). If we take into account the non-telomeric-repeat region, the TRAS1-flanked telomeric repeat should be less than 500 bp long. This result is consistent with the sequence structure, a 120- to 125-bp telomeric repeat, observed in the cloned fragments (Fig. 1). These data suggested that most

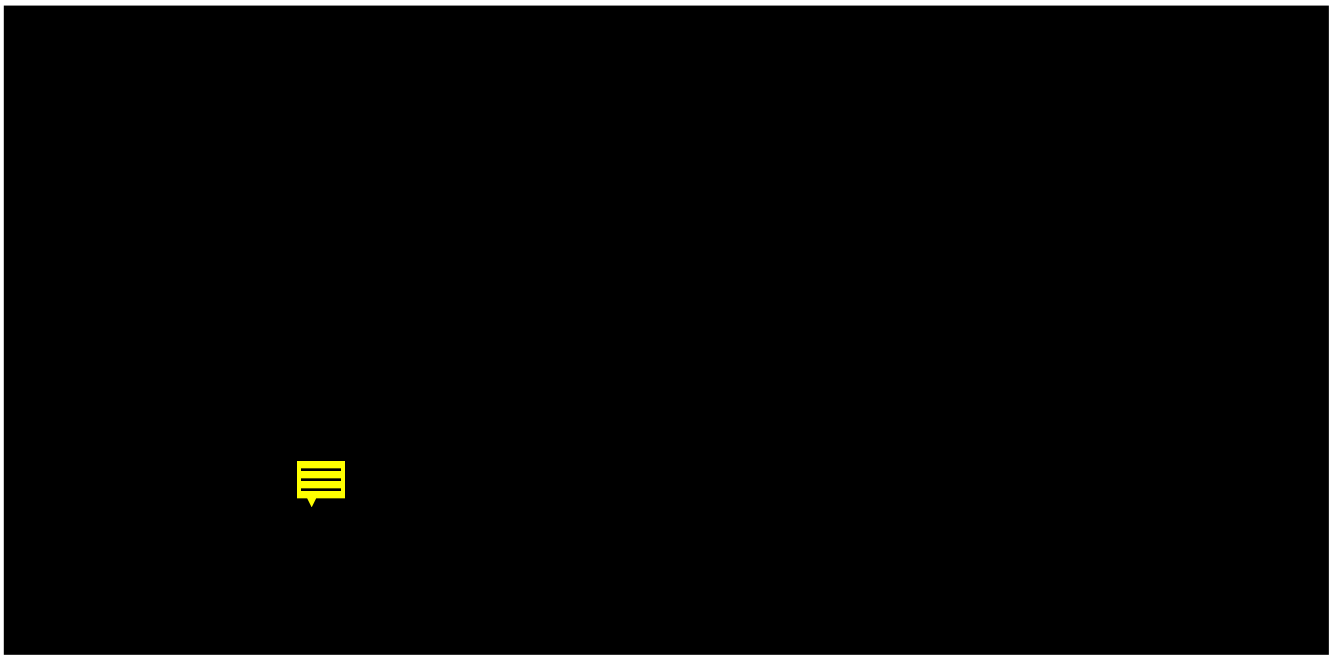

FIG. 6. Chromosomal localization of TRAS1 elements. Chromosomes were prepared from the testes of 2-day-old fifth-instar larvae of *B. mori*. The probe region used is shown in Fig. 2A. Positions of the elements are shown by yellow signals.

TRAS1 elements are not adjacent to long telomeric repeats at the chromosomal ends.

The positional relationship between telomeric repeats and TRAS1 elements was further studied by means of *Bal* 31 exonuclease sensitivity. *Bombyx* genomic DNA was digested with *Bal* 31, blotted onto nitrocellulose filters, and hybridized with the telomeric repeat and the TRAS1 element as probes. Figure 8 shows the signal intensity at each time point of *Bal* 31 digestion. The results showed that signals for the telomeric repeat were quickly weakened by *Bal* 31, whereas those for TRAS1

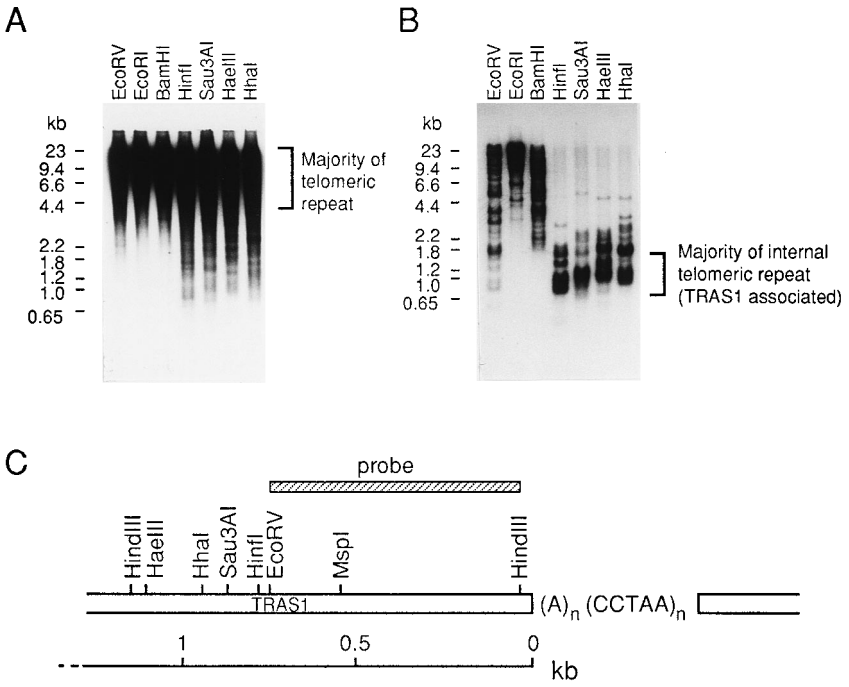

FIG. 7. Estimation of the length of telomeric repeats. (A) Southern hybridization of the *Bombyx* genomic DNA probed with the ^{32}P-labeled telomeric repeat $(TTAGG)_5$. Restriction enzymes used are indicated above each lane. (B) Southern hybridization of *Bombyx* genomic DNA probed with the 0.7-kb *Eco*RV-*Hin*dIII fragment near the 3′ end of the TRAS1 element. Restriction enzymes used are indicated above each lane. (C) Restriction map of the 3′ part of TRAS1. The scale below the map shows the distance from the 3′ end of TRAS1. The 0.7-kb *Eco*RV-*Hin*dIII fragment was used as a probe for the hybridization shown in panel B.

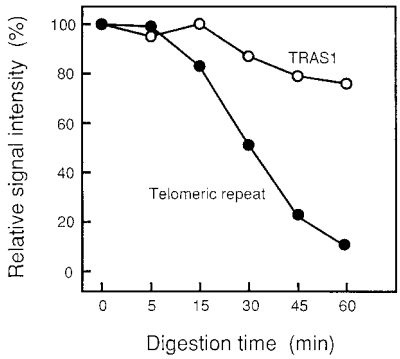

FIG. 8. *Bal* 31 exonuclease sensitivities of telomeric repeat and TRAS1. High-molecular-weight DNA from *B. mori* was subjected to *Bal* 31 exonuclease digestion for 0, 5, 15, 30, 45, and 60 min and then blotted on filters and hybridized with the telomeric repeat $(TTAGG)_5$ or the 0.7-kb *Eco*RV-*Hin*dIII fragment near the 3′ end of the TRAS1 element which is the same region used in Fig. 7B. Each signal intensity was quantified by scanning autoradiographs with a densitometer, and each signal intensity with respect to the intensity at 0 min was shown in a graph.

elements were not affected. After 60 min, most of the telomeric repeats were lost, whereas TRAS1 elements were little influenced. This result suggested that the majority of telomeric repeats are more distal in the chromosomes than TRAS1 elements. On the basis of these observations, we concluded that TRAS1 elements are clustered at subtelomeric regions, proximal to the terminal long tracts of the telomeric repeat.

DISCUSSION

We determined the complete structure of the telomeric repeat-associated retrotransposon TRAS1. TRAS1 contains *gag*- and *pol*-like ORFs which are hallmarks of retrotransposons. Southern blots showed that most copies of the TRAS1 element in the genome are conserved in structure.

ORF2 may be translated by frameshifting. The *pol*-like ORF of TRAS1 overlaps the *gag*-like ORF in the +1 reading frame (data not shown). The ORFs which encode Gag and Pol proteins often overlap in retroviruses and retrotransposons (23, 36, 41). In some of these elements, the *pol* ORF is translated as a Gag-Pol fusion protein by readthrough frameshifting (23, 36, 41). While the frameshift is often in the −1 direction as shown in Rous sarcoma virus, human immunodeficiency virus, and bovine immunodeficiency virus, there are also examples of +1 frameshifting as shown in Ty retrotransposons of *Saccharomyces cerevisiae* (10, 38, 48). We found a potential stem-loop structure in the overlapping region (Fig. 3). Similar structures are also present in many retroviruses (8, 20–22). Such stem-loop structures are thought to be involved in frameshifting by stalling translating ribosomes, thereby increasing the chance of a tRNA slippage event. Although the first ATG codon of ORF2 appears at position 64, the amino acid sequence of positions 35 to 57 of ORF2 of TRAS1 is similar to that of R1Bm or R1Dm (data not shown), suggesting that ORF2 is translated as a fusion protein with ORF1 by readthrough frameshifting.

TRAS1 is a site-specific retrotransposon closely related to R1. Whereas most retroelements are dispersed throughout chromosomes, TRAS1 is preferentially inserted at specific sites. TRAS1 is inserted into the telomeric repeat $(CCTAA)_n$, suggesting that the telomeric repetitive sequence is the target for TRAS1 elements. The insertion site of TRAS1 seems to be very strict, because the 3′ ends of three clones of TRAS1 are joined exactly with the 5′ ends of CCTAA in the same manner (Fig. 1). TRAS1 may recognize the telomeric repeat and transpose into the precise position of the sequence. Such precise integration of retrotransposons has been demonstrated in R1, R2 inserted within rDNA genes of insects (13, 15), and the retrotransposons inserted within the spliced-leader RNA genes of trypanosomatids (2, 17, 47). Each site-specific retrotransposon has a target site in the conserved sequence of the host genome. Phylogenetic analysis showed that TRAS1 and R1 are closely related (Fig. 4), suggesting an evolutionary relationship between R1 and TRAS1 elements. In R2, the functional expression of the endonuclease encoded by R2, which cleaves the target site, has been demonstrated (32, 50). We thus propose that TRAS1 may also encode a site-specific endonuclease in the ORF(s).

TRAS1 is the only known retrotransposon that is inserted into the telomeric repeat of eukaryotes. There may be other species which have retrotransposons at their telomeres. In insects, various families of non-LTR retrotransposons have been identified. More than 10 families of non-LTR retrotransposons in *D. melanogaster* have been identified, and several have been found in *B. mori* (52). Some are dispersed in the genome, and others are clustered at a defined region such as the rDNA or telomere. For instance, three families of non-LTR retrotransposons, R1, R2, and G, are inserted into *Drosophila* rDNA. It is speculated that they have independently evolved their insertion specificity for the ribosomal locus (50). The telomeric region is probably one of the most harmless targets in the complex genome. In this context, it is not surprising that other retrotransposons have evolved independently at telomeres of insects.

TRAS1 is located in the subtelomeric region. The retrotransposons HeT-A and TART are thought to be functional components of *Drosophila* telomeres. Indeed, the ends of terminally deficient chromosomes produced spontaneously or by X-ray irradiation are elongated by transfer of HeT-A or TART to the broken termini of the chromosomes (5, 6, 43a, 46). We assessed the evolutionary relationship between TRAS1 and TART by phylogenetic means (Fig. 4). The results showed that TART and TRAS1 belong to somewhat distinct lineages. The tree in Fig. 4 does not include HeT-A, because HeT-A seems to contain only a *gag*-like ORF (3, 11). When we compared the *gag*-like ORF of HeT-A with other retrotransposons, HeT-A was more similar to jockey or F than to TRAS1 or R1 (data not shown). From these results, we speculate that TRAS1 and the two retrotransposons at *Drosophila* telomeres have evolved independently to locate at telomeres of host genome.

The transpositions of HeT-A and TART are thought to be a direct addition to the ends of the chromosomes (3, 4, 28). However, this may not be the case for TRAS1, because these elements tend to be located at the proximal end of the telomeric repeat tract. This observation suggests that TRAS1 is not added to the end of the chromosome but inserts into the telomeric repeats. Figures 7 and 8 suggested that TRAS1 is inserted into the restricted regions proximal to the telomeric repeat tract. This type of insertion into the subtelomeric region has also been shown in the subtelomeric region of P elements of *D. melanogaster* (1, 26). It is unclear why TRAS1 is not inserted into the terminal long tract. One possible explanation is that the presumed specialized telomeric structure in lower and higher eukaryotes (17a, 35, 45, 48a) or the attachment to the nuclear matrix (12) may prevent the insertion of TRAS1 into the chromosomal tip. Another possibility is that the telomere elongation by telomerase or a recombinational event may rearrange TRAS1 in the subtelomeric region after random integration of TRAS1 into the telomeric repeat. The long

terminal repeats imply that the telomeric repeat tracts of *B. mori* are synthesized by telomerase and remain functional.

However, this fact does not rule out the possibility that TRAS1 elements share some functions in order to back up the telomerase-telomeric repeat system as demonstrated in the Y′ repeat of *S. cerevisiae*. Similar to TRAS1, Y′ repeats are located just proximal to the yeast telomeric repeat $(C_{1-3}A)_n$. The *est1* mutant of *S. cerevisiae* results in defective telomere elongation leading to an increase in the frequency of cell death (34). However minor *est1* mutant survivors have arisen as a result of the amplification and acquisition of Y′ repeats (33). The bulk of Y′ repeats may provide a buffer that protects the chromosome termini from progressive loss. This finding has demonstrated that even when the telomerase-telomeric repeat system is defective, there is an alternative pathway which can restore telomere function in *S. cerevisiae*. Although the propagation of Y′ may be responsible for recombination rather than transposition, Y′ has some features that shows it is or has been a mobile element (29–31). It will be of interest to study the functional role of TRAS1 in telomere formation and maintenance.

ACKNOWLEDGMENTS

This work was supported by grants for fundamental scientific research 02556008 and 03304016 from the Ministry of Education, Science and Culture of Japan.

REFERENCES

1. **Ajioka, J. W., and W. F. Eanes.** 1989. The accumulation of P-elements on the tip of the *X* chromosome in populations of *Drosophila melanogaster*. Genet. Res. Camb. **53:**1–6.
2. **Aksoy, S., S. Williams, S. Chang, and F. F. Richards.** 1990. SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. Nucleic Acids Res. **18:**785–792.
3. **Biessmann, H., L. E. Champion, M. O'Hair, K. Ikenaga, B. Kasravi, and J. M. Mason.** 1992. Frequent transpositions of *Drosophila melanogaster* HeT-A transposable elements to receding chromosome ends. EMBO J. **11:**4459–4469.
4. **Biessmann, H., B. Kasravi, K. Jakes, T. Bui, K. Ikenaga, and J. M. Mason.** 1993. The genomic organization of HeT-A retroposons in *Drosophila melanogaster*. Chromosoma **102:**297–305.
5. **Biessmann, H., J. M. Mason, K. Ferry, M. d'Hulst, K. Valgeirsdottir, K. L. Traverse, and M.-L. Pardue.** 1990. Addition of telomere-associated HeT DNA sequences "heals" broken chromosome ends in *Drosophila*. Cell **61:** 663–673.
6. **Biessmann, H., K. Valgeirsdottir, A. Lofsky, C. Chin, B. Ginther, R. W. Levis, and M.-L. Pardue.** 1992. HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster*. Mol. Cell. Biol. **12:**3910–3918.
7. **Blackburn, E. H., and J. W. Szostak.** 1984. The molecular structure of centromeres and telomeres. Annu. Rev. Biochem. **53:**163–194.
8. **Brierley, I., P. Digard, and S. C. Inglis.** 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell **57:**537–547.
9. **Burke, W. D., C. C. Calalang, and T. H. Eickbush.** 1987. The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. Mol. Cell. Biol. **7:**2221–2230.
10. **Clare, J. J., M. Belcourt, and P. J. Farabaugh.** 1988. Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast *Ty1* transposon. Proc. Natl. Acad. Sci. USA **85:**6816–6820.
11. **Danilevskaya, O. N., D. A. Petrov, M. N. Pavlova, A. Koga, E. V. Kurenova, and D. L. Hartl.** 1992. A repetitive DNA element associated with telomeric sequences in *Drosophila melanogaster*, contains open reading frames. Chromosoma **102:**32–40.
12. **de Lange, T.** 1992. Human telomeres are attached to the nuclear matrix. EMBO J. **11:**717–724.
13. **Eickbush, T. H., and B. Robins.** 1985. *Bombyx mori* 28S genes contain insertion elements similar to the type I and type II elements of *Drosophila melanogaster*. EMBO J. **4:**2281–2285.
14. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
15. **Fujiwara, H., T. Ogura, N. Takada, N. Miyajima, H. Ishikawa, and H. Maekawa.** 1984. Introns and their flanking sequences of *Bombyx mori* rDNA. Nucleic Acids Res. **12:**6861–6869.
16. **Gabriel, A., and J. D. Boeke.** 1993. Retrotransposon reverse transcription, p. 275–328. *In* A. M. Skalka and S. P. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
17. **Gabriel, A., T. J. Yen, D. C. Schwartz, C. L. Smith, J. D. Boeke, B. Sollner-Webb, and D. W. Cleveland.** 1990. A rapidly rearranging retrotransposon within the mini-exon gene locus of *Crithidia fasciculata*. Mol. Cell. Biol. **10:**615–624.

17a. **Gottschling, D. E.** 1992. Telomere-proximal DNA in *Saccharomyces cerevisiae* is refractory to methyltransferase activity *in vivo*. Proc. Natl. Acad. Sci. USA **89:**4062–4065.

18. **Higgins, D. G., A. J. Bleasby, and R. Fuchs.** 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8:**189–191.
19. **Hutchison, C. A., III, S. C. Hardies, D. D. Loeb, W. R. Shehee, and M. H. Edgell.** 1989. LINEs and related retroposons: long interspersed repeated sequences in the eukaryotic genome, p. 594–617. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
20. **Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus.** 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell **55:** 447–458.
21. **Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus.** 1988. Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. Nature (London) **331:**280–283.
22. **Jacks, T., K. Townsley, H. E. Varmus, and J. Majors.** 1987. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins. Proc. Natl. Acad. Sci. USA **84:** 4298–4302.
23. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science **230:**1237–1242.
24. **Jakubczak, J. L., W. D. Burke, and T. H. Eickbush.** 1991. Retrotransposable elements *R1* and *R2* interrupt the rRNA genes of most insects. Proc. Natl. Acad. Sci. USA **88:**3295–3299.
25. **Jakubczak, J. L., Y. Xiong, and T. H. Eickbush.** 1990. Type I (R1) and type II (R2) ribosomal DNA insertions of *Drosophila melanogaster* are retrotransposable elements closely related to those of *Bombyx mori*. J. Mol. Biol. **212:**37–52.
26. **Karpen, G. H., and A. C. Spradling.** 1992. Analysis of subtelomeric heterochromatin in the Drosophila minichromosome *Dp1187* by single *P* element insertional mutagenesis. Genetics **132:**737–753.
27. **Levis, R. W.** 1993. *Drosophila melanogaster* does not share the telomeric repeat sequence of another invertebrate, *Ascaris lumbricoides*. Mol. Gen. Genet. **236:**440–442.
28. **Levis, R. W., R. Ganesan, K. Houtchens, L. A. Tolar, and F.-M. Sheen.** 1993. Transposons in place of telomeric repeats at a *Drosophila* telomere. Cell **75:**1083–1093.
29. **Louis, E. J., and J. E. Haber.** 1990. The subtelomeric *Y′* repeat family in *Saccharomyces cerevisiae*: an experimental system for repeated sequence evolution. Genetics **124:**533–545.
30. **Louis, E. J., and J. E. Haber.** 1990. Mitotic recombination among subtelomeric *Y′* repeats in *Saccharomyces cerevisiae*. Genetics **124:**547–559.
31. **Louis, E. J., and J. E. Haber.** 1992. The structure and evolution of subtelomeric *Y′* repeats in *Saccharomyces cerevisiae*. Genetics **131:**559–574.
32. **Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush.** 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.
33. **Lundblad, V., and E. H. Blackburn.** 1993. An alternative pathway for yeast telomere maintenance rescues *est1*$^-$ senescence. Cell **73:**347–360.
34. **Lundblad, V., and J. W. Szostak.** 1989. A mutant with a defect in telomere elongation leads to senescence in yeast. Cell **57:**633–643.
35. **Makarov, V. L., S. Lejinine, J. Bedoyan, and J. P. Langmore.** 1993. Nucleosomal organization of telomere-specific chromatin in rat. Cell **73:**775–787.
36. **Marlor, R. L., S. M. Parkhurst, and V. G. Corces.** 1986. The *Drosophila melanogaster gypsy* transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell. Biol. **6:**1129–1134.
37. **McClure, M. A.** 1993. Evolutionary history of reverse transcriptase, p. 425–444. *In* A. M. Skalka and S. P. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
38. **Mellor, J., S. M. Fulton, M. J. Dobson, W. Wilson, S. M. Kingsman, and A. J. Kingsman.** 1985. A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon *Ty1*. Nature (London) **313:**243–246.
39. **Okazaki, S., K. Tsuchida, H. Maekawa, H. Ishikawa, and H. Fujiwara.** 1993. Identification of a pentanucleotide telomeric sequence, $(TTAGG)_n$, in the silkworm *Bombyx mori* and in other insects. Mol. Cell. Biol. **13:**1424–1432.
40. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. **8:**3867–3874.
41. **Saigo, K., W. Kugimiya, Y. Matsuo, S. Inouye, K. Yoshioka, and S. Yuki.** 1984. Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. Nature (London) **312:**659–661.
42. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method

for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.

43. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.

43a. **Sheen, F.-M., and R. W. Levis.** 1994. Transposition of the LINE-like retrotransposon TART to *Drosophila* chromosome termini. Proc. Natl. Acad. Sci. USA **91:**12510–12514.

44. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.

45. **Tommerup, H., A. Dousmanis, and T. de Lange.** 1994. Unusual chromatin in human telomeres. Mol. Cell. Biol. **14:**5777–5785.

46. **Traverse, K. L., and M. L. Pardue.** 1988. A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres. Proc. Natl. Acad. Sci. USA **85:**8116–8120.

47. **Villanueva, M. S., S. P. Williams, C. B. Beard, F. F. Richards, and S. Aksoy.** 1991. A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. Mol. Cell. Biol. **11:**6139–6148.

48. **Voytas, D. F., and J. D. Boeke.** 1993. Yeast retrotransposons and tRNAs. Trends Genet. **9:**421–427.

48a. **Wright, J. H., D. E. Gottschling, and V. A. Zakian.** 1992. *Saccharomyces* telomeres assume a non-nucleosomal chromatin structure. Genes Dev. **6:**197–210.

49. **Xiong, Y., and T. H. Eickbush.** 1988. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol. Cell. Biol. **8:**114–123.

50. **Xiong, Y., and T. H. Eickbush.** 1988. Functional expression of a sequence-specific endonuclease encoded by retrotransposon R2Bm. Cell **55:**235–246.

51. **Xiong, Y., and T. H. Eickbush.** 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements and mitochondrial introns. Mol. Biol. Evol. **5:**675–690.

52. **Xiong, Y., and T. H. Eickbush.** 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:**3353–3362.

53. **Young, B. S., A. Pession, K. L. Traverse, C. French, and M.-L. Pardue.** 1983. Telomere regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. Cell **34:**85–94.

54. **Zakian, V. A.** 1989. Structure and function of telomeres. Annu. Rev. Genet. **23:**579–604.