Coevolution of the Telomeric Retrotransposons Across Drosophila Species

Elena Casacuberta and Mary-Lou Pardue

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Manuscript received January 17, 2002
Accepted for publication March 13, 2002

ABSTRACT

As in other eukaryotes, telomeres in Drosophila melanogaster are composed of long arrays of repeated DNA sequences. Remarkably, in D. melanogaster these repeats are produced, not by telomerase, but by successive transpositions of two telomere-specific retrotransposons, HeT-A and TART. These are the only transposable elements known to be completely dedicated to a role in chromosomes, a finding that provides an opportunity for investigating questions about the evolution of telomeres, telomerase, and the transposable elements themselves. Recent studies of D. yakuba revealed the presence of HeT-A elements with precisely the same unusual characteristics as HeT-Aw although they had only 55% nucleotide sequence identity. We now report that the second element, TART, is also a telomere component in D. yakuba; thus, these two elements have been evolving together since before the separation of the melanogaster and yakuba species complexes. Like HeT-Aw, TARTwm is undergoing concerted sequence evolution, yet they retain the unusual features TARTwm shares with HeT-Aw. There are at least two subfamilies of TARTwm with significantly different sequence and expression. Surprisingly, one subfamily of TARTwm has ≥95% sequence identity with a subfamily of TARTwm and shows similar transcription patterns. As in D. melanogaster, other retrotransposons are excluded from the D. yakuba terminal arrays studied to date.

Studies of insect telomeres have provided the first exceptions to the general mechanism of telomere formation (reviewed in Pardue and DeBaryshe 1999). In most animals, plants, and single-celled eukaryotes an enzyme, telomerase, produces long arrays of simple DNA sequences on the ends of chromosomes. Some insects share this mechanism. For example, in the silkworm, Bombyx mori, telomeres consist of long arrays of the sequence, TTAGG, only one nucleotide different from the human repeat, TTAGGG (Okazaki et al. 1993; Sahara et al. 1999). The first exception to the general telomere mechanism was found in Drosophila melanogaster, whose telomeres are polarized head-to-tail arrays of DNA repeats generated by successive transpositions of the two telomere-specific non-LTR retrotransposons, HeT-A and TART (see Figure 1). A second exception to the general telomerase mechanism has been described in three species of Chironomus (Rosen and Edström 2000). Chironomus telomeres are also composed of head-to-tail arrays of repeats but the repeats are an order of magnitude longer than those known to be made by telomerase. These arrays undergo rapid concerted evolution, apparently by gene conversion (Kamnert et al. 1998), but it is not clear whether this process also results in the net DNA synthesis needed to counteract telomere recession or whether the repeats are generated by another mechanism.

Both telomerase and the Drosophila retrotransposons extend telomeres by copying RNA sequences, and it is possible that Chironomus telomere extension is also a variation of this basic theme. Understanding how these different telomere types are related would help us to understand how telomeres evolved. In addition, comparison of variant telomeres in different species could identify features that are universally important for telomere structure and function. The information needed will come only from the study of several branches of the phylogenetic tree, with the distance between the branches carefully chosen to maximize our understanding.

Identifying alternative Drosophila telomere types is not easy, even in species closely related to D. melanogaster, because of the rapid change of telomere sequences within each species. This difficulty was seen in using cross-hybridization to find and characterize HeT-A in D. yakuba (Danilevskaya et al. 1998a). It has been estimated that the separation between D. yakuba and D. melanogaster occurred 5–15 million years ago (Lachaise et al. 1988). Studies of HeT-A elements from these two species showed that HeT-A forms a multicopy family in each. These multicopy families have maintained sequence similarity within each species while diverging significantly from the elements in the other species, a clear example of concerted evolution. HeT-Aw and HeT-Awm have only 55% nucleotide sequence identity but the conserved features show that these are homologous elements.

The rapid sequence divergence for HeT-A was not
unexpected. The element encodes a Gag protein, and retroelement Gag protein sequences diverge more rapidly than the Pol sequences (McClure et al. 1988). The remainder of the HeT-A sequence is noncoding DNA [mostly in the 3′ untranslated region (3′ UTR)]. Noncoding DNA is also expected to evolve rapidly because it is not under the constraints that affect coding sequences. Nonetheless, the HeT-A noncoding sequences show a conserved pattern—a regular spacing of A-rich regions. Because the HeT-A 3′ UTR is abundant in heterochromatic regions, both at telomeres and in the Y chromosome, we have suggested that these regions might be involved in protein binding to form heterochromatin (Danilevskaya et al. 1998b).

Sequence differences between HeT-A\textsuperscript{ed} and HeT-A\textsuperscript{ak} are distributed fairly evenly over the element with 65% nucleotide identity in the coding region and 50% identity in the 3′ UTR. Despite these extensive sequence changes throughout the element, the features that characterize HeT-A\textsuperscript{ed} are conserved in HeT-A\textsuperscript{ak} (Danilevskaya et al. 1998a). (For example, both elements transpose only to telomeres where they form long head-to-tail arrays; both have long 3′ UTRs; and neither codes for its own reverse transcriptase.)

The conservation of these distinctive features argues that the features are important for HeT-A to function as telomeres. Thus, the rapid sequence divergence of these HeT-A elements suggests that the 5–15 million years separating D. melanogaster from D. yakuba may give meaningful information about the evolution of other aspects of telomere biology. If so, this will be very useful because much of the telomerase-mediated telomere biology has been stable over long evolutionary periods.

One of the unusual characteristics of the D. melanogaster telomere is its complexity when compared to the homogeneous simple repeats generated by telomerase. Drosophila telomeres contain mixed arrays of the two non-LTR retrotransposons. Despite their invariant association in D. melanogaster telomeres, HeT-A and TART are now thought to belong to different lineages of non-LTR retrotransposons and to have acquired their telomeric roles as the result of convergent evolution (Danilevskaya et al. 1999). The two elements are present in mixed arrays in every D. melanogaster stock that has been studied. This observation leads one to ask if both elements are present in telomeres because they cooperate, because they compete, or simply by chance. The interspersion of HeT-A and TART in telomere arrays has not allowed us to eliminate either element by genetic crosses in D. melanogaster so these questions cannot be answered directly.

Non-LTR retrotransposons are reverse transcribed onto the chromosome, primed by a 3′ hydroxyl of the target DNA (Luan et al. 1993). This mechanism for integration of retroelements is called target-primed reverse transcription. In principle, the 3′ hydroxyl primer could be either on the end of a chromosome or exposed by a nick within a DNA molecule. In practice, it appears that HeT-A and TART prime their first-strand synthesis only off the chromosome end and extend the chromosome by a mechanism very similar to that used by telomerase. In contrast, all other known non-LTR retrotransposons add only to nicks within the chromosome, although the priming mechanism appears similar. It is of interest to know whether these nonoverlapping transposition patterns are conserved in other species. Do HeT-A and TART transpose specifically to chromosome ends in other Drosophila species? Have other non-LTR retrotransposons moved into telomeres of other species?

One way to approach these questions is to study the phylogenetic distribution and conserved features of the telomeric elements. Therefore, we have extended our study of D. yakuba telomeres to examine sequences that associate with HeT-A\textsuperscript{ak} at chromosome ends. We find that TART is a telomeric element in D. yakuba and, other than HeT-A, is the only element found in its telomere arrays. As with HeT-A\textsuperscript{ak}, at least one subfamily of TART\textsuperscript{ak} displays significant sequence difference from the D. melanogaster lineage; nevertheless, the entire family conserves the unusual structural and localization characteristics of TART in D. melanogaster.

MATERIALS AND METHODS

Fly stocks: We analyzed four stocks of D. yakuba: Y-1, a stock of unknown provenance used in our earlier studies (Danilevskaya et al. 1998a); U-S180, from the Ivory Coast, and U-S181, from Kenya (both obtained from the Umeå Stock Center, Sweden); and S15, from Cameroon (obtained from M. Ashburner). All four stocks have both subfamilies of TART\textsuperscript{ak} elements. The D. melanogaster stock was Oregon R.

Southern blot hybridization: For each sample, 20 μg of genomic DNA was digested with restriction enzymes, fractionated in a 0.7% agarose gel, and transferred to Hybond-N membrane (Amersham Pharmacia Biotech). Moderate-stringency hybridization was overnight at 60° in 4× SET (1× SET: 0.15 M NaCl, 0.03 M Tris pH 7.4, 2 mM EDTA), 5× Denhardt’s solution, 0.5% SDS, and 50 μg/ml salmon sperm DNA. Washes were at 60°, 2× 20 min in 2× SSC, 0.5% SDS, and 2× 20 min in 1×
SSC, 0.5% SDS. Low-stringency hybridization was overnight at 55°C in the same hybridization solution followed by 4 × 20-min washes at 55°C with 2× SSC, 0.5% SDS. After the initial exposure of low-stringency hybridization, the filters were washed 2 × 20 min at 55°C with 1× SSC and reexposed. The filters were then washed 2 × 20 min at 65°C with 0.5× SSC and exposed for a final time. DNA probes were labeled with [32P]dATP by random primer labeling (Feinberg and Vogelstein 1983).

Library screening: A D. yakuba genomic library in Lambda Fix II (Stratagene, La Jolla, CA) was obtained from Michael Griswold (North Carolina State University). The library was screened as described in the instruction manual for Lambda Fix II, using the moderate-stringency conditions described above.

Cloning and sequencing of library clones: Inserts were mapped at high resolution by restriction digestion and hybridization with D. yakuba cloned HeT-A and TART sequences. All TART element sequences were sequenced completely and all junctions between elements were sequenced through. HeT-A elements were sequenced from both ends and the sizes of the regions between those ends were determined to ensure that only a single element was present at each site; however, not all of the central regions of HeT-A elements have been sequenced. Sequences are deposited in GenBank as AF468023–AF468026.

Northern hybridization: RNA extraction was as described by Danilevskaya et al. (1999). A total of 20 μg of total RNA per lane was treated with glyoxal, separated on a 0.7% agarose gel overnight at 35 V, and transferred to Hybond-N membrane. Hybridization was overnight at 65°C in 4× SET, 5× Denhardt’s solution, 0.5% SDS, and 50 μg/ml salmon sperm DNA. Filters were washed three times at 65°C with 1× SSC and 0.5% SDS and then treated at 37°C for 1 hr with 100 units/ml RNAseT1 (Boehringer Mannheim, Indianapolis) in buffer (10 mM Tris-HCl pH 7.5, 5 mM EDTA, 300 mM NaCl). After a rinse with 1× SSC, 0.5% SDS, filters were exposed for autoradiography. 32P-labeled RNA probes were transcribed in vitro from DNA fragments inserted into pBluescript II SK (Stratagene), according to the manufacturer’s protocol.

Probes: R1 is nucleotide (nt) 1665–3804 of GenBank no. AF468023. R2 is nt 4173–5991 of GenBank no. AF468026. FIBE is nt 1–1804 of GenBank no. AF468023.

Sequence analyses: Sequences were analyzed by Blast searches of FlyBase and GenBank. Identity percentages were calculated using the LAlign program available at the Genestream Network (IGH, Montpellier, France). Phylogenetic analyses were performed with CLUSTALW (Thompson et al. 1994) and the Mega software version 2.1 (Kumar et al. 2001). The trees were constructed on the basis of the number of differences determined by CLUSTALW alignment of the sequences, using both the neighbor-joining and the UPGMA algorithms. Bootstrap tests were performed with 500 replications and a cutoff value of 50% for the consensus tree. DotPlot (Maizel and Lenk 1981) analyses were performed using a window of 25 and a stringency of 15.

In situ hybridization: D. yakuba polytene chromosomes were hybridized as in Pardue (2000) except that probe [5 ng DNA in 10 μl hybridization buffer (50% formamide, 3× SSC, 10% dextran sulfate)] was added to each preparation before chromosomes and probe were denatured by heating slides at 95°C for 2 min. Hybridization was overnight at 37°C. DNA probes were labeled with digoxigenin-dUTP using the High-Prime kit (Roche Diagnostics, Indianapolis). Probes were detected by the enzymatic activity of antidigoxigenin-alkaline phosphatase conjugate.

PCR: Primer sequences: The Y chromosome-specific FIBE-reverse transcriptase junction was amplified with the primers F1BP (5’ GGAACCTAAGAAGCGCCGT 3’) and YRT1 (5’ GTG TCGCCTTGTAGTAGG 3’). These sequences were taken from GenBank sequence AF468023. The positive control sequence found in both sexes was amplified with primers POL2D (5’ GTACCCGAAACTAGACTAGA 3’) and POL2R (5’ GACCC TTGATCTTTGACTACA 3’). These sequences were taken from GenBank sequence AF468026. Amplification was with Taq polymerase and reaction buffer (Roche Diagnostics) for 35 cycles. The program for F1BP and YRT1 primers was 95°C for 30 sec, 58°C for 30 sec, and 72°C for 90 sec. The program for primers POL2D and POL2R was 95°C for 30 sec, 50°C for 30 sec, and 72°C for 120 sec. For both primers the final extension was increased to 7 min.

RESULTS

D. yakuba has non-LTR retrotransposons related to TARTtm: Reverse transcriptase coding sequences are the least rapidly evolving sequences in retroelements (McClure et al. 1988). Therefore, we initiated our search for TART elements in D. yakuba by probing Southern blots of D. yakuba DNA with a fragment of reverse transcriptase coding sequence from TARTtm (nt 434–2683 of GenBank accession no. U02279). When hybridization was carried out at moderate stringency, the D. melanogaster probe bound to multiple restriction fragments of D. yakuba DNA. Two of these fragments were cloned and sequenced. The sequences of the two clones were distinctly different but both were very similar to that of TARTtm with Blast scores between 3e-54 and 1e-29, depending on the TARTtm subfamily sequence used as query. No other transposable element was retrieved by the Blast search.

The cloned D. yakuba TART sequences were used to probe a library of λ phage carrying D. yakuba DNA. Eight clones were recovered. Duplicate clones were identified by restriction mapping and discarded, leaving three cloned sequences. Each of these clones was composed of a mixed array of TARTtm and HeT-A elements (Figure 2). Two of these three clones have HeT-A and TART elements in arrays like those found in D. melanogaster telomeres. The third clone closely resembles a class of nontelomeric repeats found in the heterochromatic D. melanogaster Y chromosome that was originally identified because its members contained fragments of HeT-A (Danilevskaya et al. 1993). Experiments described below show that this D. yakuba clone belongs to this class.

The D. yakuba TART sequences are not found in euchromatic regions: HeT-A and TARTtm are remarkable because they never transpose into euchromatin regions where most non-LTR retrotransposons can be found. An earlier study of HeT-A (Danilevskaya et al. 1998a) showed the same limits to transposition seen in D. melanogaster; HeT-A was never found in euchromatin. All of the TART clones analyzed here also contained HeT-A elements and therefore could not have come from euchromatin. However, this does not eliminate the possibility that D. yakuba has some TART elements that transpose into euchromatin. This possibility can be investigated at high resolution by in situ hybridization to
polytenic chromosomes because polytenization provides amplification of euchromatic sequences, making it easy to detect single-copy genes. We studied the distribution of TART\textsuperscript{vak} in D. yakuba, using probes for the reverse transcriptase. These sequences show no hybridization in the banded chromosome arms. We conclude that there are no TART elements in euchromatic regions in D. yakuba.

The probes for TART\textsuperscript{vak} reverse transcriptase bind to the most terminal band on chromosome arms, as expected for a telomere sequence (Figure 3). Both probes also bind to regions of the heterochromatic chromocenter, as do HeT-A\textsuperscript{vak} probes. This contrasts with the situation in D. melanogaster; neither HeT-A\textsuperscript{vak} nor TART\textsuperscript{vak} hybridizes with the chromocenter in D. melanogaster. The chromocenter is fused and partially underreplicated centromeric heterochromatin. Although HeT-A and TART-related sequences are in pericentric regions (Traverse and Pardue 1989; Danilevskaya \textit{et al}. 1998a; Agudo \textit{et al}. 1999; Sirriaco \textit{et al}. 2002), under stringent hybridization conditions these are not detected in polytenic chromosomes. The chromocenter must also contain the tiny short arms of chromosomes X and 4 although it is not known whether these are polytenized in either or both species. The structure of the chromocenter is amorphous, with no landmarks to distinguish the regions where HeT-A and TART bind. Thus, the chromocentric hybridization in D. yakuba may be due to telomeres on the short arms of chromosomes X and 4. However, there is also a class of sequences that we designate HeT-TART-related mosaic sequences. These sequences, found in nontelomeric heterochromatin, contain fragments of HeT-A and TART mixed with other sequences. We discuss below a clone of these mosaic sequences from the D. yakuba Y chromosome (see last section of \textit{Results}). Y chromosomes are not polytenized and therefore Y chromosome sequences would not be banded chromosome arms. We conclude that there are no TART elements in euchromatic regions in D. yakuba.

\textit{D. yakuba telomeres contain mixed head-to-tail arrays of HeT-A and TART}: The two larger cloned sequences consist of mixtures of HeT-A\textsuperscript{vak} and TART\textsuperscript{vak} elements with no other sequences interspersed (see Figure 2A). As in D. melanogaster telomeres, all elements are oriented in the same direction. Such polar arrays are presumably generated by successive target-primed transpositions onto the end of the chromosome. The elements in these clones are truncated by varying amounts at their 5' ends; similar 5' truncations are seen in D. melanogaster telomere arrays and are common for non-LTR elements generally. This truncation is thought to be due to failure to complete reverse transcription. For HeT-A and TART, truncation could also result from end erosion of the chromosome before the addition of the next retrotransposon. Neither of these two clones contains any additional sequence interrupting the HeT-A and TART arrays. As in the telomere arrays of D. melanogaster, HeT-A\textsuperscript{vak} and TART\textsuperscript{vak} elements associate only with each other.

\textit{TART\textsuperscript{vak} elements form subfamilies that differ in se-
sequence and in degree of similarity to TART\textsuperscript{mel} subfamilies: The cloned sequences contain four TART\textsuperscript{mel} elements (Figure 2). The three elements in telomere arrays are partial, truncated at the 5' end by attachment of another element or at the 3' end by the cloning vector. All junctions of the telomeric TART\textsuperscript{ak} elements are with He\textsuperscript{T}-A\textsuperscript{st}. The Y chromosome TART\textsuperscript{ak}1 (see last section of results) is truncated on both ends.

The TART\textsuperscript{ak} elements can easily be divided into two subfamilies, 1 and 2, on the basis of the sequence of the 3' UTRs. These sequences are so different that it is not possible to do a meaningful alignment of the 3' UTRs to compare the subfamilies. Precisely the same situation is seen with the TART elements in D. melanogaster, where three subfamilies, A, B, and C, have been identified on the basis of significant differences in the sequence of their 3' UTRs. Although these differences are too great to allow alignment of 3' UTR sequences of the TART\textsuperscript{ak} subfamilies, dot matrix comparisons reveal some sequence similarity in the 3'-most kilobase of the element in all of the D. melanogaster subfamilies (data not shown). No region of similarity is seen in dot matrix comparisons of the TART\textsuperscript{ak} subfamilies.

Because of so little evidence of sequence conservation of the 3' UTR among elements within either species, it was a surprise to find that the 3' UTR of TART\textsuperscript{mel} subfamily 2 (TART\textsuperscript{mel2}) is highly similar to that of the C subfamily of TART\textsuperscript{ak}. The two 3' UTRs have 95% nucleotide identity over 2040 bp (Figure 4). The comparison is limited on the 5' end because only the available TART\textsuperscript{melC} sequence is truncated. On the 3' end the TART\textsuperscript{mel} sequence extends for 1381 bp beyond the end of TART\textsuperscript{mel2}. It does not appear that the TART\textsuperscript{mel} is truncated because it has the typical 3' oligo(A) stretch. The lack of a TART\textsuperscript{mel} counterpart of the 3'-most terminal region of TART\textsuperscript{ak} is puzzling because the terminal region is the one showing some similarity between the TART\textsuperscript{ak} subfamilies.

As expected from the D. melanogaster data, the TART\textsuperscript{ak} subfamilies have much more sequence similarity in the coding regions than in the 3' UTRs. Nevertheless, analyses of both DNA and protein sequences show the same pattern of subfamily divergence that we see in the 3' UTR sequence (Figure 5). To analyze the relationships of the TART\textsuperscript{ak} subfamilies, coding sequences from all available TART\textsuperscript{ak} and TART\textsuperscript{mel} subfamilies were compared. TART\textsuperscript{melC} was not included because the only available sequence is a portion of its 3' UTR. Two mechanisms that can facilitate concerted evolution are recombination and gene conversion. These mechanisms might be expected in telomeres because the TART\textsuperscript{ak} subfamilies are mixed in these arrays, and both they and their neighboring He\textsuperscript{T}-A elements have sufficient sequence similarity to encourage both recombination and gene conversion events involving TART elements. We looked for interchange between the TART\textsuperscript{ak} subfamilies by analyzing three regions of the coding sequence independently. We divided the open reading frame (ORF) 2 region into the endonuclease domain and the reverse transcriptase domain. We used only the 3' end of ORF 1
because all of the cloned elements were truncated at the 5’ end.

The coding regions were compared as both nucleotide and amino acid sequences using neighbor-joining and UPGMA algorithms. Only the neighbor-joining trees of nucleotide sequence analyses are shown (Figure 5), but all analyses yield the same conclusions: All three regions of TART\textsuperscript{yak2} elements are more similar to the corresponding TART\textsuperscript{mel} regions than to those of TART\textsuperscript{yak1} even though the three regions of the ORFs diverge at different rates, as is found for other retroelements (McClure et al. 1988). Therefore, at least in our limited sample, the TART\textsuperscript{yak} subfamilies appear to be maintained as intact units without sequence exchanges.

Although there is evidence that recombination and/or gene conversion can occur during the repair of short, recently healed D. melanogaster telomeres (Kahn et al. 2000), our observation that TART subfamilies appear to be evolving as units suggests that in the normal course of events, replacement of elements by new transposition, rather than by recombination or gene conversion between elements, may be the predominant force in sequence change. This suggestion is similar to the one made by Perez-Gonzalez and Eickbush (2001) to explain the evolution of multiple lineages of R1 and R2 elements within the rDNA locus.

The relative abundance of the two D. yakuba subfamilies can be evaluated by Southern blot analysis. When the hybridization is performed at medium stringency, bands corresponding to both subfamilies cross-hybridize (Figure 6) but the relative strength of the signal depends on the subfamily member used as a probe (compare bands marked with an asterisk in Figure 6). Higher-stringency washes eliminate most of the cross-hybridization (not shown). These analyses show that the TART\textsuperscript{mel} subfamily is more abundant in D. yakuba than the TART\textsuperscript{yak} family.

The two subfamilies of TART\textsuperscript{mel} differ somewhat in expression and possibly in translation: The D. yakuba TART elements produce both sense and antisense transcripts, as do the TART elements in D. melanogaster.

**Figure 5.**—Phylogenetic relationships among TART and HeT-A coding regions in D. melanogaster and D. yakuba. Nucleotide and amino acid sequences of all available elements were aligned using CLUSTALW and then analyzed in MEGA2.1 software. Neighbor-joining trees for the nucleotide sequences are shown. (The UPGMA trees yield the same relationships, as do the amino acid trees.) Bootstrap tests were performed with 500 replications and a cutoff value of 50% for the consensus tree. Numbers indicate bootstrap values ≥ 50% in the corresponding node. When comparing coding regions that were not complete, the smallest one was used to trim the others. Additional analyses were performed without the shortest sequence to see whether it was biasing results; however, relationships did not change. Bars indicate the number of changes between the sequences. The only sequence from a Y chromosome mosaic repeat is the RT from TART\textsuperscript{mel}. Note that the only available TART Gag sequences were from the more conserved 3’ part of the coding region. The HeT-A Gag sequences, added for comparison, are all from complete coding regions.
TART elements yield many more antisense transcripts than sense-strand transcripts (Danilevskaya et al. 1999). Probes for the TART subfamily, the subfamily most like TART in sequence, also detect a large excess of antisense RNA (Figure 7). The sense transcripts of this family are much less abundant and blots show much background due to the long exposure needed (data not shown). In contrast, we find approximately equal amounts of sense and antisense RNA from TART. Although our clones do not have any complete TART elements, the sizes of the major bands in the Northern blots are comparable to the sizes of TART elements. We presume that these large transcripts come from full-length elements not present in our clones.

Non-LTR retrotransposons should require only sense-strand RNA because this strand serves as both mRNA and the transposition template. As expected, Het-A yields only sense transcripts. In contrast, TART and a few elements in other organisms make both sense and antisense transcripts. The function of the TART antisense RNA is unknown, but TART shows structural similarity (Danilevskaya et al. 1999) to the Dictyostelium element, DRE, which requires both strands of RNA for replication (Schumann et al. 1994). Although the ratios of the two strands differ in the D. yakuba subfamilies, the ability to yield both strands has been conserved.

A second characteristic of TART sequence that is conserved in TART, but not TART, is the sequence joining the gag coding region (ORF 1) to the pol coding region (ORF 2). Retroelements tend to translate ORF 2 as part of a polyprotein linked to the product of ORF 1 either by a frameshift or by readthrough of a leaky stop codon (Jack 1990). Nevertheless at least one RNA virus, hepatitis C virus, has been shown to translate ORF 2 independently of ORF 1 (Brown et al. 1992), and other retroelements, e.g., human LINE-1 (McMillan and Singer 1993) and the L element of D. melanogaster (Bouidel et al. 1994), have been shown to be capable of internal initiation to translate ORF 2. TART sequences suggest that TART also uses internal initiation to translate ORF 2 because the arrangement of stop codons in the three frames between the ORFs would require complex ribosome movements to link the translation products (Figure 8). Both of the TART sequences reported here show the same arrangement of stop codons seen in TART, although there are differences in
Figure 8.—Alignment of the nucleotide sequences linking TART ORF 1 and ORF 2 in elements from both D. yakuba and D. melanogaster. (A) For each of the elements, the nucleotide sequence begins at the stop codon of ORF 1 (TAA in boldface type) and continues to the start codon of ORF 2 (ATG in boldface type). Gaps in the alignment are indicated by a dash (–). (B) Translation of the above sequences in the two relevant frames. Stop codons are represented by an asterisk (*). The first amino acid (M) in ORF 2 is indicated in boldface type for each element. In frame 1, all sequences are shown from the final stop codon of ORF 1 but only the element TART yak1 begins translation of ORF 2 in this same frame. (This element has two extra nucleotides that change the frame relative to the other sequences.) The other sequences all have at least one more stop in the nine codons beyond those shown here for frame 1. All elements except TART yak1 begin translation of ORF 2 in the second frame but have two stop codons between the stop in frame 1 and the ATG in frame 2.

The third D. yakuba clone differs markedly from the other two in sequence arrangement: It has a partial TART yak1 element in one orientation and a partial HeT-A yak element in the opposite orientation (Figure 2B). Furthermore, each of these elements is truncated near its 3′ end where it attaches to the other element. Neither tail-to-tail attachments nor 3′ truncations are found in telomeric regions. Tail-to-tail attachments should not be found in telomeres because telomere extension is by target-primed reverse transcription onto the chromosome end and priming from the chromosome end dictates uniform polarity. Truncation at the 3′ end is not expected because this transposition mechanism requires the extreme 3′ end sequences of the element (Luan et al. 1993).

In addition to the 3′ truncations, both elements in this third clone are truncated at the 5′ end by attachment of unrelated sequences. The 5′ end of the TART yak1 element is associated with a novel 1.8-kb sequence (FIBE) with no protein-coding regions or similarity to known transposable elements. The 5′ end of the HeT-A yak element is associated with ~5 kb of a second novel sequence (FIBT) that makes up the rest of the clone. We have sequenced the 1.8-kb FIBE sequence. It has no ORFs but has significant similarity to a scaffold sequence of unknown function in the euchromatin of chromosome 3 of D. melanogaster. Sequences of several fragments from both ends and the center of the 5-kb FIBT fragment adjacent to the HeT-A element all show high similarity to the same scaffold sequence located in the euchromatin of chromosome 2 of D. melanogaster. The region of the D. melanogaster scaffold with similarity to the FIBT sequence has no ORFs. The atypical features of this third clone suggested that it is derived from a class of nontelomeric sequence mosaics that we initially identified in D. melanogaster. In that study (Danilevskaya et al. 1993), these mosaics were found in several families.
of tandem repeats. In each family the unit repeat con-
tained scrambled fragments of HeT-A and other unre-
related sequences, some known and some unknown; we
therefore called them HeT-A-related repeats. We now
know that some families of these repeats also contain
fragments of TART; thus, we prefer the term HeT-TART-
related mosaic repeats. Repeats from two of the families
on the D. melanogaster Y chromosome have been se-
quenced (Danilevskaya et al. 1993). One, the 356 re-
peat, has two fragments of HeT-A 3’ UTR joined to a
fragment of TART 3’ UTR. The other, the 665 repeat,
contains sequence from the HeT-A 3’ UTR with frag-
ments of Stellate, a gene located in the euchromatic
polytene region 12E, and the transposable element
Copia. These repeats are present in the nontelomeric
heterochromatin of the Y chromosome, with possibly
the largest array being pericentric (Agudo et al. 1999).
In situ hybridization experiments also suggest that similar
repeats are present in the pericentric heterochromatin
of the autosomes but these sequences have not been
characterized (Traverse and Pardue 1989). The hy-
pothesis that the third D. yakuba clone derives from a
mosaic repeat on the Y chromosome leads to three
testable predictions. First, individual sequences in the
clone, although present at other sites, should be more
abundant in male DNA because they are repeated on the
Y chromosome. Second, the junctions between the
HeT-A or TART fragments and the nontelomeric com-
ponents of the repeat will be found only in male DNA
because this mosaic is only on the Y chromosome. Third,
mosaic sequences other than the HeT-A and TART frag-
ments will be found only in nontelomeric regions by in
situ hybridization to polytene chromosomes. All three
of these predictions are satisfied by the D. yakuba clone.

The abundance of the repeat sequence in DNA from
males and females was measured by hybridizing South-
ern blots with a probe for the FIBE sequence because this
sequence is less repeated in the genome than are
HeT-A and TART and thus easier to measure on the
Southern blot. The probe hybridized with DNA from
both sexes but there was a significant excess of hybridiz-
ing sequence in DNA from males (Figure 9A), as ex-
pected for a sequence repeated on the Y chromosome
but also present elsewhere in the genome.

If FIBE and TART are associated only in the mosaic
fragment on the Y chromosome, the junction between
these two sequences should be found only in DNA from
males. As predicted, analytical PCR experiments with
male DNA amplified a fragment spanning the FIBE-
TART junction, whereas a control fragment was am-
plified equally from DNA of both sexes (Figure 9B).
The junction PCR primers were chosen to amplify a
1.6-kb fragment extending from one primer in FIBE to
a second primer in the TART sequence (Figure 9, B
and C). Even in overloaded gel lanes, this 1.6-kb frag-
ment was not found in PCR products of female DNA.
As a control to test the ability of the female DNA to
serve as a template for PCR we used primers from within
a telomeric TART that should be present in DNA from
the two sexes.

In situ hybridization to polytene chromosomes can-
not be used to map sequences on the Y chromosome be-
cause the Y does not polytenize and is therefore unde-
tectable in these nuclei. However, in situ hybridization
can find other sites occupied by the fragments that make
up the mosaic. Southern hybridization had shown that
some of the FIBE sequence was present in females. On
polytene chromosomes, the FIBE probe hybridized to
only one site, a large band in a euchromatic region of
one chromosome. There was no hybridization to any
telomeric site (data not shown). Thus, FIBE is not associ-
ated with HeT-A or TART in any of the sites detected
in polytene chromosomes; the association is found only
on the Y chromosome. In situ hybridization to a second
D. yakuba stock (not shown) shows the same chro-
mosomal site, supporting our conclusion that it is not a
mobile element. As discussed above, there is precedent
for finding euchromatic sequence in HeT-TART-related
mosaics; fragments of the Stellate gene are found in the
D. melanogaster 665 repeat family.

DISCUSSION

HeT-A and TART occupy the same niche in the genomes
of D. yakuba and D. melanogaster: The D. melanogaster ge-

nome contains a number of families of non-LTR retro-
transposons, including HeT-A and TART. These elements
have several well-conserved features atypical of non-LTR
retrotransposons; these features are presumably related
to their exclusive association to telomeres and hetero-
chromatin. Target-primed reverse transcription, by which
non-LTR elements insert into new sites, explains the
ability of HeT-A and TART to add to chromosome ends
where they form arrays in which the 3’ ends are oriented
toward the centromere. Other non-LTR retrotranspo-
sons use the same mechanism to transpose into many sites
in euchromatic regions of the chromosomes but are
never found in telomeric arrays. HeT-A and TART
are the only elements found in telomere arrays and
they are never found in euchromatin.

The only D. yakuba telomeric element previously char-
acterized is HeT-A, which shows the same pattern of telo-
mere-specific transposition seen in D. melanogaster. The
TART clones characterized here show that HeT-A also
shares its telomeric sites with TART in D. yakuba. Neither
of these telomeric elements is found in euchromatin in
D. yakuba, nor do we detect new retrotransposons that
have acquired telomeric specificity.

Long runs of sequence that do not code for proteins
needed for transposition are rare in retrotransposons,
yet both HeT-A and TART elements have large regions
of noncoding DNA. This DNA is undergoing concerted
evolution but there are underlying patterns of sequence
conservation. The conservation suggests function, al-
though function other than protein coding is not easily deciphered from sequence. An illustrative example of marked change in a noncoding sequence of defined function is seen in the gene for the RNA template used by telomerase to extend telomeres. Telomerase RNAs from different organisms vary greatly in both size and sequence but some sequence conservation preserves the folding pattern of the transcript (Chen et al. 2000). Presumably, this folding pattern is important for interactions between the RNA and protein components of telomerase. It is possible that the conserved sequences in the 3' UTRs of the telomeric retrotransposons play a similar role in the transposon RNA, that they are involved in heterochromatization of the telomere, or that they are needed for protein binding after incorporation into the heterochromatic DNA of the telomere. In any case, these conserved sequences are likely to be important for interactions with other molecules.

**D. yakuba TART elements are undergoing concerted evolution in at least two subfamilies:** One somewhat unexpected finding of this study is that the two subfamilies of TART\textsuperscript{mel} are evolving separately. A surprising finding is the high similarity between the sequence of TART\textsuperscript{mel} and TART\textsuperscript{yak}.

Such high sequence conservation, especially in a non-coding region, raises the possibility of horizontal transmission. Transmission of a TART\textsuperscript{mel} element from *D. melanogaster* could explain the high similarity between TART\textsuperscript{mel} and TART\textsuperscript{yak}. It is known that transposable elements can transpose horizontally between species (Kidwell 1992). Transmission between species appears to be more frequent for DNA transposons (Clark and Kidwell 1997), but has also been described for RNA transposons (Kidwell 1992; Jordan et al. 1999). The presence of the non-LTR retroelement jockey in *D. melanogaster* and *D. funebris*, but not in the intervening species, has led to a proposal of horizontal transfer (Mizrokhi and Mazo 1990); however, the possibility that jockey has been lost in the intermediate species cannot be ruled out (Malik et al. 1999). Invasion of *D. melanogaster* by P\textsubscript{mel} elements has been demonstrated because this element is not present in stocks placed in laboratories before 1950 (Ashburner 1989). We have analyzed the four available *D. yakuba* stocks and found that all have significant numbers of TART\textsuperscript{yak} elements and that these elements are transcribed. If this element has invaded *D. yakuba*, the element has now spread through populations in the geographical regions sampled (see MATERIALS AND METHODS for origin of stocks).

Although we cannot rule out horizontal transmission
of TART\textsuperscript{mel} and TART\textsuperscript{yak} sequences, rather than TART\textsuperscript{mel} sequences, cross-hybridize with more distant species of Drosophila suggests that characterization of these species may reveal a pattern of evolution similar to that of R1.

**In conclusion:** Comparison of the mechanisms used by telomerase and the reverse transcriptases of non-LTR retrotransposons suggests that little, if any, modification would be necessary to enable retrotransposons to extend telomerases (Pardue et al. 1996). Evidence that the two telomeric transposons in D. melanogaster have different origins is consistent with this supposition. However, acquisition of the ability to extend telomerases must be a rare event because HeT-A and TART are the only known telomic retrotransposons. These studies of D. yakuba show that the telomere specificity of both these elements must have developed before the separation of the D. melanogaster and D. yakuba species complexes. Since that time, none of the other non-LTR elements now active in Drosophila genomes has joined the telomere, and despite much sequence change, the basic features of the Drosophila telomere have been conserved.

We thank Michael Griswold for the generous gift of the D. yakuba clone library and Michael Ashburner for one of the D. yakuba stocks. Members of the Pardue laboratory, Ky Lowenhaupt, and Josep Casacuberta have provided much useful discussion and helpful comments on the manuscript. This work has been supported by grant GM50315 from the National Institutes of Health.

**LITERATURE CITED**


Lowenhaupt, K., A. Rich and M. L. Pardue, 1989 Non-randdom distribution of long mono- and di-nucleotide repeats in Drosophila...


Communicating editor: J. A. Birchler