**Drosophila telomeres: A variation on the telomerase theme**

The MIT Faculty has made this article openly available. *Please share* how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://www.landesbioscience.com/journals/fly/article/6393">http://www.landesbioscience.com/journals/fly/article/6393</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Landes Bioscience</td>
</tr>
<tr>
<td>Version</td>
<td>Author’s final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Thu Jan 03 00:42:44 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/76259">http://hdl.handle.net/1721.1/76259</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>

**Please share** how this access benefits you. Your story matters.
DROSOPHILA TELOMERES: A VARIATION ON THE TELOMERASE THEME

Mary-Lou Pardue and P. Gregory DeBaryshe
Biology Department
Massachusetts Institute of Technology
Cambridge, MA 02139

Keywords
Anopheles gambiae, ATM, Bombyx mori, Cdc3, Chironomus, coevolution, Diptera, Drosophila americana, Drosophila melanogaster, Drosophila virilis, Endopterygota, endonuclease, EST1, HeT-A, HP1, Ku70, Ku80, mu2, MRE11, P-element, RAD50, retrotransposon, TAHRE, TART, TAS, Tegenaria ferrugenea, telomere, telomere associated repeats, telomere repeats, telomeric retrotransposons, Triboleum castaneum, TTAGG repeats, Uvir

Abstract
In Drosophila, the role of telomerase is carried out by three specialized retrotransposable elements, HeT-A, TART, and TAHRE. Telomeres contain long tandem head-to-tail arrays of these elements. Within each array, the three elements occur in random, but polarized, order. Some are truncated at the 5’ end, giving the telomere an enriched content of the large 3’ untranslated regions which distinguish these telomeric elements from other retrotransposons. Thus, Drosophila telomeres resemble other telomeres because they are long arrays of repeated sequences, albeit more irregular arrays than those produced by telomerase. The telomeric retrotransposons are reverse-transcribed directly onto the end of the chromosome, extending the end by successive transpositions. Their transposition uses exactly the same method by which telomerase extends chromosome ends - copying an RNA template. In addition to these similarities in structure and maintenance, Drosophila telomeres have strong functional similarities to other telomeres, and, as variants, provide an important model for understanding general principles of telomere function and evolution.

Introduction
There appear to be only a few ways to build a eukaryotic telomere.
The concept of the telomere was derived from analysis of early studies on Drosophila chromosomes. In 1938 Herman Muller noted that these chromosomes could survive many kinds of breakage, exchange, and rejoining, but simple terminal deletions were never found.¹ He concluded that chromosome ends are capped by special structures and that broken chromosomes could not survive if they did not acquire a cap from another chromosome. He named the caps telomeres and
noted that these regions had heterochromatic morphology.\textsuperscript{1,2} We now understand that this cap is what distinguishes a chromosome end from a break in the chromosome. Breaks in chromosomes activate a checkpoint response that prevents the cell from proceeding through the cell cycle. Telomere caps allow cells to pass the checkpoint but we do not understand the mechanism involved.

Also in 1938, Barbara McClintock\textsuperscript{3} reported that ends of broken chromosomes in corn tended to fuse with ends of other broken chromosomes, forming dicentric chromosomes. Dicentrics broke again when the two centromeres tried to enter different daughter nuclei at metaphase. These early studies suggested that telomeres in flies and corn were very similar, as might be expected if the basic structure of telomeres arose early in the evolution of linear nuclear chromosomes and had been conserved.

This evolutionary conservation is now strongly supported by studies of the molecular structure of the ends of eukaryotic chromosomes. Beginning with the work of Elizabeth Blackburn and colleagues on Tetrahymena,\textsuperscript{4} chromosomes in all animals, plants, and unicellular eukaryotes studied have been found to consist of long arrays of tandem repeat sequences\textsuperscript{5} (which we will call telomere repeats) and, frequently, repeated telomere-associated sequences (TAS).\textsuperscript{6}

The large amount of DNA found in the telomere arrays was unexpected because DNA viruses have much more economical telomere mechanisms and need only a few nucleotides to prevent erosion of the ends of their linear genomes (see ref. 7). In contrast, multicellular eukaryotes tend to have ten or more kilobases of these sequences per end, and even unicellular eukaryotes have a few hundred base pairs of telomere repeats on each chromosome. This large investment of cellular resources is less surprising now that we know that eukaryotic telomeres have many roles beyond simply capping the end of the DNA. For example, the arrays are important in cell maintenance, senescence, genomic stability, and oncogenesis in ways that are not understood but that are related to the length of the arrays.\textsuperscript{8-10} The significance of array length and the mechanisms by which it is regulated are major questions in the field today.

For the vast majority of eukaryotes, telomere arrays are composed of 5-10 bp repeats added to the chromosome end by the enzyme complex, telomerase. The repeat sequence for each organism is determined by the RNA template used by its telomerase. Organisms with telomerase are also able to extend telomere arrays by a recombination-based mechanism. Recombination provides a back-up mechanism when telomerase is lost but seems to be of minor importance in cells with active telomerase. After loss of telomerase in budding yeast populations undergoing senescence, rare spontaneous survivors use recombination to elongate telomeres.\textsuperscript{11,12} A significant fraction of human tumors and immortalized cell lines lack active telomerase and instead use recombination-based Alternative Lengthening of Telomeres (ALT) to extend telomeres.\textsuperscript{13,14}
Telomeres maintained by recombination in both yeast and human cells are very heterogeneous in length but return to normal when telomerase activity is restored experimentally, although vestiges of the recombination system remain.\textsuperscript{12, 15, 16}

In most organisms the telomerase catalytic subunit and its template RNA seem to be encoded by single copy genes. This is a boon for biological research because knocking out either gene eliminates telomerase activity. It is also puzzling that the telomerase mechanism has persisted with very little change throughout the evolution of eukaryotes despite being so easily susceptible to experimental knockout.

Telomerase activity has been found almost everywhere in the Eukaryota. Indeed, we think it likely that the primary mechanism for maintenance of all eukaryotic telomeres utilizes reverse transcription of RNA templates. Unfortunately, it is difficult to test this possibility because of problems in identifying telomere sequences in new organisms. Telomeres, like all complex repeat sequences, present nearly insurmountable technical problems both for cloning and for correct sequence assembly. As a result, sequence databases contain little, if any, of these sequences, even for organisms with complete assembly of the non-heterochromatic genome. Thus it is not possible to do informatic surveys for telomere sequences. However, the sequence of the telomerase RNA template has been so strongly conserved that in situ hybridization has identified telomere repeats on the chromosomes of many organisms. Surveys of the insects have been especially interesting.

Many insects have TTAGG telomerase repeats, one nucleotide different from the vertebrate TTAGGG. This TTAGG sequence hybridizes with telomeres of many different species but neither TTAGG nor TTAGGG hybridizes to chromosomes in species in several branches of the insect phylogenetic tree (Fig. 1), including branches of the most successful lineage of insects, the superorder Endopterygota. This lack of hybridization suggests that telomerase has been lost, or at least modified, several times in insect evolution.\textsuperscript{17, 18}

It is of considerable interest to learn what has happened to telomeres in organisms that do not appear to use telomerase, but only a few of these species have been studied. In one case the change has been relatively minor; the flour beetle, \textit{Triboleum castaneum}, uses a telomerase template, TCAGG, which does not cross-hybridize with TTAGG.\textsuperscript{19} It is not impossible that recombination has completely replaced telomerase in other organisms. For example, telomeres in the midge, \textit{Chironomus}, end in repeats of 176, 340 or 350 bp, depending on the species.\textsuperscript{20} There is clear evidence that these repeats maintain their homogeneity by recombination, although it is not clear that recombination also compensates for sequence erosion. Because extrachromosomal RNA-DNA complexes containing long runs of the telomere repeats have been found in all three studied species of \textit{Chironomus},\textsuperscript{21} \textit{Chironomus} may also use an RNA template to extend its telomeres. In the
mosquito, *Anopheles gambiae*, analysis of a transgene on the end of a broken chromosome showed that chromosomes can be elongated by unequal recombination; however the endogenous telomere sequences in this organism have not yet been characterized.

A strikingly different mechanism of telomere maintenance has been found in the genus *Drosophila*. In all studied species of this genus, specialized retrotransposons extend the telomeres. These retrotransposons provide a robust mechanism which may well be used by other members of the Endopterygota. The *Drosophila* telomere-specific retrotransposons also provide an unexpected link between chromosome structure and transposable elements, raising important questions about the evolution of both.

**Drosophila telomeres are maintained by specialized non-LTR retrotransposons.**

The three retrotransposons that maintain *Drosophila* telomeres are all non-LTR (Long Terminal Repeat) retrotransposons (Fig. 2). Non-LTR elements differ from their relatives, LTR retrotransposons and retroviruses, in the way in which the RNA transposition intermediate is converted to chromosomal DNA. Of course, like all transposons, they also differ from retroviruses by lacking the viral envelope gene. LTR retrotransposons and retroviruses are reverse transcribed in the cytoplasm, transported to the nucleus and inserted into the chromosome as double stranded DNA.

In contrast, non-LTR elements enter the nucleus as RNA, the 3’ end associates with a nick in the chromosome, and reverse transcriptase uses the 3’ OH on the nicked DNA to initiate reverse transcription of the RNA. Thus, new DNA is linked to the chromosome. The 5’ end of the reverse-transcribed DNA is then joined to the other side of the nick in the chromosome, to complete the insertion. At least for some elements, the insertion is then converted to double-stranded DNA by the same reverse transcriptase, although it is also possible that the second strand synthesis might be accomplished by normal DNA synthesis.

The mechanism by which non-LTR elements are incorporated into chromosomes is significant because it is basically equivalent to that used by telomerase (Fig. 3). We have postulated that telomeric retrotransposon RNA associates with the end of the DNA, rather than with an internal nick like other elements. Thus each transposition of a telomeric element adds a new end to the DNA, extending the chromosome (Fig. 3). Analysis of sequence at junctions between elements in *Drosophila* arrays shows that these elements do not use the precise sequence pairing that telomerase uses to align each repeat. Such alignment is necessary to obtain a consistent sequence when the template is only a few nucleotides in length. On the other hand, the *Drosophila*
templates range from six to thirteen kb so imprecise alignment at junctions would have little impact, especially since each junction starts with a variable run of As.

Our working model of the Drosophila telomere (Fig. 4) draws from retrovirus and yeast biology, as well as our own results. 28 Transcription of elements in telomere arrays produces RNA which is transported to the cytoplasm where the coding regions are translated to yield the structural protein, Gag, and, except for HeT-A, the enzymatic protein, Pol. These proteins associate with their own RNA and move into the nucleus where the Gag protein is targeted to telomere regions. We suggest that this targeting is directed by an end-associated protein, perhaps analogous to Est1 or Cdc13 in budding yeast.

**Drosophila telomere retrotransposons have special features.**

Retrotransposon telomeres were discovered in *D. melanogaster* and have been most extensively characterized in this species. We will confine discussion in this section to *D. melanogaster* elements and discuss other species later.

There are three telomere-specific retrotransposons in *D. melanogaster*, HeT-A, TART, and TAHRE (Fig. 2). Sequences of their gag and pol coding regions group these elements in the jockey clade of non-LTR retrotransposons. 29, 30 This clade also contains several of the parasitic elements abundant in non-telomeric parts of the *D. melanogaster* genome (e.g., jockey, Doc, and F). The primary structural feature that distinguishes the telomeric elements from other transposable elements is a very long 3’ UTR (untranslated region). Their relatives in the jockey clade, like transposable elements in general, have very little sequence that does not code for something needed for transposition.

The many available sequences of HeT-A and TART show that each element has multiple forms that differ both by nucleotide differences and insertions/deletions and yet appear to be fully functional. There are not yet enough TAHRE sequences to determine whether the TAHRE sequence is equally variable.

HeT-A is unusual because it does not encode its own reverse transcriptase. Nevertheless HeT-A is the most abundant of the elements 31 and the one most often found transposing to heal broken ends. 27, 32 It must obtain the necessary enzyme activity from another source – either another element or a nuclear reverse transcriptase. HeT-A expression is developmentally regulated. 33, 34 It is not expressed in polyploid cells and therefore there is little expression during most of larval growth. Also, HeT-A Gag is efficiently targeted to telomeres in interphase diploid nuclei, but not in polyploid cells. 28, 35 This targeting appears to be important for telomere-specific transposition.
TART encodes both Gag and Pol and is always less abundant than HeT-A. Both elements are present in every Drosophila stock, cell line, and species we have studied. Measurements across different stocks and cell types\textsuperscript{31} show a strong correlation in the relative abundance of the two elements, even where the total number of elements differs significantly (Fig. 5). Unlike HeT-A, TART produces both sense and antisense transcripts. Antisense transcripts are much more abundant and display little, if any, developmental regulation.\textsuperscript{33, 36} TART Gag, like HeT-A Gag, is efficiently transported to interphase nuclei; however, it does not associate with telomeres by itself. When HeT-A Gag is present the two proteins colocalize to telomeres.\textsuperscript{28} This suggests that the two elements collaborate, with HeT-A providing telomere specificity and TART providing reverse transcriptase. Such collaboration could explain why the two elements are found together in every stock or cell line studied.

TAHRE (Telomere-Associated and HeT-A-Related Element) was discovered recently\textsuperscript{37} and has been less studied than the others. However much of its sequence is so strongly related to HeT-A that some conclusions can be drawn from experiments using HeT-A probes. The 5' and 3' UTR and gag coding regions are very similar to HeT-A while the pol coding region resembles, but is less closely related to, TART. TAHRE is very rare. One complete and three truncated copies were reported from a study of BACs made for the D. melanogaster Genome Project. The expression pattern has not been reported but a BLAST search of the database of cDNA clones found no sequences except those that were also found with HeT-A sequence. If abundance in the telomere array is determined by the ability to transpose, it is surprising that an element combining HeT-A’s ability to target telomeres with TART's reverse transcriptase is not more abundant.

The powerful genetic tools available only for Drosophila have made it possible to produce chromosomes with broken ends that evade checkpoints and can then be retained in the genome by strong selection. These experiments have shown that broken ends can be healed by transposition of telomere retrotransposons. More recently, it has been shown that the rate of healing is affected by specific genes and that these genes are acting through components of the RNAi machinery.\textsuperscript{38, 39}

Telomere retrotransposons are almost completely segregated from other transposable elements in the genome.

In spite of their similarity to other retroelements in the D. melanogaster genome, the telomere retrotransposons differ markedly from those elements in transposition targeting. As a result, there is little mixing of the two types of elements. The euchromatic regions of the D. melanogaster genome have been sequenced completely. No sequence with significant similarity to HeT-A, TART, or TAHRE is found in these regions except for the pol coding region of BS, a non-LTR element found at
several sites in euchromatin, which has a small region with similarity to 90 bp of TART pol. Thus there is no evidence that telomere elements can transpose into these euchromatic regions, although other retrotransposons are found at many sites in this part of the genome.

The exclusion of telomere elements from euchromatin may be explained by their specific targeting to ends. (Telomere elements do transpose onto euchromatin if a chromosomal break causes that euchromatin to be at the end of the chromosome.) Even if targeting is not perfect, internal insertion might also be forbidden for other reasons. For example, telomere elements may unable to form a junction at the 5′ end after reverse transcription, and the resulting loss of the part of the chromosome distal to an internal insertion would likely be lethal.

As noted above, telomere arrays and telomere-associated sequences present formidable challenges for correct assembly of sequence. However, the Drosophila Heterochromatin Genome Project now has assembled sequence extending into the telomeres on the right end of chromosome 4 (4R) and the left end of the X chromosome (XL). These assemblies (Fig. 6) contain 75,946 bp of telomere transposons on 4R and 19,199 bp of these sequences on XL. It might be supposed that these long telomere arrays would be safe landing sites for parasitic elements because they contain no vital genes to be disrupted. This does not seem to be the case. In both XL and 4R, the interior of the chromosome, peppered with non-telomeric elements, is separated from the distal array of uninterrupted telomere elements by a short transition region comprised of mixed fragments of telomere and non-telomeric elements. On XL the transition region is approximately 300 bp and the assembled portion of the distal array of uninterrupted HeT-A elements is ~19 kb. On 4R the transition zone is 5.4 kb and the assembled HeT-A/TART array is 70.6 kb. The transition zones show that the distal separation is not due to incompatibility of the DNA sequences. Thus the lack of mixing in the distal arrays is probably due to a specific chromatin structure in the distal regions which has been partially invaded by non-telomeric elements at the proximal end.

The distinction between distal telomere arrays and telomere-associated sequences affects not only elements that are long-time residents in the D. melanogaster genome but may have at least a partial effect on the recent invader, the P-element. P-elements were discovered because they disrupt chromosomes when they invade a naïve host and have, therefore, become a powerful tool for geneticists wishing to manipulate chromosomes. Attempts to insert P-elements in telomere regions have found hotspots for insertion in telomere-associated sequences but there is only one report of inserts in the telomere array. The collection of ~20,000 random P-element inserts produced by the Drosophila genome project was screened for elements flanked by telomere retrotransposon sequence. Seven inserts were identified that mapped to telomere arrays. Most were inserted in a short region very near the 3′ end of TART; a smaller hotspot for insertion was seem near the 3′ end
of HeT-A. Thus the P-element may be revealing the beginning of a footprint of the telomere chromatin structure. It is interesting that the only retrotransposon that has been found within the telomere array, a roo element, was found near one of the P-element inserts, suggesting that the P-element may have affected chromatin structure, allowing entry of roo.

The apparent exclusion of non-telomere elements from the distal array is also seen in telomeres of the silkworm, Bombyx mori, which has extremely long TTAGG arrays made by telomerase. B. mori has two families of retrotransposons, TRAS and SART, which insert at specific nucleotides in the TTAGG sequence. TRAS and SART are abundant in proximal parts of the TTAGG array but are not found in the distal six to eight kb. As with the Drosophila telomere, terminal insertions in B. mori seem to be prevented by something other than lack of DNA insertion sites, but nothing is known about its chromatin structure. Other insect species have yet to be studied.

**Very long 3’ UTR sequences seem to have a role in forming heterochromatin structure.**

One of the distinguishing features of telomere retrotransposons is their very long 3’ UTR (Fig. 2). We have suggested that this sequence plays a role in forming telomeric chromatin which, as Muller observed, is heterochromatic. The 3’ sequence is overrepresented in telomere arrays because some of the elements are truncated at the 5’ end, either because of incomplete reverse transcription or because of erosion during the time when they form the extreme end.

This 3’ sequence is also found in another class of D. melanogaster heterochromatin, the Y chromosome. Like other heterochromatin, the Y presents significant problems for sequence assembly; however sequence scaffolds of seven Y chromosome genes have been assembled. Four of these contain fragments of HeT-A or TART. These fragments are distinguished from sequences in telomere arrays in two important ways. First, they have been inserted into the interior of the chromosome, rather than added to the end. Second, the fragments contain only sequence from the 3’ UTR and some do not contain the extreme 3’ sequences thought necessary for reverse transcription. This suggests that the sequences have been transposed to the Y by some other mechanism. Much of the Y sequence has not been assembled. There is evidence that there is more 3’ sequence in unassembled parts of the Y and perhaps other unassembled heterochromatic regions of the genome.

These observations of the segregation of 3’ UTR sequence into telomeres and other heterochromatic regions shows that these sequences have a special relation to this type of chromatin, possibly because they are involved in forming its structure.

**Telomere retrotransposons have a symbiotic relationship with Drosophila cells.**
Our hypothesis that telomeric transposons are targeted to telomeres by Gag proteins (Fig. 4) initially had two bases. First, although gag genes of individual HeT-A and TART elements differ by both indels and nucleotide changes, all these coding regions are open, suggesting that each element must be successfully translated in order to transpose.\(^{31}\) Second, retroviral Gags had been shown to escort viral RNA through the transport path specific for their virus.\(^{49}\) HeT-A and TART Gags share important motifs with retroviral Gags and we surmised they also share functions. Transient expression in cultured cells of HeT-A and TART Gags tagged with Green Fluorescent Protein supported this hypothesis; in these cells, which normally express both telomeric transposons, Gags of both elements are efficiently transported into the nucleus where HeT-A Gag moves to telomeres and also directs TART Gag to the same targets.\(^{28, 50, 51}\)

These studies of Gag localization show that the telomere elements have co-evolved with their hosts an ability to interact beneficially with cellular components. This ability is not seen in their non-telomeric relatives. Direct comparison with Gags from jockey, Doc, and I factor showed that these proteins were mostly, if not entirely, constrained to remain in the cytoplasm.\(^{51}\) We suggest that cytoplasmic retention is a reflection of the cell’s efforts to keep non-telomeric elements out of their chromosomes.

We have also expressed HeT-A Gag in live flies from a transgene driven by a promoter that is active in all tissues of the fly. Normally, HeT-A is active in diploid cells and, in these cells, transgenic Gag forms dots as it does in cultured cells. Polyploid cells do not express HeT-A and, when transgenic HeT-A Gag is expressed in polyploid tissues, the protein does not enter the nucleus. Instead large amounts of the protein accumulate in cytoplasmic regions that differ from tissue to tissue.\(^{35}\) These transgenic experiments show that cells actively regulate Gag targeting in a cell-type-specific manner.

**Retrotransposon telomeres probably predate the genus Drosophila**

Telomere retrotransposon sequences diverge rapidly. For example, the six complete HeT-A elements found in the assembled 4R and XL sequence from D. melanogaster have between 68% and 99% nucleotide identity when pairwise comparisons are made. Even the coding sequence has only 80% to 100% nucleotide identity (76% to 100% amino acid identity), depending on the elements compared.\(^{31}\) This divergence makes it difficult to search genomes of other species on the basis of sequence homology. Nevertheless we have found both HeT-A and TART homologues in every Drosophila species we have studied,\(^{52-55}\) including D. virilis, a species originally reported to depend on recombination to maintain its telomeres.\(^{56}\)
Our search for *D. virilis* telomere elements was initiated by using the most conserved part of the *D. melanogaster* TART pol gene in low stringency hybridization experiments to search for a fragment that could then be used to probe a lambdaphage library of *D. virilis* DNA. We found a cross-hybridizing fragment in *D. americana*, a species closely related to *D. virilis*. With this fragment, two phage in the *D. virilis* library were found and sequenced. Both had tandem copies of TART, showing that our strategy had been successful. The 3' end of one TART was joined to 5' sequence of an unidentified element which had been truncated by cloning. This 5' sequence was used to reprobe the lambda phage library. Two new phage were selected and sequenced. Both contained tandem copies of *HeT*-A, revealing that the truncated element in the TART clone was *HeT*-A. The tandem array of elements in one of these phage also contained a novel element, *Uvir*.

*Uvir* looks like a non-LTR retrotransposon that has 5' and 3' UTRs very similar to those of *HeT*-A but lacks a Gag coding region; instead it has a coding region that most closely resembles the Pol coding region of *jockey*. Because it encodes Pol, but not Gag, *Uvir* represents a new kind of non-LTR element and it is not clear that it is a successful one. Searches through the *D. virilis* genome database show only a few partial copies. However, it is important to note that even the *D. melanogaster* genome, the first and most thoroughly sequenced Drosophila genome, is now the focus of a large Heterochromatin Genome Project to sequence the repeated parts of the genome. The project is finding new sequence and correcting assembly of other sequence in *D. melanogaster*. Much less is known about other Drosophila genomes.

Hybridization to total *D. virilis* DNA shows that there are very few copies of the *Uvir* pol gene. We have suggested that the *Uvir* ORF might come from a cellular reverse transcriptase. If so, that cellular gene might be the ancestral source of enzyme for *HeT*-A. Indeed, it might still be functioning in *HeT*-A transposition.

Although the *D. virilis* elements, *HeT*-A<sup>vir</sup> and TART<sup>vir</sup>, differ markedly from their homologues in other species (Fig. 7), their sequences also group in the telomere element group of the jockey clade of non-LTR elements. The cloned sequences are found in mixed tandem arrays and hybridize only to telomere regions of *D. virilis* polytene chromosomes. Thus there is strong evidence that these *D. virilis* elements are true homologues.

In addition, *HeT*-A<sup>vir</sup> has the same structural features that distinguish *HeT*-A<sup>mel</sup> from other non-LTR elements; it has unusually long 3' UTRs with an irregular pattern of A-rich repeats and does not encode reverse transcriptase. TART<sup>vir</sup> differs more markedly from TART<sup>mel</sup>, having a significantly shorter 3' UTR and a large domain of unknown function (the X domain) at the C-terminal end of its Pol coding region. A similar C-terminal domain is seen in *D. americana* but not in other species. TART<sup>vir</sup>, like TART in other species, yields both sense and antisense transcripts. The conservation of
unusual features in spite of marked sequence change suggests that these features are important for telomere function.

One of the remarkable properties of HeT-A\textsuperscript{mel} is the specific localization of its Gag protein to telomeres in interphase cells.\textsuperscript{28} It seems likely that this localization is important for targeting transposition to chromosome ends. HeT-A\textsuperscript{vir} Gag shows similar telomere localization in spite of the large difference in the amino acid sequences of the two proteins. These species-specific differences in amino acid sequence might be driven by need to coevolve with the various cellular components that Gag must interact with as it moves to telomeres, but this does not seem to be the case, HeT-A\textsuperscript{vir} Gag forms Het dots when it is expressed in \textit{D. melanogaster} cells and HeT-A\textsuperscript{mel} forms Het dots in \textit{D. virilis} cells. Thus both proteins interact appropriately with cellular targeting proteins in the other species.\textsuperscript{57}

Gag sequences of \textit{TART}\textsuperscript{vir} and HeT-A\textsuperscript{vir} are about equally diverged from their \textit{D. melanogaster} homologs but \textit{TART}\textsuperscript{vir} Gag targeting differs from that of \textit{TART}\textsuperscript{mel} Gag.\textsuperscript{57} In both \textit{D. melanogaster} and \textit{D. virilis} cells, \textit{TART}\textsuperscript{mel} Gag enters the nucleus and interacts with HeT-A\textsuperscript{mel} Gag to be directed to telomere regions. These localization experiments were done with tagged proteins transiently expressed in cultured cells. In similar experiments, \textit{TART}\textsuperscript{vir} Gag entered the nucleus only if the last \textasciitilde 200 amino acids had been deleted. We believe that the behavior of \textit{TART}\textsuperscript{vir} Gag is a reflection of the experimental design, rather than its normal behavior, but, in either case, \textit{TART}\textsuperscript{vir} Gag differs more from \textit{TART}\textsuperscript{mel} Gag than HeT-A\textsuperscript{vir} Gag differs from HeT-A\textsuperscript{mel} Gag.

Given that \textit{D. melanogaster} and \textit{D. virilis} are about as widely separated as any members of the Drosophila genus (~ 60 My),\textsuperscript{58} it is reasonable to assume that retrotransposon telomeres antedate the genus and will eventually be found in other Diptera.

\textbf{Drosophila telomeres resemble other telomeres both structurally and functionally}

Transposable elements are generally considered to be parasitic DNA. HeT-A, \textit{TART}, and \textit{TAHRE} are the first elements that appear to be entirely beneficial to the cell. At first glance, Drosophila telomeres seem very different from those produced by telomerase but in fact the two telomeres are basically very similar. Both are extended by reverse transcription of RNA templates that produces long arrays of tandem repeats. In fact, telomerase appears to be closely related to the reverse transcriptase of non-LTR retrotransposons.\textsuperscript{59} Both kinds of telomeres can be extended by recombination-based mechanisms but, for both, this mechanism appears to be primarily a back-up mechanism.\textsuperscript{11-14, 60-62}

The three elements that are the Drosophila repeats are much longer than repeats produced by telomerase but the total length of the telomere array is similar to telomere arrays in other
multicellular eukaryotes. The lengths of these arrays fluctuate around a set point and, in organisms as different as yeast and man, that set point can be changed by genetic background and environment.\textsuperscript{5, 63, 64} There are technical difficulties in accurately measuring Drosophila telomere length; however, several genes have been shown to affect the length set point or to affect the rate of transposition to a broken chromosome end.\textsuperscript{31, 32, 38, 61, 62} Thus, length regulation is seen with both types of telomeres. It may be that RNA-templated extension is the predominant mechanism for telomere maintenance because it can be easily regulated and produce rapid change in length.

An increasing number of proteins are considered to be telomere-associated either because they have been found on telomeres or because mutation of the protein causes chromosome end stickiness. Many proteins that are telomere-associated in other organisms are also telomere-associated in Drosophila. The list includes proteins involved in DNA damage response and repair, such as ATM, RAD50, MRE11, Ku70 and Ku80 and chromatin structure, such as HP1 (see ref. \textsuperscript{65} for review).

It is sometimes incorrectly said that Drosophila telomeres are unusual because they do not need special telomere sequences; this is based on a misinterpretation of some experiments that have been done with Drosophila. The straightforward experiment of simply inducing a chromosome break in a mitotic cell results in activation of a checkpoint and eventual cell death for all organisms studied, including Drosophila.\textsuperscript{66} In contrast to this simple experiment, Drosophila experiments that allow recovery of broken chromosomes without a cap of telomere DNA are multigenerational experiments that involve checkpoint evasion followed by strong selection to maintain the broken chromosome in subsequent generations. These experiments utilize genetic tools not available in other organisms and, in fact, offer clues to telomere behavior that may also pertain to other organisms.

Drosophila geneticists have found two ways to produce broken chromosomes that evade checkpoints. Breaks can be induced in the ovaries of \textit{mu2/mu2} females\textsuperscript{67} or they can be induced by \textit{P-element} transposition.\textsuperscript{68} In either case, once through the checkpoint, the broken end will not activate a checkpoint in subsequent cell cycles. Broken chromosomes that slip through one checkpoint have received scant attention in other organisms; however, there is one study in budding yeast\textsuperscript{69} that suggests that checkpoint-evaders will not be stopped at that checkpoint in subsequent divisions in organisms other than Drosophila. In these yeast experiments, after being held by a check point for a period of time, some broken chromosomes managed to complete the cell cycle. Although those that completed the cell cycle did not activate the checkpoint in later cell divisions, they showed a relatively high rate of loss in subsequent divisions. Sandell and Zakian point out that this experiment demonstrates two critical functions of telomeres in yeast: telomeres distinguish ends
from breaks and also prevent chromosome loss. These two functions are separated in time; end identification is needed only before the first checkpoint is passed while the tendency for broken chromosomes to be lost continues through subsequent cell generations.

Similarly, in the Drosophila experiments, end identification is needed only for passing the first checkpoint because broken ends induced in a mu2/mu2 background can be maintained in a wild type background after the first checkpoint. The loss of broken chromosomes in these Drosophila experiments was prevented by other genetic tools that select for the broken chromosome. For example, a broken X chromosome can be retained by a mating scheme that causes the broken X to be passed only from father to son. Because there is only one X in males, this broken chromosome is essential for maintenance of the stock and all surviving males will carry the broken chromosome. In these experiments, broken chromosomes shorten by an average of seventy nucleotides per fly generation, a rate consistent with loss of the terminal RNA primer in DNA replication.

The experiments just described dissect some aspects of telomere function, giving clues that probably apply to other organisms. Telomeres distinguish ends from breaks, but only until the first checkpoint is passed. Lack of telomere sequences does not necessarily make a “sticky end” because surviving broken chromosomes do not form end-to-end attachments. In fact the presence or lack of telomere DNA may not be relevant to “sticky ends” because both Drosophila and human chromosomes can form end-to-end junctions that contain significant amounts of telomere DNA. Without a telomere, chromosomes shorten but this happens so slowly that it could be many generations before a vital gene activity is lost, even if the end is not healed by transposition of retrotransposons.

It should be noted that these Drosophila which carry a broken chromosome do not test several important aspects of telomere function. These flies are living in relatively non-stressful conditions and, importantly, the broken chromosome has no competition from a telomere-containing homolog until a healed chromosome begins to take over the line. Thus only the most drastic effects on fitness will be detected. In addition, all of the other chromosomes have normal telomeres so the loss of one telomere may have little effect on phenotypes that depend on the total amount of cellular telomere sequence or on the organization of all chromosomes in the nucleus.

Evolution of retrotransposon telomeres

Understanding the origin of Drosophila telomeres would be a significant step toward understanding the evolution of both telomeres and transposable elements. There are several possibilities: 1) Telomerase could be the ancestral mechanism and the Drosophila telomeres could have evolved from telomerase, 2) Drosophila telomeres could be remainders of the ancestral mechanism, 3) Drosophila telomeres could be derived from transposable elements that had no
relation to telomeres until they were co-opted to substitute for a lost telomerase. None of these explanations can be eliminated with confidence. However we will offer one possible scenario that is consistent with what is now known and also satisfies the requirements of Occam’s razor in that only one reassortment of existing genes is required for the minimally needed functionality; other changes would be the natural result of long-term coevolution with the Drosophila cell.

We have suggested that telomerase is the ancestral mechanism and that telomeric retrotransposons are derived from telomerase.\textsuperscript{73} We speculate that somewhere in the lineage leading to Drosophila, the gene for one of the proteins required to deliver telomerase to its target fused to the gene for the telomerase RNA template. This would be analogous to having a translocation fuse the 3’ end of the *Cdc13* gene to the *Tlc1* gene in yeast. Transcripts of this compound gene would still be translated to yield a protein designed to take the template RNA to the telomere where it would be reverse transcribed by the catalytic subunit of telomerase. The compound gene, comprised of a coding region (derived from the telomerase-related protein) and a long 3’ UTR (derived from the telomerase RNA template), could be the ancestor of *HeT-A*.

This hypothesis provides a relatively easy transition between telomerase and telomeric retrotransposons. The two mechanisms might even coexist for some time in diploid cells because both the retrotransposon-encoded protein and its RNA should still be able to interact with cellular components necessary for telomerase function. Our hypothesis also suggests that the gene for the telomerase catalytic subunit might persist in the Drosophila genome. The *D. melanogaster* genome has no sequence with all the hallmarks of telomerase but these might not have been conserved after the RNA template changed so drastically.

The invariant partnership between *HeT-A* and *TART* suggests that *TART* subsequently might have been coopted to take over the enzymatic function from the telomerase catalytic subunit. *TART* differs from *HeT-A* in its promoters, the organization of its untranslated regions, and its pattern of transcription. In fact, the predominant similarity between the two elements is in their Gag proteins, which target them to telomeres, this suggests that *TART* may have been a preexisting retrotransposon that acquired Gag coding from *HeT-A* and therefore become targeted to telomeres. The partnership with *TART* might be favored because it offers *HeT-A* more sources of enzyme activity than the single copy telomerase catalytic subunit. Recent evidence that transposition and expression of *HeT-A* and *TART* are sensitive to disruptions in the RNAi pathway suggests another possible advantage of a *HeT-A/TART* partnership.\textsuperscript{38, 39} *TART* appears to be the principal target of this newly recognized regulation mechanism with *HeT-A* possibly controlled by *TART*. As mentioned above, the number of *HeT-A* and *TART* elements per genome varies in a correlated manner between different stocks and cell lines.\textsuperscript{31}
It has recently been reported that LINE-1 elements lacking endonuclease activity can transpose in an orientation-specific manner onto telomere ends in Chinese Hamster cells that have dysfunctional telomeres caused by loss of DNA-PKcs. This study and earlier work showing that these endonuclease-independent elements can integrate into internal DNA lesions provide support for the authors’ suggestion that non-LTR retrotransposons served to repair DNA lesions before these elements acquired endonuclease activity. It would be tempting to suggest that the telomere retrotransposons owe their end-specificity to lack of endonuclease activity. However, the pol gene sequences of TART, TAHRE and Uvir all have well-conserved endonuclease coding sequences, suggesting that these sequences are important even for transposition to chromosome ends.

It has been suggested that TAHRE is an ancestral element from which HeT-A was derived by loss of the Pol coding. However, HeT-A is much more abundant than TAHRE and thus apparently more successful. It is not obvious why loss of an essential function should make the progeny more successful than the parent; although such an outcome might be evolutionarily favored, if cellular well-being requires closely regulated rapid changes in telomere length. On the other hand, TAHRE could have arisen from HeT-A by acquiring a pol gene, just as retroviral oncogenes have been acquired from the cellular genome. If so, the scarcity of TAHRE suggests that combining the two activities is not beneficial in this environment.

Both TAHRE in D. melanogaster and Uvir in D. virilis have 5' and 3' UTR sequences very closely related to the HeT-A elements of their respective species (Fig. 8). Although it is not possible to say whether coding region differences in these elements are the result of gain or loss, taken together, HeT-A, TAHRE, and Uvir show that sequence changes within HeT-A-related UTR sequences may be relatively frequent on an evolutionary time scale. Because these three of the four elements found in telomere arrays have related 3' and 5' UTR sequences, these sequences must have a special role in telomeres. It will be important to look for more elements so that we can use sequence analysis to deduce the history of their components.

Conclusions

It is intriguing that retrotransposons have so completely adapted to an essential cellular role. Although the ways in which HeT-A and TART have coevolved to perform these functions are interesting in their own right, and important to our understanding what is essential to the role of the telomere, we find the more general evolutionary implications most fascinating.

It should be noted that the scenario for deriving retrotransposons from the telomerase machinery, described above, could also occur in organisms that do not lose telomerase. As noted above, if the organism is diploid, modification of one copy of telomerase components leaves the other
copy in the genome functional. The modified copy of the telomerase components might be then be lost from the population after the newly generated retrotransposon has occupied other sites in the genome. Of course, telomerase components may not be the only cellular genes that can give rise to retrotransposable elements. Thus there may have been multiple sources of retrotransposons in different organisms. The variant telomeres of Drosophila raise many questions about the evolution of telomeres and of transposable elements, two topics about which we know little.

ACKNOWLEDGEMENTS
Work in the authors’ laboratory is supported by National Institutes of Health Grant 50315
FIGURE LEGENDS

Figure 1. Some species in the Insecta do not use telomerase to maintain telomeres. Phylogeny of insect orders where one or more species have been analyzed for the presence of telomeric TTAGG repeats. (+): all species studied have TTAGG; (-): all species studied lack TTAGG; (+/-): some species studied have TTAGG, others do not. Only species from three genera without TTAGG have been studied further, Drosophila, Chironomus, and Anopheles. All are Diptera (see text for discussion). At least one non-insect, the spider Tegenaria ferrugenea also lacks TTAGG. Tree based on Frydrychova, et al.17

Figure 2. The three D. melanogaster telomere retrotransposons drawn as their putative RNA transposition intermediates. Coding regions, Gag and Pol, are labeled. Gray regions indicate 5’ and 3’ untranslated regions. AAAA Indicates the 3’ poly(A) tail on each RNA. It is the source of the (dA/T)n that joins each DNA copy to the chromosome when the element transposes. Sizes are only approximate because individual elements can differ in length of both coding and non-coding regions. HeT-A elements are ~ 6 kb. The 5’ end of TART has not been completely defined but subfamilies appear to be 10-13 kb. TAHRE is ~ 10.5 kb.

Figure 3. Telomere element retrotransposition resembles telomere extension by telomerase. In both cases the catalytic subunit (gray) with its RNA template (black wavy line) associates with the end of the chromosome. Telomerase aligns the first nucleotides of the sequence to be copied with their complement in the chromosome before copying sequence on to the end of that complement, thus assuring precise replication for each addition. HeT-A, and other telomere elements, do not require complementary sequence for alignment so the sequence to which the initial Ts are added is shown as NNNN. Retrotransposon additions copy variable amounts of the poly(A) tail at each transposition.

Figure 4. Model for maintenance of chromosome ends by telomeric retrotransposons. Retrotransposons yield sense-strand transcripts that serve as both mRNAs and transposition intermediates. This diagram shows our current model for the path of these RNAs from transcription until they are reverse-transcribed to add another repeat onto the telomere array. Gray arrows represent HeT-A (dark) and TART (light) elements attached to the end of the chromosome. A poly(A) sense strand RNA is transcribed from a member of the array (step 1). For the telomeric retrotransposons there is evidence suggesting that this RNA must be translated (step 2) before serving as a template (step 3) for telomere addition. This suggestion is now supported by the finding
that translation products (Gags) of these RNAs appear capable of delivering the transposition template specifically to its target at the telomere. Gray circles in the diagram represent Gags of either Het-A or TART. Analogy with retroviruses suggests that reverse transcriptase is also included in the Gag-RNA complex; however, there is no evidence on this point. Reproduced from J. Cell Biol. (2002) 158:398 by copyright permission of the Rockefeller University Press.

Figure 5. The number of Het-A and TART elements per genome are correlated in D. melanogaster stocks, cells, and tissue types. This figure shows data for a cultured cell line (S2), for diploid cells, and for polytene salivary gland cells from four stocks (Oregon R, 2057, Su(var) 2054, and G3). When analyzed, the data show that the number of the two elements present are linearly correlated at better than the 95% confidence level when polytene salivary gland measurements are compared, when diploid cell measurements are compared, or when all data is pooled and compared.31 The solid black line is the best linear fit to the data, the dotted line is the best quadratic fit. If the data point in the extreme right-hand upper corner is omitted from the analysis, the best linear fit for the remaining data is indistinguishable from the quadratic fit at the level of detail in this plot.

Figure 6. The assembled regions of telomere arrays from two D. melanogaster telomeres. The figure, not drawn to scale, illustrates sequence organization from the most distal assembled sequence (left end) to the most distal gene (right end). The more terminal sequences have not been assembled. Upper diagram: array from the left end of the X chromosome. Lower diagram: array from the right end of chromosome 4. Black boxes: full-length telomeric elements. Gray boxes: partial telomeric elements. Boxes marked T are TART elements, all other boxes are Het-A elements. All elements have 3’ end toward chromosome interior. All partial elements are truncated at the 5’ end except for one in the transition zone. Differences in size of partial elements not shown. White regions: other transposable elements or regions rich in these elements. Striped regions: the most distal gene on each chromosome. trans zone: the tiny (320 nt) transition zone on the XL telomere.

Figure 7. Comparisons of Het-A and TART elements from three Drosophila species. Elements are drawn approximately to scale but individual elements vary in length of both coding and non-coding regions. Because individual elements differ in sequence, % identity differs depending on elements compared. Numbers shown are typical. Dark gray: 5’ and 3’ untranslated regions. Light gray: Gag and Pol coding regions. Scale on right indicates separation of species in millions of years. For Het-A (top): Full length arrows between species indicate % nucleotide identity between elements.
in that pair of species. Note that, although *D. virilis* is much more distant from *D. melanogaster* than is *D. yakuba*, the *D. virilis* element shows significant nucleotide identity, showing strong conservation of sequence in the 5' and 3' untranslated regions. Note also that in both comparisons the nucleotide sequence of the Gag coding region is more conserved than is the amino acid sequence. For TART (bottom): Only the coding regions are compared and the X domain found only in *D. virilis* Pol is not included. The available sequences of *D. yakuba* TART are 5'-truncated so only a partial *gag* sequence is shown. The untranslated regions are too different for any meaningful alignment of any two species. ? indicates that the 5' end of TART*mel* has not been completely defined.

**Figure 8.** *HeT-A* sequences are strongly conserved in *TAHRE* and *Uvir*. Elements are drawn approximately to scale. In *D. melanogaster*, *TAHRE* sequence is highly similar to *HeT-A* *mel* in the *gag* gene (light gray) and most of the untranslated regions (horizontal black stripe). In *D. virilis*, *Uvir* is highly similar to *HeT-A* *vir* in the 5' UTR and the last ~600 nt of the 3'UTR (horizontal black stripe). Dark gray regions indicate 3'UTR sequences specific to *TAHRE* or *Uvir*.

**REFERENCES**


Figure 2

HeT-A

TART

Tahre

Gag   Pol   AAAAAA

Gag   AAAAAA

Gag   Pol   AAAAAA
Fig 5
Figure 7
Figure 8

D. melanogaster

HeT-A
Gag
AAAAA

Tahre
Gag
Pol
AAAAA

D. virilis

HeT-A
Gag
AAAAA

Uvir
Pol
AAAAA