Intracellular targeting of telomeric retrotransposon Gag proteins of distantly related Drosophila species

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://dx.doi.org/10.1073/pnas.0702566104">http://dx.doi.org/10.1073/pnas.0702566104</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Nov 06 12:56:49 EST 2017</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/76237">http://hdl.handle.net/1721.1/76237</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>
Title: Intracellular targeting of telomeric retrotransposon Gag proteins of distantly related Drosophila species

Authors: Elena Casacuberta\textsuperscript{1}, Fernando Azorín Marín\textsuperscript{1} and Mary-Lou Pardue\textsuperscript{2}

Authors affiliation:\textsuperscript{1} Institute of Molecular Biology of Barcelona, CSIC and Institute for Research on Biomedicine of Barcelona (IRB).

\textsuperscript{2} Department of Biology, Massachusetts Institute of Technology

Cambridge, MA 02139

Corresponding Author: Dr. Mary-Lou Pardue, Biology Department, 68-670 Massachusetts Institute of Technology

Cambridge, MA 02139 Phone: 617-253-6741, FAX: 617-253-8699

e-mail: mlpardue@mit.edu

Manuscript information: Text pages: 16

Figures: 5

Word and character count: Abstract: 240 words

Characters and Spaces: 38942 (text) plus 7734 (figures)

Abbreviation footnote: MHR: major homology region
ABSTRACT

The retrotransposons that maintain telomeres in *D. melanogaster* have unique features that are shared across all Drosophila species but are not found in other retrotransposons. Comparative analysis of these features provides insight into their importance for telomere maintenance in Drosophila. Gag proteins encoded by *HeT-A*<sup>mel</sup> and *TART*<sup>mel</sup> are efficiently and cooperatively targeted to telomeres in interphase nuclei, a behavior that may facilitate telomere-specific transposition. *D. virilis*, separated from *D. melanogaster* by 60 million years (MY), has telomeres maintained by *HeT-A*<sup>vir</sup> and *TART*<sup>vir</sup>. The Gag proteins from *HeT-A*<sup>mel</sup> and *HeT-A*<sup>vir</sup> have only 16% amino acid identity, yet several of their functional features are conserved. Using transient transfection of cultured cells from both species, we show that the telomere association of *HeT-A*<sup>vir</sup> Gag is indistinguishable from that of *HeT-A*<sup>mel</sup> Gag. Deletion derivatives show that organization of localization signals within the two proteins is strikingly similar. Gag proteins of *TART*<sup>mel</sup> and *TART*<sup>vir</sup> are only 13% identical. In contrast to *HeT-A*, surprisingly *TART*<sup>vir</sup> Gag does not localize to the nucleus, although *TART*<sup>vir</sup> is a major component of *D. virilis* telomeres and localization signals in the protein have much the same organization as in *TART*<sup>mel</sup> Gag. Thus the mechanism of telomere targeting of *TART*<sup>vir</sup> differs, at least in a minor way, from that of *TART*<sup>mel</sup>. Our findings suggest that, despite dramatic rates of protein evolution, protein and cellular determinants that correctly localize these Gag proteins have been conserved throughout the 60 MY separating these species.

Introduction

*D. melanogaster* has unusual telomeres formed by retrotransposons (reviewed in 1). Because most species have telomerase-maintained telomeres, it is important to determine how *D. melanogaster* has come to use a variant mechanism for serving the same telomere functions. To learn more about the evolution of the retrotransposon telomere we have studied *D. virilis*, separated from *D. melanogaster* by 60 MY (2, 3, 4).

*D. melanogaster* telomeres are composed of two non-LTR retrotransposons, *HeT-A* and *TART*, with few copies of *Tahre* (5), an element that appears to combine parts of both *HeT-A* and *TART*. We have found homologs of both *HeT-A* and *TART* in *D. virilis* (Fig. 1). Because retrotransposons have rapidly changing sequences, we used low stringency hybridization with the most conserved part of the *D. melanogaster* telomere sequences, the *TART pol* gene, to clone a fragment of *D. virilis* DNA. This homologous probe was used to screen a genomic library and obtain large clones containing tandem arrays of *TART*<sup>vir</sup> elements. One *TART*<sup>vir</sup> was adjacent to a *HeT-A*<sup>vir</sup> element. This *HeT-A*<sup>vir</sup> sequence was used to obtain clones with more
elements in tandem arrays. We have not detected a *Tahre* homolog, although we have found one copy of *Uvir*, a novel element apparently formed by replacement of the *HeT-A* Gag gene with sequence encoding reverse transcriptase. The *Uvir* element was found within a tandem array of *HeT-A*\textsuperscript{vir} elements.

*HeT-A* and *TART* elements have undergone significant sequence divergence between *D. melanogaster* and *D. virilis*; nevertheless, in both species these elements transpose specifically forming head-to-tail arrays on chromosome ends. Furthermore, in analyses of the relationship of these sequences to other elements from the same phylogenetic clade, the resulting tree shows that these elements are true homologs of *HeT-A* and *TART*. Although there may be other subfamilies of these elements in the *D. virilis* genome, we have not found any in the sequence now in the database.

Studies of the telomere retrotransposons in *D. melanogaster* show that *HeT-A* and *TART* both have unusual features not seen in other non-LTR retrotransposons (Fig.1). Both elements have unusually long 3’ untranslated regions (3’ UTRs). *HeT-A*’s 3’ UTR has an irregular pattern of A-rich repeats. *HeT-A* does not encode reverse transcriptase. *TART* has large Perfect Non-Terminal Repeats (PNTR, sequence at the 5’ end that is perfectly repeated near, but not at, the 3’ end). An important feature of *HeT-A*\textsuperscript{mel} and *TART*\textsuperscript{mel} is the interactive targeting of their Gag proteins to chromosome ends in interphase nuclei, providing a potential explanation for the telomere-specific transpositions of these elements. This targeting depends on sequence in several parts of the Gag proteins (6). It also depends on cell-type-specific contributions of host cells. For example, polyploid larval cells do not support nuclear localization of *HeT-A*\textsuperscript{mel} Gag, which instead accumulates in cell-type-specific cytoplasmic locations (7).

The telomere transposons of *D. virilis* have conserved many of the unusual features (Fig. 1). At the level of resolution in figure 1, *HeT-A*\textsuperscript{vir} is nearly identical to *HeT-A*\textsuperscript{mel}, *TART*\textsuperscript{vir}, on the other hand, differs from *TART*\textsuperscript{mel} in several ways. It lacks the PNTRs in the 5’ and 3’ UTRs of *TART*\textsuperscript{mel}. It has a much shorter 3’UTR and its *pol* gene encodes an extra domain of unknown function, the X domain (seen in the closely-related *D. americana* but not in *TART* from other species (2)). This description is based on two complete and two partial *TART*\textsuperscript{vir} elements from two phage clones. As mentioned, repeated searches in the database of *D. virilis* sequences have not revealed other subfamilies of *TART*\textsuperscript{vir}.

The extensive sequence divergence between *D. melanogaster* and *D. virilis* is at odds with the important function of these elements in Drosophila species. In the studies reported here we investigate how this extensive divergence has affected function by studying the intracellular targeting of *HeT-A* and *TART* Gags in *D. virilis* and *D. melanogaster* cells.
Results

The *D. virilis* WR Dv-1 cell line has *HeT-A* and *TART* elements similar to those in *D. virilis* flies.

Our early studies on the telomere retrotransposons in *D. virilis* were on flies (2, 3). For the transfection experiments reported here we used the cultured *D. virilis* cell line, WR Dv-1. All of the *D. melanogaster* cell lines that we have studied have multiple *HeT-A* and *TART* elements. Genomic hybridization shows that WR Dv-1 cells also have *HeT-A* and *TART* elements. As expected from the low level of sequence homology between these elements in the two species of flies, neither *HeT-A* nor *TART* give detectable cross-hybridization with DNA of cells from the other species (Supplemental Fig. 1).

*HeT-A*\(^\text{vir}\)* Gag is targeted to telomeres in both *D. virilis* and *D. melanogaster* cells.

One of the remarkable properties of the *HeT-A*\(^\text{mel}\)* element is the specificity with which its Gag protein localizes to telomeres in interphase cells (Fig. 2A). This localization has not been seen for Gag proteins from non-telomeric retrotransposons (8) and is likely to reflect *HeT-A*’s role in telomere maintenance. To see whether this property has been conserved in *HeT-A*\(^\text{vir}\)* we transiently transfected *D. virilis* cells with a construct expressing *HeT-A*\(^\text{vir}\)* Gag protein fused to green fluorescent protein (GFP). As in our experiments with *D. melanogaster*, the Gag protein efficiently entered the nucleus and formed many small dots (Het dots), strongly suggesting that telomere targeting of this protein has been conserved in *D. virilis* (Fig. 2B).

*D. melanogaster* studies have shown that the Het dots in that species are associated with telomeres because the dots colocalize with the telomere-associated protein, HOAP (9). We do not have useful markers for *D. virilis* telomeres; however, the similarity of the localization of *HeT-A*\(^\text{vir}\)* Gag with that of *HeT-A*\(^\text{mel}\)* Gag suggests that *HeT-A*\(^\text{vir}\)* Gag is also targeted to telomeres. The conclusion that Het dots are associated with telomeres is supported by analysis of the number of nuclear dots produced by the two proteins in both *D. virilis* and *D. melanogaster* cells. The analysis was done by confocal microscopy because the fluorescence microscope used in the other experiments captures only a fraction of the dots in each optical section. The multiple planes obtained with the confocal microscope revealed more dots, and showed that some of the telomeres were under others and therefore obscured in a non-confocal image. The number of nuclear dots should be related to the number of telomeres. Therefore, if
both proteins give similar numbers of dots in the same cell type, we can assume that both are at the telomeres.

*D. virilis* cells have more telomeres than *D. melanogaster* cells. *D. virilis* has 6 chromosomes (24 ends in a diploid cell before replication), while *D. melanogaster* has 4 chromosomes (16 ends in a diploid cell). As expected from non-confocal pictures, confocal pictures show that the two Gags form similar Het dots in nuclei of either cell line; however the number of dots depends on the species of the transfected cell, not the species of the *Het-A* Gag. In both cell types some dots are bigger than others; probably due to failure to resolve closely spaced or fused dots. Fused dots might be expected because the Gag proteins in these experiments are overexpressed and have a tendency to form aggregates. Nevertheless, we found that confocal images of both *Het-A* Gag proteins showed substantially more dots in *D. virilis* cells (Fig. 2B) than in *D. melanogaster* cells (Fig. 2A).

*Het-A* is telomere-specific in *D. virilis*, as it is in *D. melanogaster*, so Het-dot targeting of its Gag protein is not surprising. However we were surprised to find that *Het-A<sup>vir</sup>* Gag formed Het dots when it was expressed in *D. melanogaster* cells (Fig 2C) and that *Het-A<sup>mel</sup>* formed Het dots in *D. virilis* cells (Fig 2D). Thus, in spite of their very different amino acid sequences both proteins interacted appropriately with cellular targeting proteins in the other species. We therefore conclude that the two *Het-A* Gag proteins are targeted to telomeres in both homologous and heterologous cells.

*Het-A<sup>vir</sup>* does not form cytoplasmic Het-bodies in either *D. virilis* or *D. melanogaster*.

In addition to forming nuclear Het dots, *Het-A<sup>mel</sup>* Gag forms a characteristic cytoplasmic structure, a Het body, in some transfected cells. Het bodies are round or oval with smooth edges. Typically there is only one per cell and there is no association with a known cytoplasmic organelle to suggest function for the bodies. These bodies may indicate saturation of some part of the system for transporting Gag into the nucleus because Het bodies are seen only in cells that have formed nuclear dots and are larger in cells with more Gag protein. We have not seen similar structures from Gags of other elements, including *TART*. *Het-A<sup>mel</sup>* Gag forms Het bodies in both *D. melanogaster* (8) and *D. virilis* cells (Fig. 3A). Surprisingly *Het-A<sup>vir</sup>* Gag does not form Het bodies in either cell line. This is the most obvious difference in the behavior of the two proteins. The two proteins appear to express at similar levels in both cells and therefore it is possible that the formation of the Het bodies is a particularity of *Het<sup>mel</sup>* Gag.
TART\textsuperscript{vir} Gag does not appear to enter nuclei in \textit{D. virilis} cells, although TART\textsuperscript{mel} Gag moves efficiently into nuclei in both \textit{D. melanogaster} and \textit{D. virilis} cells.

In contrast to \textit{HeT-A}\textsuperscript{vir} Gag, TART\textsuperscript{vir} Gag does not behave like its \textit{D. melanogaster} homolog. In studies of \textit{D. melanogaster}, TART\textsuperscript{mel} Gag moved efficiently into the nucleus but formed loose clusters that did not preferentially associate with telomeres unless coexpressed with \textit{HeT-A}\textsuperscript{mel} Gag (9). Surprisingly, the localization of TART\textsuperscript{vir} Gag is very different. TART\textsuperscript{vir} Gag made characteristic clusters that spread throughout the cytoplasm but did not seem to enter the nucleus (Fig. 3B). The difference between the two Gags is determined by the proteins themselves, not the cells in which they are expressed. TART\textsuperscript{vir} Gag clusters were cytoplasmic in both \textit{D. virilis} (Fig. 3B) and \textit{D. melanogaster} cells, while TART\textsuperscript{mel} Gag clusters were nuclear in both cell types (not shown).

Organization of functional regions within the Gag proteins resembles that of their homologs in the other species.

Deletion derivatives of the \textit{D. melanogaster} Gag proteins have been used to identify regions of the protein capable of activities such as nuclear localization or clustering of the protein. These studies showed that \textit{HeT-A}\textsuperscript{mel} Gag and TART\textsuperscript{mel} Gag had a similar organization of these regions (6). To compare the internal organization of \textit{HeT-A}\textsuperscript{vir} and TART\textsuperscript{vir} Gag proteins with that of their \textit{D. melanogaster} homologs, we constructed and expressed deletion derivatives (Fig 4. see 4C and F for maps). Intracellular localizations of these deletion proteins showed striking similarities to those of analogous regions of the \textit{D. melanogaster} homologs. \textit{D. virilis} deletion proteins had the same localizations in both \textit{D. virilis} and \textit{D. melanogaster} cells, showing their ability to interact with cells of either species.

\textit{HeT-A}\textsuperscript{vir} Gag contains nuclear localization signals in both N-terminal and C-terminal ends as does the orthologous \textit{HeT-A}\textsuperscript{mel} protein, HM(1-534). However there were some differences in localization. N-terminal HV(1-462) was largely concentrated just inside the nuclear membrane (Fig. 4A), while HM(1-534) was more evenly distributed in the nucleus. C-terminal HV(463-907), efficiently entered the nucleus (Fig. 4B), where it formed many clusters. This localization resembled that of the analogous HM(589-921).

The N-terminal derivative of TART\textsuperscript{vir} Gag, TV(1-440), entered the nucleus, spreading without visible clusters (Fig. 4D), while the C-terminal fragment of TART\textsuperscript{vir} Gag, TV(441-1039), remained in the cytoplasm (Fig. 4E). This behavior was like that of the analogous TART\textsuperscript{mel} fragments. However the C-terminal end of TART\textsuperscript{vir} Gag has an activity not seen for TART\textsuperscript{mel} Gag; it determines the cytoplasmic localization of intact TART\textsuperscript{vir} Gag. A deletion derivative
lacking the very last ~200 amino acids, TV(1-703), entered the nucleus and spread evenly (Fig. 5 C and D). This C-terminal region may act as a nuclear export or a cytoplasmic retention signal, or alternatively affect the conformation of the full length protein to block the nuclear localization signal at the N-terminus.

The C-terminal region contains the largest cluster of poly-glutamine repeats in \textit{TART}^{\text{vir}} \text{Gag}. Poly-glutamine repeats are a distinguishing feature of both \textit{D. virilis} Gag proteins and have no counterparts in their \textit{D. melanogaster} homologs. These repeats are also found in other \textit{D. virilis} proteins. Their functions are not known but such repeats have been implicated in protein-protein interactions. It is interesting that removal of ~200 amino acids containing a significant repeat causes the protein to spread evenly in the nucleus rather than clustering in the cytoplasm. We have not determined whether poly-glutamine is responsible for either the clustering or the cytoplasmic localization but it is the most obvious sequence difference from the equivalent region of \textit{TART}^{\text{mel}} \text{Gag}.

\textit{TART}^{\text{mel}} \text{Gag} interacts with \textit{HeT-A}^{\text{mel}} \text{Gag} in both \textit{D. melanogaster} and \textit{D. virilis} cells, while \textit{TART}^{\text{vir}} \text{Gag} and \textit{HeT-A}^{\text{vir}} \text{Gag} interact only if \textit{TART}^{\text{vir}} \text{Gag} is forced to enter the nucleus.

Although \textit{TART}^{\text{mel}} \text{Gag} moves into the nucleus very efficiently it does not localize to telomeres unless coexpressed with \textit{HeT-A}^{\text{mel}} \text{Gag}. In \textit{D. melanogaster}, \textit{HeT-A}^{\text{mel}} \text{Gag} moved \textit{TART}^{\text{mel}} \text{Gag} into telomeric Het dots (9), showing that the two proteins interact. This interaction and telomere localization also occurred in \textit{D. virilis} cells (Fig. 5A), suggesting that either the two proteins interact without intervention of other cellular components or that necessary cellular components have been functionally conserved for the 60 MY that separate the two Drosophila species.

The interaction between the two \textit{D. melanogaster} Gags suggested that the two \textit{D. virilis} Gags might interact and that this interaction might be necessary to move \textit{TART}^{\text{vir}} \text{Gag} into the nucleus, as well as localizing it to telomeres. However, this does not seem to be the case. When \textit{TART}^{\text{vir}} \text{Gag} was co-transfected with \textit{HeT-A}^{\text{vir}} \text{Gag}, each of the proteins had the same localization seen in single transfections, showing no evidence that the proteins interact (Fig. 5B). There was also no interaction between \textit{HeT-A}^{\text{vir}} \text{Gag} and \textit{TART}^{\text{vir}} \text{Gag} when coexpressed in \textit{D. melanogaster} cells.

Comparisons of the Gag proteins show that the most conserved amino acid sequences of both \textit{HeT-A} Gag and \textit{TART} Gag are in the region resembling the MHR (major homology region)-zinc knuckle region of retroviral Gag proteins (Supplemental Figs. 2 and 3). This region
has been shown to be necessary for interaction between \textit{HeT-A}^{mel} Gag and \textit{TART}^{mel} Gag (6). The conservation of this region suggested that the two \textit{D. virilis} Gag proteins might interact if \textit{TART}^{vir} Gag could be moved into the nucleus. To accomplish this, we coexpressed \textit{HeT-A}^{vir} Gag with a deletion derivative of \textit{TART}^{vir} Gag, TV(1-703), containing the putative interaction region but lacking the C-terminal region that prevents entry into the nucleus.

TV(1-703) entered the nucleus but showed no tendency to form dots, although it has the MHR-zinc knuckle region and the equivalent TM(1-793) does form clusters. TV(1-703) localization was not affected by coexpression with either \textit{HeT-A}^{vir} Gag or \textit{HeT-A}^{mel} Gag (Figs. 5C and D). However there was evidence that TV(1-703) interacted with \textit{HeT-A}^{vir} Gag because in the coexpression \textit{HeT-A}^{vir} Gag was inhibited from forming nuclear dots. This interaction between the two \textit{D. virilis} proteins was seen in both \textit{D. virilis} and \textit{D. melanogaster} cells.

Importantly, TV(1-703) did not interact with \textit{HeT-A}^{mel} Gag (Fig. 5D). \textit{HeT-A}^{mel} Gag formed characteristic Het dots while the \textit{TART}^{vir} deletion remained dispersed in the nucleus. In cells with cytoplasmic Het bodies there was no detectable \textit{TART}^{vir} associated with this structure. This result argues that although the putative interaction region does show amino acid sequence conservation, changes have occurred that allow \textit{HeT-A} and \textit{TART} Gags to interact within a species but not between species.

\section*{Discussion}

\textbf{The telomeric targeting of \textit{HeT-A}^{vir} Gag is conserved in spite of a very high level of amino acid sequence difference.}

\textit{HeT-A} and \textit{TART} are remarkable because their coding sequences show a close relationship to members of the jockey clade of non-LTR retrotransposons in \textit{Drosophila} (4, 10), yet their transposition targets are the opposite of those of other members of the clade. \textit{HeT-A} and \textit{TART} never transpose into the euchromatic chromosome arms yet other elements are found at many sites in these arms. Conversely, other elements never transpose into the long \textit{HeT-A/TART} arrays which make up the telomeres (11). The basis for telomere-specific transposition is not known; however, the unique telomere targeting of the \textit{HeT-A} Gag protein in \textit{D. melanogaster} seems to offer at least a partial explanation. The work reported here demonstrates that this telomere targeting of \textit{HeT-A} Gag is conserved in the distantly related \textit{D. virilis}, despite significant differences in amino acid sequence.

In \textit{HeT-A}^{mel} Gag the determinants of telomere targeting are not confined to a specific region of the protein but are distributed along the protein (6). Therefore conservation of telomere targeting in \textit{HeT-A}^{vir} Gag must respond to strong selective pressure over each of these multiple
determinants. It might be supposed that the divergence in the amino acid sequence is driven by need to coevolve with cellular components acting in the several steps by which Gag moves from the ribosome to the telomeric Het dot. However, our results argue that this is not a major factor in the sequence changes; HeT-A Gag from either species shows appropriate targeting in cells of both species. In this case, conservation of protein structure may be more important than conservation of amino acid sequence.

Studies on *D. melanogaster* show that homomultimerization of HeT-A Gag depends on the MHR-zinc knuckle region of the protein to form Het dots. However, these studies give no hint to which interactions are involved in telomere association (6). The only *D. melanogaster* protein known to remain predominantly telomere-associated in interphase nuclei is HOAP (9, 12). Het dots are telomere-associated only in interphase; colocalization of Het dots with clusters of HOAP provides strong evidence that the Het dots are at telomeres. However these studies provide two kinds of evidence that HeT-A Gag does not bind HOAP (7). First, although both Het dots and HOAP clusters vary in size, the size of a Het dot frequently does not correlate with the associated HOAP cluster. If the proteins bind, we would expect their amounts to vary together. Second, when cells were centrifuged onto slides to break nuclei and spread the DNA, Het dots and HOAP still associated closely, showing that the association can withstand the force that broke the nuclei. However this association does not appear to be due to direct contact between Gag and HOAP. The two proteins remain close together but can be slightly separated by the force of centrifugation, suggesting both contact something else. The HOAP gene is one of the most rapidly changing genes known (13) and no *D. virilis* equivalent has been reported. Thus we cannot currently test for an association of HeT-A\textsuperscript{vir} Gag with HOAP from *D. virilis*. However the observation that the HeT-A\textsuperscript{vir} protein forms Het dots that are indistinguishable in number and distribution from those formed by HeT-A\textsuperscript{mel} Gag in the same cell line strongly suggests that HeT-A\textsuperscript{vir} Gag localizes at the telomeres in both cell types.

**The mechanism of TART localization is less conserved.**

HeT-A and TART have been found together on telomeres in all Drosophila species and stocks that have been studied. These two non-LTR elements have closely related Gag proteins but are different in ways that suggest that they evolved from different ancestors. We have suggested TART became a telomere-specific element when it acquired a Gag protein that allowed it to be targeted to telomeres and that the most likely source of this protein is transfer from a telomere-specific HeT-A (14), although it could have been inherited from a common ancestor (15). We suppose that TART acquired the new Gag well before the separation of *D.*
melanogaster and D. virilis because both elements have been found at telomeres in all Drosophila studied. HeT-A and TART are interdependent. TART\textsuperscript{mel} Gag needs HeT-A\textsuperscript{mel} Gag for correct localization to telomeres in D. melanogaster cells, while HeT-A elements do not encode a reverse transcriptase and might therefore depend on TART for this activity.

As expected from their apparent relationship, the Gag proteins from both TART\textsuperscript{vir} and HeT-A\textsuperscript{vir} are about equally diverged from their D. melanogaster homologs (4). However, unlike HeT-A\textsuperscript{vir}, TART\textsuperscript{vir} has differences from its homolog that can be seen by inspecting its sequence organization (see Fig. 1).

The work presented here shows additional differences between TART\textsuperscript{vir} and TART\textsuperscript{mel}. TART\textsuperscript{mel} Gag moves efficiently into the nucleus, while TART\textsuperscript{vir} Gag remains in clusters through the cytoplasm. We believe that TART\textsuperscript{vir} Gag proteins in our experiment are fully functional. They were expressed from two full-length gag genes, one from each of the two phage clones isolated in our screen. It seems unlikely that both genes would be defective. Neither protein was seen to enter the nucleus. Nor was HeT-A\textsuperscript{vir} Gag able to transport TART\textsuperscript{vir} Gag into the nucleus. We note that in neither of these transfection experiments could we be certain that no TART\textsuperscript{vir} Gag entered the nucleus. However in equivalent experiments with D. melanogaster elements TART\textsuperscript{mel} Gag is efficiently moved into the nucleus, showing a significant difference from the localization of TART\textsuperscript{vir} Gag. This difference is inherent in the Gag proteins since the behavior is not changed by expression in cells of the other species.

The difference between the localization of TART\textsuperscript{vir} Gag and that of TART\textsuperscript{mel} Gag is surprising. Sequences of the two proteins have diverged significantly but the difference is about equal to the difference between HeT-A\textsuperscript{vir} Gag and HeT-A\textsuperscript{mel} Gag and the latter proteins have strikingly similar patterns of localization. Sequence comparisons show that TART\textsuperscript{vir} Gag is evolving under selection and coevolving with HeT-A\textsuperscript{vir} Gag along the MHR and zinc knuckle regions (4). This pattern of sequence change suggests that Gag is important to TART\textsuperscript{vir} and that it acts in partnership with HeT-A\textsuperscript{vir} Gag. If the partnership resembles the partnership in D. melanogaster, the interaction should be detected in the nucleus. However we were only able to observe significant amounts of TART\textsuperscript{vir} Gag into the nucleus by deleting the C-terminal end of the protein, which had been shown to prohibit nuclear import or retention of TART\textsuperscript{vir} Gag. This deleted protein retained the most conserved region of the protein, the MHR-zinc knuckle region which is involved in the HeT-A\textsuperscript{mel}/TART\textsuperscript{mel} interactions. This shorter protein did interact with HeT-A\textsuperscript{vir} Gag but the interaction prevented HeT-A\textsuperscript{vir} Gag from localizing to telomeres. The mislocalization may have been the result of defective heteromultimers caused by the deleted TART\textsuperscript{vir} Gag. Nevertheless the experiment demonstrates an interaction between the two
proteins. Importantly, this interaction is species-specific; the deleted protein did not affect localization of \textit{HeT-A}^{mel} Gag when the two were coexpressed. This specificity is reminiscent of the selectivity of retroviral Gags in forming heteromultimers (16, 17). We note that the \textit{TART}^{vir} Gag derivative used in these experiments contained the MHR-zinc knuckle region important for multimerization of retroviral Gag proteins (18, 19, 20), and also the interaction of \textit{HeT-A}^{mel} Gag with \textit{TART}^{mel} Gag (6).

One possible explanation for the unexpected \textit{TART}^{vir} Gag localization might be that localization was studied by overexpressing a tagged protein. Overexpressed \textit{TART}^{mel} Gag moves into the nucleus. Perhaps at some point in its journey to the nucleus \textit{TART}^{vir} Gag requires a cell component that is present in limiting amounts and that component is overwhelmed by the overexpression of Gag. If this explanation is correct; \textit{TART}^{vir} Gag localization may be accomplished by only a minor modification of that used by \textit{TART}^{mel} Gag. If a complete \textit{TART}^{vir} Gag could enter the nucleus, it might interact appropriately with \textit{HeT-A}^{vir} Gag. (This discussion is based on the hypothesis that \textit{TART} Gags must enter the nucleus to enable transposition of this element. If \textit{TART}^{vir} can transpose without a nuclear Gag, the cytoplasmic localization seen in these experiments may be appropriate.) The results reported here allow us to conclude only that the mechanism of \textit{TART} Gag localization has diverged more than that of \textit{HeT-A} Gag.

We note that \textit{TART}^{vir} may not be the only source of enzyme activity for transposition of \textit{HeT-A}^{vir}. The \textit{U}^{vir} element contains a \textit{pol} gene flanked by 5’ and 3’ UTR sequences highly similar to \textit{HeT-A}^{vir}. This association suggests that \textit{U}^{vir} Pol could also supply enzyme activity for \textit{HeT-A}^{vir}. We have suggested that \textit{U}^{vir} may have been fortuitously formed from the mRNA of a cellular gene (3), and not a bona fide element. Currently available \textit{D. virilis} sequence is still consistent with this possibility. The putative cellular gene might have been the original source of enzyme activity for \textit{HeT-A} and, in fact, may still provide this activity in \textit{D. virilis}.

\textbf{In spite of significant amino acid sequence divergence among the telomere retrotransposon Gags, the organization of functional domains tends to be conserved}

Comparison of the amino acid sequences for \textit{HeT-A} and \textit{TART} Gags shows the highest similarity in the regions of two motifs, the MHR and the zinc knuckles, that characterize retroviral Gag and are found in the jockey clade of non-LTR retrotransposons to which \textit{HeT-A} and \textit{TART} belong (6). The zinc knuckles are located roughly one third of the length of the protein from the C-terminal end and the MHR is slightly N-terminal of this. Sequences before and after these regions are more diverse and lack motifs that would give clues to function. However study of
deletion derivatives of HeT-A<sup>mel</sup> Gag and TART<sup>mel</sup> Gag identified functions associated with different domains of the sequence and showed that the linear order of these domains in the two proteins was similar.

Analogous deletion derivatives of the D. virilis Gag proteins reveal similarity of organization, both between the two Gags within a species and between homologs in the two species. N-terminal fragments of all four Gags localize to the nuclei, as do C-terminal fragments of the HeT-A Gags. C-terminal fragments of TART Gags remain in the cytoplasm. N-terminal derivatives of all four Gag proteins tend to spread evenly in the nucleus while C-terminal derivatives have tendency to form clusters; HeT-A Gags form clusters in the nucleus and TART Gags form clusters in the cytoplasm. This organization indicates that sequence divergence is occurring without much rearrangement of regions of peptide sequence, suggesting that this organization is important for proper function of the telomere Gags. In fact we believe that the conservation of several small motifs along the proteins might be responsible for the grouping of the Gag proteins of all HeT-A and TART homologs onto the same subclade (bootstrap value 82, see supplemental Fig. 4) when compared with other Gag proteins from the Jockey clade (10).

**Concluding Remarks**

Comparisons of the telomeres in different species of Drosophila show that, although the mechanism of telomere maintenance is largely conserved, both nucleotide sequence and amino acid sequences change rapidly. As mentioned earlier, protein structure may be more important than amino acid sequence in this case. The low percentage of identity among the different homolog sequences was unexpected for genes that are directly related to an essential cellular role as complex as telomere maintenance. This presents an interesting parallel to the rapidly evolving sequences of centromeres (21, 22). Perhaps rapid evolution is a clue to the mechanism by which telomeres are maintained. If the telomere Gag proteins need to interact with rapidly evolving proteins, such as HOAP, in order to fulfill their telomere function, the importance of the telomere would explain the strong conservation of function while the sequence change would be driven by the need to interact with other rapidly changing proteins.

**Materials and Methods**

**Cell lines:** The *D. melanogaster* SL2 cell line and the *D. virilis* cell line WR Dv-1 were established by I. Schneider (23). Cells were maintained in Schneider media (Sigma for SL2 and Invitrogen for WR Dv-1) supplemented with 10% fetal bovine serum (Gibco) at 25°. WR Dv-1 cells were from the Drosophila Genomics Resource Center.
**Southern hybridizations:** Genomic DNA was extracted from SL2 and WR Dv-1 cells. Aliquots of DNA were digested with Sal I + EcoRI and Hind III. Southern blot hybridization was under low stringency conditions (24), with sequence of the Gag genes as probes. Probes were labeled with $^{32}$P dCTP, Amersham Ready to Go Labeling Beads.

**Recombinant DNA:** Fusion protein constructs were made by PCR amplification of Gag sequence from phage clones V8, V2 and V3 (see 2 and 3) with specific primers. 5' primers had a methionine codon added if needed. 3' primers had the stop codon deleted and minimum changes introduced in the sequence to obtain the desired fusion protein. Constructs were cloned in the pPL17 vector that encodes the desired fluorescent protein, GFP, CFP or YFP under the armadillo promoter (see 8).

**Transient transfections:** Cells were seeded at 2x10$^6$ cells/ml twelve hr prior to transfection and transfected using the calcium phosphate method (25) with 5 to 10 μg of DNA. For WR DV-1 cells, a 2.5 min DMSO shock (10% DMSO in cell culture media) was performed 5-7 hr after transfection. Slides were made 48 and 72 hr after transfection.

**Slide preparation and microscopy:** Cell suspensions were placed on concavalin treated coverslips, fixed 10 min with 4% paraformaldehyde in PBT, washed with PBT and with PBS, and mounted with mowiol/ DAPI medium. Fluorescent fusion proteins were analyzed in a Nikon Eclipse E1000 fluorescent microscope and a Leica confocal microscope TCS-SP2-AOBS. Black and white micrographs were colored with Metamorph 6.3r1 software (Molecular Devices) and adjusted with Adobe Photoshop CS version 8.0.

**Lamin immunoblotting:** Transfected cells, fixed as above, were incubated overnight with LaminDmO antibody (DSHB, ADL67-10). Cy3 anti-rabbit antibody was from Jackson ImmunoResearch.

**Acknowledgments**

We thank Lidia Bardia Sória for great help with microscope analyses. We thank the Pardue lab and Ky Lowenhaupt for critical reading of the manuscript. This work was supported by 31090 International Reintegration Grant, Marie Curie Action and by BFU2006-13934/BMC of the Spanish Ministry of Education and Sciences to Elena Casacuberta, by BC2003-00243 of the Spanish Ministry of Education and Sciences to Fernando Azorín and by GM50315 from the National Institutes of Health to Mary-Lou Pardue.
References
Figure Legends

Figure 1. Comparison of telomere retrotransposons in *D. melanogaster* and *D. virilis*, showing conserved features and nucleotide identities of specific regions. Diagrams of (A) *HeT*-A and (B) *TART* homologs. Drawn approximately to scale. Solid bars on right indicate phylogenetic relationships. MY, million years: %, percentage of nucleotide identity of sequence between arrowheads, for *HeT*-A both the Gag coding region and the entire element are compared, for *TART*, only the Gag and Pol regions are similar enough to calculate % identity: dashed lines under *HeT*-A indicate irregular A-rich repeats; striped lines under *TART*^mel^ indicate position of Perfect NonTerminal Repeats (PNTR); dotted lines under *TART*^vir^ genes indicate concentrations of poly-glutamine repeats. X, extra domain of *pol* coding region. (*D. melanogaster* has multiple subfamilies of *TART*. *TART*^mel^ represents a consensus of these subfamilies). Note that the non-coding regions of *HeT*-A are more highly conserved than the coding regions.

Figure 2. Confocal images of cells transiently transfected with *HeT*-A Gag-GFP constructs. In each experiment transfected cells have Gag-GFP localized to dots in the nucleus. The number of *HeT*-A dots is smaller in SL2 cells (A and C) than in WR Dv-1 cells (B and D), whether the cell is expressing *HeT*-A^mel^ Gag or *HeT*-A^vir^ Gag. DNA in all cells is stained with DAPI (blue). Left panel: summation of all planes taken with the GFP filter and merged with the DAPI image. Right panel: transmission image merged with DAPI.

Figure 3: WR Dv-1 cells expressing *HeT*-A^mel^ Gag-GFP or *TART*^vir^ Gag-GFP. (A) Daughters of a cell expressing *HeT*-A^mel^ Gag-GFP show nuclear dots and also very large cytoplasmic Het bodies. It appears that the Het bodies did not segregate at cell division but both remained in the lower cell. These are examples of very large Het bodies and they are overexposed to reveal the nuclear dots. (B) Cells expressing *TART*^vir^-GFP have no detectable nuclear label. Left panel; GFP+DAPI. Right panel: DIC+DAPI.

Figure 4. WR Dv-1 cells expressing deletion derivatives of *HeT*-A^vir^ Gag-GFP or *TART*^vir^ Gag-GFP: Deletion derivatives show that both ends of *HeT*-A^vir^ Gag enter the nucleus, although their distribution within the nucleus differs (A and B). Only the N-terminal end of *TART*^vir^ Gag enters the nucleus (C). The C-terminal end forms cytoplasmic clusters (D).
*HeT-*A\textsuperscript{vir} Gag, HV(1-462). GFP; lamin (red); merged GFP, lamin and DAPI; merged DIC and DAPI. B) HV(463-907), GFP; merged DIC and DAPI. C) Diagrams of *HeT-*A\textsuperscript{mel} and *HeT-*A\textsuperscript{vir} Gag proteins showing MHR (single gray box), zinc knuckles (three white boxes), and glutamine-rich region (black dots below line). Gray bars below each protein show sequence in each deletion derivative with first and last amino acids. D) *TART*\textsuperscript{vir} Gag, TV(1-440) GFP; lamin (red); merged GFP, lamin (red) and DAPI; merged DIC and DAPI. E) TV(441-1039) GFP; merged DIC and DAPI. Note that lamin staining in A and C is exterior to the GFP signal indicating that the Gag-GFP fusion is in the nucleus. F) Diagrams of *TART*\textsuperscript{mel} and *TART*\textsuperscript{vir} Gag proteins with deletion derivatives, marked as in C. Transient transfection of the construct TV(1-703) is shown in Fig. 5, panels C and D.

**Figure 5.** WR Dv-1 cells expressing co-transfected Gag-CFP and Gag-YFP proteins. Co-localization shows that *HeT-*A\textsuperscript{mel} Gag interacts with *TART*\textsuperscript{mel} Gag to move *TART*\textsuperscript{mel} Gag into Het dots (A), but *HeT-*A\textsuperscript{vir} Gag does not affect localization of *TART*\textsuperscript{vir} Gag (B). The deletion protein TV(1-703) interacts with *HeT-*A\textsuperscript{vir} Gag to interfere with its formation of Het-dots (C) but TV(1-703) does not affect the localization of *HeT-*A\textsuperscript{mel} Gag (D).
Fig. 1

HeT-A<sub>mel</sub> 5' Gag 3' AAA

HeT-A<sub>vir</sub> 5' Gag 3' AAA

TART<sup>mel</sup> 5' Gag Pol 3' AAA

TART<sup>vir</sup> 5' Gag Pol X 3' AAA

30% 50% 26% 45% 60MY
Fig. 2
Fig. 3

HeT-A<sup>mel</sup>  WRDv-1

A

B

TART<sup>vir</sup>  WRDv-1
Fig. 4
Fig. 5
Legends for Supplemental Figures

Figure 1. HeT-A and TART elements show little sequence conservation between D. melanogaster and D. virilis. DNA from D. melanogaster SL2 cells and D. virilis W Dv-1 cells (labeled Mel and Vir above lanes) was digested with Sal I plus EcoRI (left lane in each pair) and Hind III (right lane in each pair) and probed with $^{32}$P-labeled sequence of (A) HeT-A$^{mel}$ Gag, (B) HeT-A$^{vir}$ Gag, (C) TART$^{mel}$ Gag, and (D) TART$^{vir}$ Gag. Each probe labels multiple bands in its homologous DNA but, although hybridization was done at low stringency (8), there is little, if any detectable hybridization to DNA of the other species.

Figure 2. Multiple alignment amino acid sequences of HeT-A$^{vir}$ Gag (V3A and V7A) with sequence of six HeT-A$^{mel}$ Gag proteins (#4491-36518). HeT-A$^{vir}$ sequences from Genbank (V3A=AY369259, V7A=AY369260. HeT-A$^{mel}$ sequences from (13). Red bar underlines MHR region. Yellow bars underline zinc knuckles.

Figure 3. Multiple alignment amino acid sequences of TART$^{vir}$ Gag with sequences of TART$^{mel}$ Gag proteins from TART subfamilies A, B, and C). TART$^{vir}$ sequence (AY219708). TART$^{mel}$ subfamily A (AY561850), subfamily B (U14101), subfamily C (AY600955). Red bar underlines MHR region. Yellow bars underline zinc knuckles.

Figure 4. Phylogenetic relationships of Gag sequences from the Jockey clade. Neighbor-joining tree is shown. (UPGMA tree gave the same result). Bootstrap values (at corresponding nodes) calculated with 500 replications and a cutoff value of 50%. Scale bar indicates number of differences per residue. jockey$^{mel}$ Ac: M22874, jockey$^{fun}$ Ac: PIR B38418, Doc Ac: CAA35587, X Ac: AF 237761, TART$^{mel}$, Ac: U14101, TART$^{mel}$: F. Sheen and R. Levis, TART$^{Cmel}$: L. Tolar, J. Stolk, and R. Levis, TART$^{vir}$ Ac: AAO67564, TART$^{ame}$ Ac: AAO67565.
<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
<th>130</th>
</tr>
</thead>
<tbody>
<tr>
<td>2V3</td>
<td>16768</td>
<td>2V2</td>
<td>16768</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4491</td>
<td>16768</td>
<td>2V2</td>
<td>16768</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4491</td>
<td>16768</td>
<td>2V2</td>
<td>16768</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16768</td>
<td>16768</td>
<td>2V2</td>
<td>16768</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16768</td>
<td>16768</td>
<td>2V2</td>
<td>16768</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16768</td>
<td>16768</td>
<td>2V2</td>
<td>16768</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16768</td>
<td>16768</td>
<td>2V2</td>
<td>16768</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16768</td>
<td>16768</td>
<td>2V2</td>
<td>16768</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16768</td>
<td>16768</td>
<td>2V2</td>
<td>16768</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16768</td>
<td>16768</td>
<td>2V2</td>
<td>16768</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Supplemental Fig 2