Supporting Information

George et al. 10.1073/pnas.1015926107

SI Text

Promoter structures for antisense RNA for *Drosophila melanogaster* and *Drosophila virilis TART* are shown in Fig. S2 and Fig. S3, respectively. *TART* is the only retrotransposon in *D. melanogaster* with very abundant, nearly full-length, antisense transcripts (4 to 10 times more abundant than sense strand). New classes of RNAs are emerging rapidly, but none suggest a role for these *TART* transcripts. Conservation of the unusual transcripts in distantly related *D. virilis* (1) adds to their interest but does not suggest a function. The conservation is particularly striking because the architecture of the sense strand promoters in the two species has diverged dramatically. To investigate the possibility that the divergence of the sense strand promoters is correlated with divergence of the antisense promoters, we have mapped the antisense promoters with our reporter assay. We find that antisense promoters for the large RNAs are similar in the two species. Each has a small cluster of transcription start sites in the 3′ UTR several hundred nucleotides from the end of the element. In each species the most active antisense construct was approximately twice as active as the major sense strand construct. The difference in activity is much less than the difference in the abundance of endogenous sense and antisense RNA, suggesting that the endogenous RNA may be regulated by both transcription and turnover. Our 5′ RACE clones from *D. virilis TART* showed consistent splicing of antisense, but not sense strand, RNA in *D. virilis TART*. The significance of this splicing is unclear, but it is conserved in *D. melanogaster*: Maxwell et al. (2) reported finding splicing of antisense but not sense transcripts from all three *TART* subfamilies. The splice sites detected do not markedly improve protein coding.

1. Casacuberta E, Pardue ML (2003) Transposon telomeres are widely distributed in the *Drosophila* genus: TART elements in the *virilis* group. *Proc Natl Acad Sci USA* 100: 3363–3368.

2. Maxwell PH, Belote JM, Levis RW (2006) Identification of multiple transcription initiation, polyadenylation, and splice sites in the *Drosophila* melanogaster TART family of telomeric retrotransposons. *Nucleic Acids Res* 34:5498–5507.

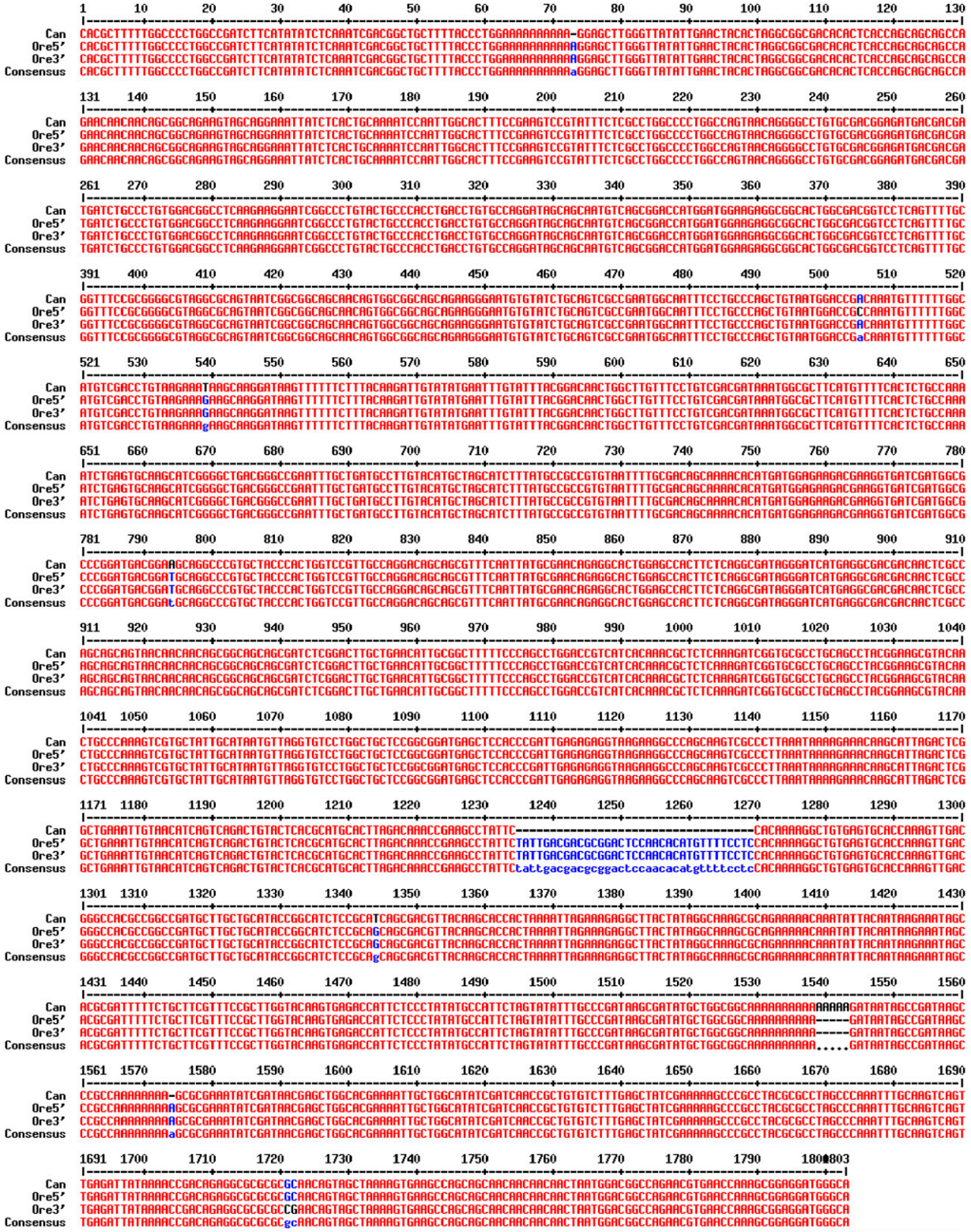

Fig. S1. Evidence that 5′ and 3′ PNTRs (perfect nonterminal repeats) are evolving together. A multiple alignment using MultAlin [Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16:10881–10890.] of the regions of overlap in the 5′ and 3′ PNTRs of the two longest available *TART* elements shows that 45 of the 48 nt differences between the two elements are present in both the 5′ and 3′ PNTR. Can: nt 2208–3974 of element AJ566116 and the identical sequence in its 3′ UTR. Ore5′: nt 1–1798 of element AY561850. Ore3′: nt 11179–12976 of element AY561850. Although the PNTRs of these two elements overlap by 1,850 nt, we have omitted the last few nt where all PNTRs are a little ragged. For simplicity we have compared only the first 1,798 nt of each of the Ore PNTRs to the equivalent sequence of Can because in this region the two Can PNTRs are completely identical.

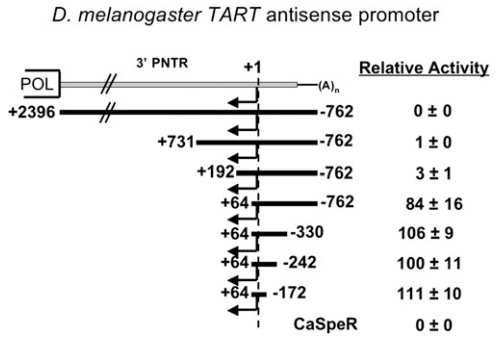

Fig. S2. Relative promoter activity of sequences from *D. melanogaster TART* elements tested in *D. melanogaster* cells. Diagram shows the 3′ end of *TART*, ending in the oligoA. Gray bar represents the 3′ PNTR (perfect nonterminal repeat), arbitrarily terminated within ORF 2 because the 5′ limit of the PNTR is undetermined. The thin black line at the 3′ end marks the post-PNTR, ending with the terminal $(A)_n$. Shown below are sequences tested, identified by nucleotide number at either end. Nucleotides are numbered from the transcription start site (arrow) identified by Maxwell et al. [Maxwell PH, Belote JM, Levis RW (2006) Identification of multiple transcription initiation, polyadenylation, and splice sites in the *Drosophila* melanogaster TART family of telomeric retrotransposons. *Nucleic Acids Res* 34:5498–5507.]. Positive numbers run downstream of the start, and negative numbers run upstream because this is antisense RNA. Activity (±SD) was determined as in Fig. 2 (main text). The smallest promoter construct tested was one of the most active. It had only 172 nt of sequence upstream of the start site and 64 nt downstream of the start site identified by Maxwell et al. Surprisingly, adding more sequence downstream of the start rapidly diminished the apparent activity of the promoter. We believe this apparent loss of activity is an artifact of our assay, which requires translation of the reporter protein. There is no indication that *TART* antisense RNA encodes a protein, and the sequence that seems to inhibit reporter protein activity may be doing so because its usual role is to sequester the antisense RNA in a nontranslatable form or location.

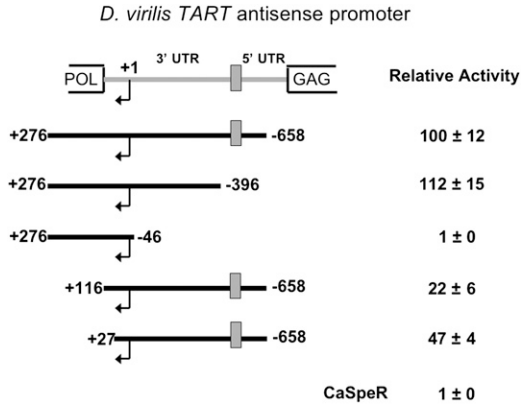

Fig. S3. Relative promoter activity of sequences from *D. virilis TART* elements tested in *D. virilis* cells. Diagram shows junction between two elements with the 3′ UTR of the upstream element and the 5′ UTR of the downstream element. Constructs are shown below, with starting and ending nucleotide indicated. Numbering is from the most used site in the cluster of three transcription starts (arrow) determined by 5′ RACE of RNA from larvae and cultured cells. Positive numbers run downstream of these starts, and negative numbers run upstream because this is antisense RNA. Gray box indicates string of eroded tags on the downstream element. Activity (±SD) was determined as in Fig. 2. Promoter activity requires sequence 5′ of the transcription start (in the antisense strand) and is enhanced as the construct is extended into the coding region: it is not increased by addition of sequence from the 5′ UTR of the downstream element.