LETTERS

Genomic Analysis of Drosophila melanogaster Telomeres: Full-length Copies of HeT-A and TART Elements at Telomeres
José P. Abad,*1 Beatriz de Pablos,*1,2 Kazutoyo Osoegawa,3 Pieter J. de Jong,3 Antonia Martín-Gallardo,2 and Alfredo Villasante1

1Centro de Biología Molecular Severo Ochoa (CBM), CSIC-UAM, Madrid, Spain; 2Servicio Interdepartamental de Investigación (SidI), UAM, Madrid, Spain; 3Children’s Hospital Oakland Research Institute (CHORI), Oakland, Calif.

The repetitive nature of heterochromatin hampers its analysis in general genome-sequencing projects. Specific studies are needed to extend the sequence into telomeric and centromeric heterochromatin. Drosophila telomeres lack the telomerase-generated repeats that are characteristic of other eukaryotic chromosomes. Instead, they consist of tandem arrays of HeT-A and TART elements. Herein, we present the genomic organization of the telomeres in the isogenic strain (y; cn bw sp) that was used for the Drosophila melanogaster sequencing project. The data indicate that the canonical features of telomere organization are widely conserved in evolution. In addition, we have identified full-length elements, likely competent elements, for HeT-A and TART.

Introduction

The heterochromatin in any eukaryotic species is formed by highly repetitive and middle repetitive sequences or both. Chromosomal elements crucial for chromosome behavior, such as centromeres and telomeres, are included in the heterochromatic regions of the chromosomes. Even though much research has been done to pursue the knowledge of the structure and function of such regions, our understanding of them is still quite poor. The Genome projects, which have reported to us the so-called “complete sequence” of several organisms, are not actually complete, since in most cases the heterochromatic regions have not been sequenced or assembled. This is the consequence of these regions being, in most cases, difficult to clone or to assemble owing to their repetitive nature.

The telomeres of Drosophila melanogaster do not have typical telomerase repeats. Instead, two families of non-LTR retrotransposons, HeT-A and TART, maintain telomere length by occasional transposition to the chromosome ends. It has been proposed that HeT-A and TART, which transpose only to chromosome ends, do this by target-primed reverse transcription (Biessmann et al. 1992b; Levis et al. 1993). The targeting does not seem to depend on the DNA sequence at the transposition site (Traverse and Pardue 1988; Biessmann et al. 1990; Biessmann et al. 1992a; Sheen and Levis 1994). TART has the two open reading frames (ORFs) typical of many non-LTR retrotransposons: ORF1 and ORF2 (also called gag and pol because of sequence similarities to retroviral gag and pol genes). The ORF2 has both endonuclease and reverse transcriptase (RT) domains. HeT-A is an atypical element because it has no ORF2, so the RT for its transposition is produced in trans from an unknown source. As a distinctive feature, HeT-A and TART have an unusually long 3’ untranslated region (UTR) that in the case of HeT-A contains imperfect repeats. The transcription of HeT-A seems to require a promoter located in the 3’ UTR of another upstream element (Danilevskaya et al. 1997). It follows that an HeT-A element can act as a competent copy only if it is adjacent to HeT-A 3’ end sequences. Although HeT-A elements are transcribed from the sense strand, TART elements yield both sense and antisense transcripts from unidentified promoters (Danilevskaya et al. 1999). HeT-A is several times more abundant than TART, and the two sequences appear to be randomly mixed in head-to-tail arrays. Sequences from these telomeric elements are also found in centric heterochromatin (Traverse and Pardue 1989; Danilevskaya et al. 1993; Levis et al. 1993; Losada, Abad, and Villasante 1997; Agudo et al. 1999; Losada et al. 1999), but never in euchromatin. Immediately proximal to the terminal retrotransponson array are several kilobases of a complex minisatellite called telomere-associated sequences (TAS) (Karpen and Spradling 1992; Levis et al. 1993; Walter et al. 1995). Sequence similarities between TAS of XL, 2R, and 3R have been shown by in situ hybridization (Karpen and Spradling 1992).

Telomeric copies of HeT-A and TART may not be capable of transposition, as the chromosome tip is progressively truncated by incomplete terminal replication. It has been speculated that the nontelomeric copies of HeT-A and TART in centric heterochromatin could behave as the templates for transposition. While studying the centromeric region of the Y chromosome, we have detected putative ancestral telomeres in internal positions. These internal regions consist of arrays of HeT-A and TART elements (including coding regions) that have decayed by mutations, deletions, transposon invasion, and regional amplification (Losada, Abad, and Villasante 1997; Agudo et al. 1999). It seems clear that these internal regions of the Y chromosome are unlikely to retain complete retrotranspositional elements.

Materials and Methods

Strain and Telomere Mapping

In this work, three genomic DNA BAC libraries from the D. melanogaster isogenic strain yellow (y1); cinnabar (cn1) brown (bw1) speck (sp1) were used: one prepared from EcoRI digested genomic DNA, library RPCI-98 (http://bacpac.chori.org/dromel98.htm) with an average

Key words: telomeres, telomeric retrotransposons, Drosophila.

E-mail: avillasante@cbm.uam.es.

Molecular Biology and Evolution vol. 21 no. 9 © Society for Molecular Biology and Evolution 2004; all rights reserved.

* These authors contributed equally to this work.
insert size of 165 kb; and two probes from shearedin genomic DNA, libraries CHORI221 (http://bacpac.chori.org/drososmall221.htm) and CHORI223 (http://bacpac.chori.org/drososmall223.htm) with average insert sizes of 115 kb and 48 kb, respectively. The libraries were screened by using colony hybridization with DNA fragments from telomeric retrotransposons and probes isolated by use of polymerase chain reaction (PCR) from the distal ends of BACR45O17 (tel 3L) (primers pair 5'-CGATTAGTT-GACACTATAG-3' and 5'-AACAATAGAGAAGC-GAG-3'; nucleotides 1 to 701 from GenBank entry AC105293), and of BACR222J20 (tel 4R) (primers pair 5'-TATAACGCTACTATAGG-3' and 5'-TCTATAATACTCCTCCCTCG-3'; nucleotides 151992 to 152629 from GenBank entry AC010577). The HeT-A probes were contributed by M-L. Pardue: probe 2b is a 1.3-kb fragment from the 3' UTR of another HeT-A element (U06974), and probe ORF 1+II is a 2.6-kb fragment containing the coding region of another HeT-A element (U06920). The TART probes were contributed by R. Levis: a 2.2-kb ScaI fragment from the ORF2 of a TART A element (U02279) and a 3.7-kb ScaI fragment from the 3' UTR of another TART A element (Oregon R). Membranes containing the arrays of clones from each library were provided by the Children’s Hospital Oakland Research Institute BACPAC resources. Contigs were constructed by restriction enzyme fingerprinting and Southern hybridization. The restriction enzymes used were BamHI, EcoRI, HindIII, KpnI, PstI, and NotI. For restriction enzyme analysis of BACs both conventional 0.8% agarose gel electrophoresis and pulse-field gel electrophoresis (PFGE) were used. PFGE was performed in a CHEF-DR II apparatus (Biorad, Hercules, CA) on 1% agarose gels in 0.5X TBE (1X TBE is 90 mM Tris-borate/2 mM EDTA) for 20 h using 150 V and a 14-s pulse time. Probes were 32P-labeled by random-priming, and colony hybridization and Southern blot analysis were performed overnight at 68°C by the Church and Gilbert method (Church and Gilbert 1984). TAS length was measured by conventional gel electrophoresis after digestion of BACs with a restriction enzyme that does not cut within the repeat but cuts frequently within the insert: AluI was used for 2L and 3L TAS, and HaeIII was used for 2R and 3R TAS. DNA Sequencing and Sequence Analysis Most sequencing was done on BAC DNA with the appropriate primers. In a few cases, owing to the redundancy of sequences in some BACs, a particular fragment was subcloned or amplified by PCR and sequenced using custom designed primers. Primers used for PCR amplification and sequencing are shown in supplementary table 1. All sequencing was performed using big dye-termination reagents and ABI/PE 377 automated sequencers. Sequences were analyzed by Blast searches. Alignments were performed with ClustalX (Thompson et al. 1997), followed by minor manual adjustments of gaps. Results and Discussion Until this study, the distal-most regions of the telomeres of the y; cn bw sp strain had not been cloned. While working on the HeT-A- and TART-rich regions of the Y chromosome we found that the conventional YAC and BAC libraries are underrepresented for telomeric sequences. In other eukaryotic organisms, telomeres have been cloned by functional complementation in yeast of a half-YAC telomere cloning vector (Riethman et al. 1989). This strategy is not feasible in Drosophila because its atypical telomeres would not function as a yeast telomerase substrate. To overcome this problem we used two Drosophila BAC libraries (CHORI221 and CHORI223) generated from randomly sheared DNA because we have previously found that the heterochromatin sequences are better represented in this new type of libraries (Osoegawa et al., unpublished data) and, in theory, these libraries should permit the cloning of double-stranded terminal chromosomal fragments. In practice, we were able to isolate sequences distal to those previously characterised by the Drosophila Genome Project (Celniker et al. 2002; Release 3.1 BDGP 2003). We have constructed BAC-based physical maps by restriction enzyme fingerprinting and hybridization analysis. The BAC end-sequences of all isolated clones and the partial sequencing of particular ones have allowed us to determine the structural organization of the telomeres of the y; cn bw sp isogenic strain of D. melanogaster (fig. 1). In those telomeres from which multiple BAC clones have been isolated, the distal sequences of the most distal clones cluster within a narrow terminal region in agreement with the narrow length heterogeneity at the telomere ends (fig. 1, panels tel XL, tel 2L, tel 2R, and tel 3L). This suggests that the chromosome ends have been cloned. The telomeres show a large variation in length, the biggest being the XL telomere (147 kb), followed by 2R (90 kb), 2L (47 kb), 4R (43 kb), 3R (28 kb), and 3L (26 kb). As expected, the HeT-A elements are more abundant than TART and there are no TART elements in some telomeres. Interestingly, we did not find telomeric retrotransposons at the 3L telomere in the y; cn bw sp strain (fig. 1, panel tel 3L). This lack of telomeric elements at one specific telomere may be more common than expected because terminal chromosome deficiencies occur in wild populations (Walter et al. 1995). The TAS are 26 kb in length in 3L, 19 kb in 2L, 17 kb in 2R, and 16 kb in 3R, but they are absent from the XL and 4R telomeres. Lack of TAS in the XL telomere is a specific feature of the y; cn bw sp strain, since these sequences have been found in the X chromosome of Oregon R (Benos et al. 2001) and in the Dp1187 minichromosome (Karpen and Spradling 1992) (fig. 1, see bottom of panel tel XL). The absence of TAS at different telomeres may arise after transposition of telomeric retroelements to chromosome ends that have gradually lost the HeT-A–TART array and TAS, or to ends that have experienced a terminal deletion within the TAS followed by further recession into the subterminal euchromatic region. The polymorphic distribution of TAS shares features with the organization of subtelomeric repeats in yeast and human genomes (Horowitz, Thorburn, and Haber 1984; Brown et al. 1990). The analysis of the subtelomeric minisatellites in the y; cn bw sp strain (2R TAS AJ549606, 15; 3R TAS AJ564677; 2L TAS AJ564667, 8; 3L TAS AJ549618, 19, 22, 24, 26, 29, 31, 35, 38, and 40) shows
fig. 1.—Structure of the telomeres of the D. melanogaster y; cn bw sp isogenic strain. Numbers next to clones indicate the number of nucleotides from the distal boundary of the most distal clone. Vertical lines indicate EcoRI sites except for those labeled N, which indicate NotI sites. Telomeric retrotransposons labeled with * have been detected as full-length elements. There is evidence of insertional polymorphism in the 2R telomere (inverted triangles in panel tel 2R) observed by the different size (about 6 kb) of the corresponding restriction fragments in the BAC contigs. In the 4R telomere, there are insertions of Cr1a and Hoppel (green and brown boxes, respectively), and there is a pseudogene in the XL telomere (grey thin bar in the drawing at the bottom of panel tel XL) corresponding to 242 bp of a cDNA from l(3)S112803 gene. The red and yellow boxes correspond to HeT-A and TART A elements, respectively. The blue boxes correspond to TAHRE, a novel telomeric retrotransposon (Abad et al. unpublished data).

A significant sequence similarity (98% identity) between the 2R and 3R TAS and shows that the 3L TAS is a 459-bp repeat remarkably similar (99.5% identity) to the 2L TAS repeat (Walter et al. 1995). The sequence similarity between telomere-associated DNA is thought to reflect exchange of DNA sequences between nonhomologous telomeres that lie nearby as a result of the Rabl configuration (polarized organization) of the chromo-
somes in the nucleus (Horowitz, Thorburn, and Haber 1984).

Previous studies have shown the presence of two Het-A subfamilies in the same genome (Biessmann et al. 1994; Danilevskaya, Lowenhaupt, and Pardue 1998). In the y; cn bw sp strain we found four Het-A subfamilies that we have called A, B, C, and D. The previously described Het-A elements 9D4 (Biessmann et al. 1992a), RT394 (Biessmann et al. 1992b) and 23Zn1 (Danilevskaya et al. 1994) belong to Het-A A subfamily, and the Het-A elements RT473 (Biessmann et al. 1992b), 17B3 (Biessmann et al. 1994), and 23Zn3 (Danilevskaya, Lowenhaupt, and Pardue 1998) belong to the Het-A C subfamily. The alignment of the amino acid sequences of the Gag proteins of the isogenic strain subfamilies shows certain variability in the amino-terminal half of the proteins and little sequence variability around the zinc knuckle region (fig. 2) (AJ549588, 97, 603, 605, and 609; AJ635224; AC022280; AF043258). In the alignment, the Het-A A subfamily has the same gap described previously in the 23Zn-1 sequence (Oregon R stock) as a major length polymorphic region (Pardue et al. 1996). The Het-A B subfamily presents a different length polymorphic region closer to the N-terminal of the protein. In addition to these major gaps, there are multiple amino acids changes and small gaps in all Het-A subfamilies, including the Het-A from Drosophila yakuba, which do not change their reading frame, suggesting that there is selection for the incorporation of Het-A elements with functional Gag proteins. It has been proposed that this selection could occur by a mechanism of cotranslational association of the nascent protein with the RNA transposition intermediate (Biessmann et al. 1994; Pardue et al. 1996).

We detected at least seven full-length Het-A elements (fig. 1) that might represent functional copies of four Het-A subfamilies (fig. 2). In several cases, two of these full-length elements, not necessarily from the same subfamily, occur in an array of three elements. Rather than by recombination (Kahn, Savitsky, and Georgiev 2000), the generation of tandem arrays of complete elements could be better explained by the continuous synthesis of cDNA on two or more RNA templates, as has been recently shown for the reverse transcriptase of the R2 non-LTR retrotransposon (Bibillo and Eickbush 2002). This transposition behavior would allow the existence of protected and transpositionally competent Het-A elements at telomeres.
The 5' UTR of the few complete HeT-A elements that have been described, carries two imperfect tandem repeats of an AT-rich 325-bp sequence (Biessmann et al. 1994). In addition, the most 5' region carries one or more copies of a short sequence with an oligo (A) tail that is identical to the extreme 3' end of the 3' UTR (Biessmann et al. 1994). It has been proposed that these sequences derive from the 5' end of the transposing RNA initiated at the promoter located in the 3' end of the upstream HeT-A element (Danilevskaya et al. 1997). We have found that these sequences are also present between the duplicated AT-rich repeats and that related ones are found several times throughout the repeats (fig. 3). The implication is that much of the 5' UTR derives from the 3' UTR by the promoting in tandem mechanism. Such reiterated incorporations of these short sequences would be a consequence of the mechanism of propagation of the several competent HeT-A elements and imply that any complete HeT-A element could behave as a competent copy if an upstream promoter sequence is available.

In the y; cn bw sp strain we found TART A sequences both at telomeres and in the centric heterochromatin of the Y chromosome. TART A elements from Y heterochromatin differ from the telomeric copies in being degenerate elements. Because no complete TART A element has been documented in the databases, we started analyzing the most 5' region of all TART elements. Only two telomeric copies extend into the 5' UTR, very little in one case and about 4 kb in the other (fig. 4). At the XL telomere we have found the longest TART A sequence to date (AJ566116), which may correspond to a TART A master copy (fig. 4). The 5'-3' direct repeats of this element are 4 kb in length and slightly diverged at one end (fig. 4). The presence of analogous repeats in TART B and C subfamilies (Sheen and Levis 1994) implies that the 5' ends of TART elements may have derived from the 3' UTR of an adjacent element following a deletion event (fig. 5).

Our telomere analysis, together with earlier molecular and cytogenetic data (Levis et al. 1993; Walter et al. 1995), shows that every D. melanogaster strain has different telomeres. This variability of telomeres is also common in yeast and humans. Thus, a main conclusion of our experiments is that the organization of Drosophila
telomeric sequences is similar in principal to the organization of telomeric sequences in yeast, mammals, and other multicellular eukaryotes. It seems likely that comparative genomic analysis of the unorthodox telomeres from different Drosophila species will continue to yield important clues in the understanding of the molecular basis of telomere function.

Acknowledgments

We are grateful to R. Levis and M.-L. Pardue for providing us with transposable element probes. We thank R. Levis, D. Gubb, and C. H. Langley for critical comments and suggestions. This work was supported by grant BMC2002-02632 from the Direccio´n General de comments and suggestions. This work was supported by


Literature Cited


Pierre Capy, Associate Editor

Accepted May 25, 2004