Nucleic Acids Research, Vol. 20, No. 14 3731–3736

Structural organization and diversification of *Y*-linked sequences comprising *Su(Ste)* genes in *Drosophila melanogaster*

M.D.Balakireva, Yu.Ya.Shevelyov, D.I.Nurminsky, K.J.Livak[1,*] and V.A.Gvozdev
Institute of Molecular Genetics, Kurchatov Square, 123182, Moscow, Russia and [1]The Du Pont Merck Pharmaceutical Co., Experimental Station, PO Box 80328, Wilmington, DE 19880-0328, USA

Received March 9, 1992; Revised and Accepted June 12, 1992 EMBL accession nos X59157, Z11734 and Z11735

ABSTRACT

Expression of the *X*-linked repeated *Stellate* (*Ste*) genes, which code for a protein with 38% similarity to the β-subunit of casein kinase II, is suppressed by the *Su(Ste)* locus on the *Y* chromosome. The structure and evolution of the *Y*-linked repeats in the region of the *Su(Ste)* locus were studied. The 2800 bp repeats consist of three main elements: the region of homology to the *Ste* genes, an adjacent AT-rich, *Y*-specific segment, and mobile element 1360 inserted in the *Ste* sequence. Amplification of repeats was followed by point mutations, deletions, and insertions of mobile elements. DNA sequencing shows that these repeats may be considered as *Ste* pseudogenes or as damaged variants of a putative gene(s) encoding a protein quite different from the *Ste* protein as a result of an alternative splicing pattern. A comparison of 5 variants of the *Y-Su(Ste)* repeats shows a number of recombination events between amplified and diverged sequences that could be due to either multiple unequal mitotic sister-chromatid exchanges or to gene conversion. It is a first demonstration on a molecular level of these processes occurring in heterochromatic non-rDNA tandemly organized sequences in an eukaryotic genome.

INTRODUCTION

In *Drosophila melanogaster* males lacking a *Y* chromosome, primary spermatocytes contain either needle- or star-shaped proteinaceous crystals. Crystal morphology is determined by the *Stellate* (*Ste*) locus, which maps to position 45.7 on the *X* chromosome (1). This locus has been cloned and studied (2–4). The *Stellate* locus contains tandemly repeated genes whose transcription has only been detected in testes. Copy number of the *Stellate* gene varies among strains, with the highest being approximately 200 copies found in Oregon R. Low copy number corresponds to the Ste^+ allele which produces needle-shaped crystals; and high copy number corresponds to the *Ste* allele and formation of star-shaped crystals in the absence of a *Y* chromosome.

The presence of a specific region of the *Y* chromosome prevents the appearance of crystals in spermatocytes. This *Y*-linked locus was called *Su(Ste)* because it suppresses the *Stellate* phenotype. *Su(Ste)* maps to a region of the *Y* chromosome long arm (1). The *Su(Ste)* region contains moderately repeated sequences homologous to *Ste* that have been postulated to regulate activity of the *X*-linked *Stellate* genes (4). Lack of the *Su(Ste)* region results in reduced fecundity and aberrant meiotic behavior of chromosomes in males.

The mode of the *Su(Ste)-Ste* interaction remains enigmatic and

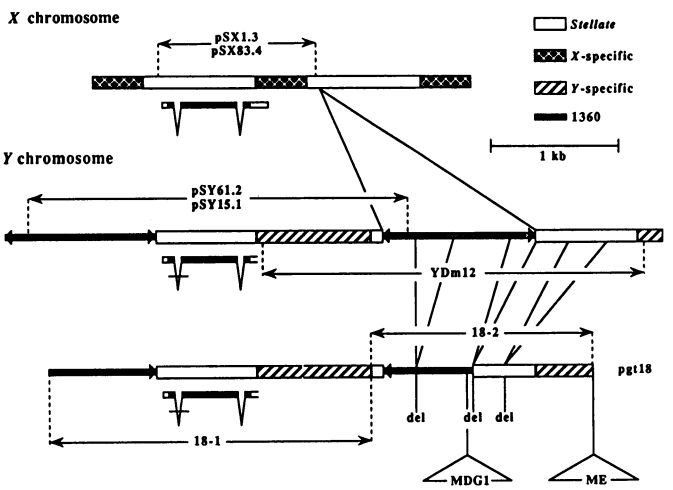
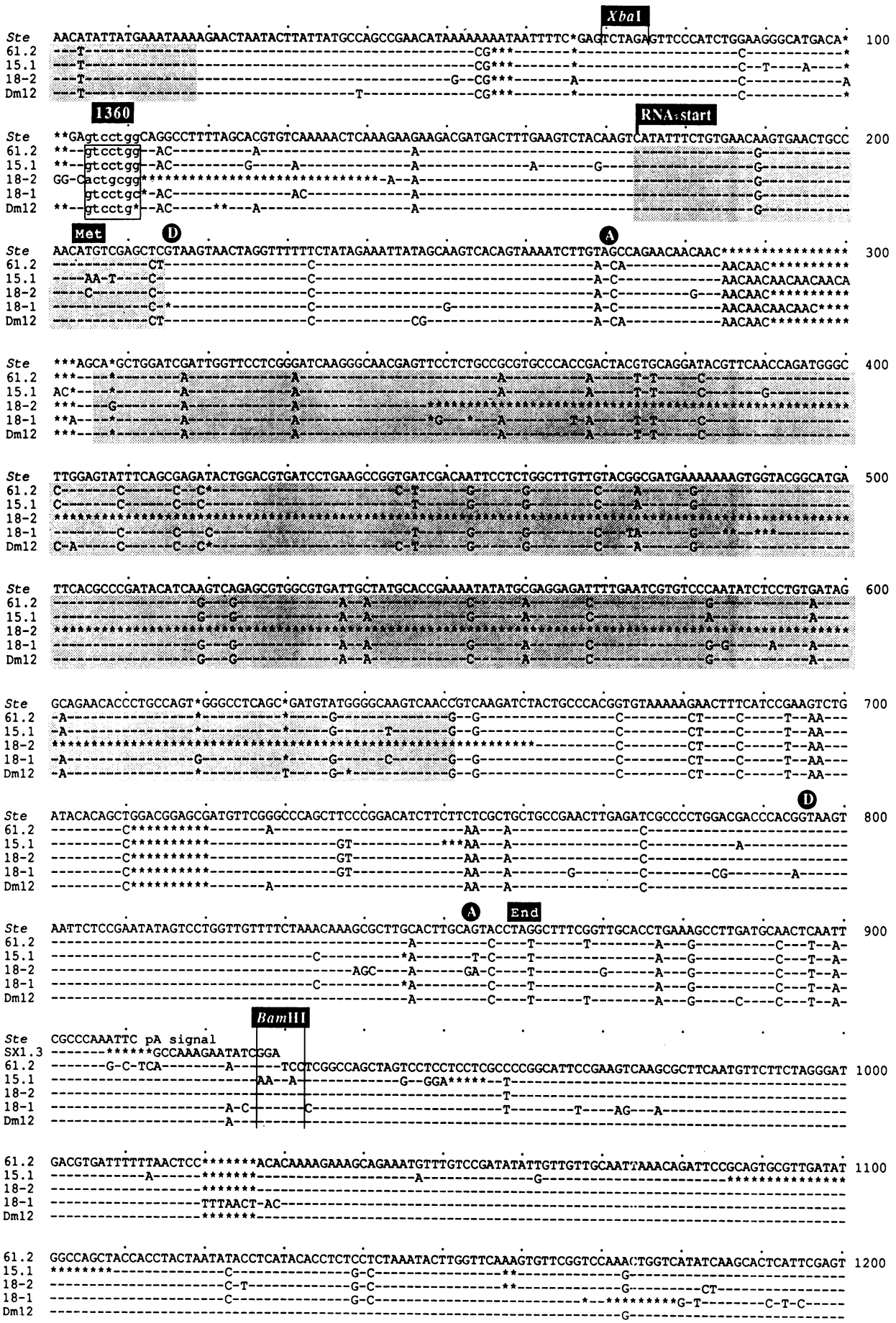


Figure 1. Structural organization of the *Ste* and *Su(Ste)* repeats. As indicated by the key, the arrangements of *Stellate*-homologous, *X*-specific, *Y*-specific, and 1360 elements on the *X* and *Y* chromosomes are shown. The double-headed arrows denote the segment of a repeat that is actually present in the indicated clone. Plasmids pSX1.3 and pSX83.4 contain *XbaI Ste* fragments; pSY61.2 and pSY15.1 contain *MluI Su(Ste)* fragments; and YDm12 contains a *Bam*HI *Su(Ste)* fragment. The location of the *Ste* exons and introns is indicated below each chromosomal segment. The solid triangles denote the inverted repeats of the 1360 mobile element. Deletions (del) and insertions (MDG1 and ME) in the 18-2 segment are shown. MDG1 is a copia-like mobile element and ME indicates an uncharacterized mobile element.

* To whom correspondence should be addressed


XbaI
1360
RNA start
Met
D
A
End
BamHI
pA signal
Ste
61.2
15.1
18-2
18-1
Dm12
SX1.3

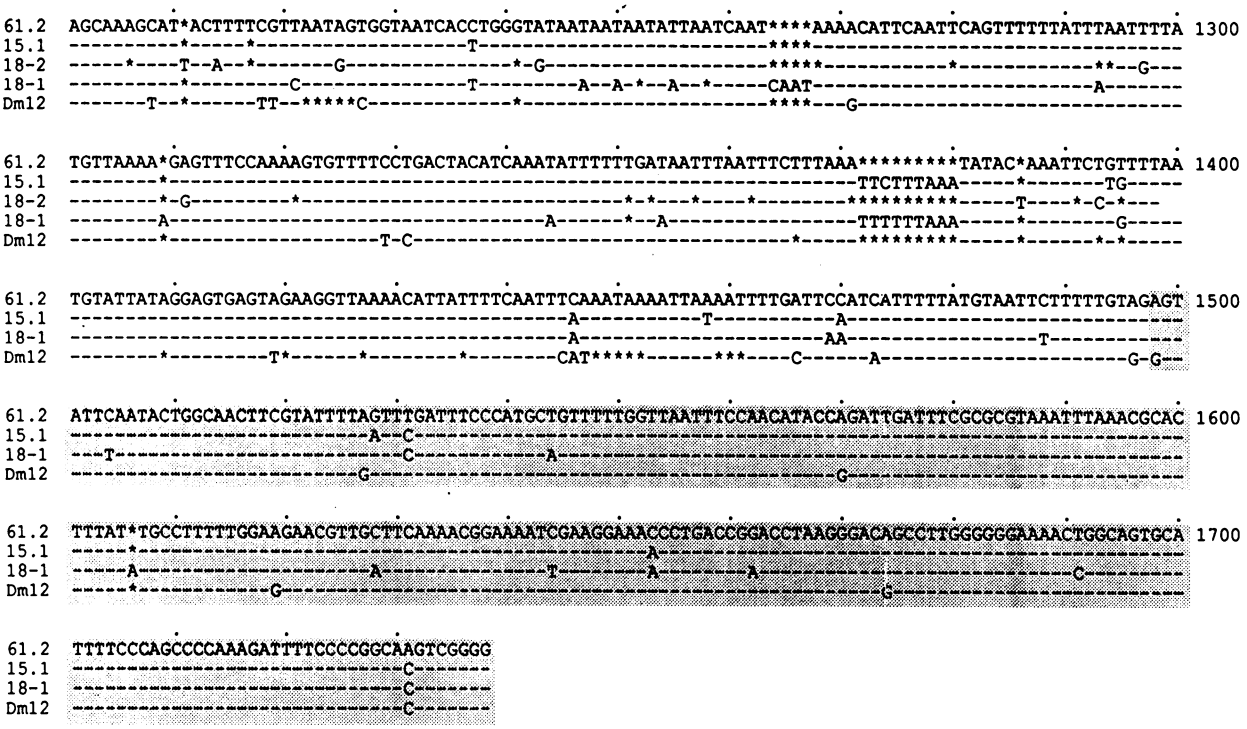


Figure 2. Comparison of five *Y*-linked *Su(Ste)* sequences with the *X*-linked *Stellate* gene. Based on the tandem arrangement of *Ste* and *Su(Ste)* repeats, the sequences of pSX83.4, pSY61.2, pSY15.1, pgt18, and YDm12 have been circularly permuted in order to align them beginning at the start of the *Ste*-homologous element. A hyphen means that the base is the same as in the *Ste* sequence of pSX83.4 or as in the *Y*-specific sequence of pSY61.2. Asterisks mark gaps introduced to maximize the alignment. The region 915–929 in the *Y*-linked sequences shows homology to a region of the *X*-linked variant pSX1.3 which is deleted in pSX83.4. The 7 bp target for the insertion of the 1360 mobile element is shown in lower case letters. The target sequences in the box should be replaced with the 1360 sequences of Figure 3 in order to generate the complete sequence of the *Y*-linked *Su(Ste)* repeats. For the *Ste* gene, the RNA start, the start and end of translation, the GT of splice donors (D) and the AG of splice acceptors (A) are designated according to Livak (4). The shaded sequences indicate the putative exons discussed in the text. The *Xba*I site used to clone the *Ste* fragments of pSX83.4 and pSX1.3 and the *Bam*HI site used to clone the *Su(Ste)* fragment of YDm12 are marked. The pSX83.4 and pSX1.3 sequences are from Livak (4). The YDm12 sequence is from Kholodilov *et al.* (10), where the opposite strand was shown. The EMBL accession numbers for the remaining sequences are: Z11734 (pSY61.2), Z11735 (pSY15.1) and X59157 (pgt18).

detailed studies of the *Su(Ste)* sequences may be considered as a first step to clarify this intriguing type of gene relationship. At the same time questions concerning the evolution, amplification and peculiarities of diversification of repeated heterochromatic sequences represent a separate, unresolved topic.

MATERIALS AND METHODS

Plasmids pSY15.1 and pSY61.2 contain *Mlu*I fragments, each corresponding to a single 2.8 kb *Y* unit homologous to *Ste* (4). The pgt18 cosmid consists of several *Y-Ste* repeats interrupted by insertions of mobile elements (5). Two of these repeats were subcloned in pUC19 and sequenced by the method of Maxam and Gilbert (6) using the modified nested deletion technique (7). Briefly, the 4K element from pUC4K (8) was inserted into a plasmid as a protector to prevent vector degradation by Bal31 nuclease. After linearization of the plasmid between inserted and protector sequences and limited Bal31 digestion, the remainder of the protector was removed and plasmid DNA was recircularized. Other standard procedures were according to Sambrook *et al.* (9).

RESULTS

It was known that the *X*-linked repeated *Ste* genes are characterized by a 950-bp *Cfo*I fragment, while the homologous *Y*-linked sequences by an 800-bp *Cfo*I fragment (3,4). According to this characteristic, five variants of the repeats from the *Y* chromosome were selected to study their structure and diversity as compared to the homologous *Ste* sequences from the *X* chromosome. The *Y* repeats will be designated *Su(Ste)*-repeats although only circumstantial evidence indicates that these are the *Y* sequences responsible for suppression of the *Stellate* genes (4).

The DNA sequences were derived from the following clones: pSY61.2 and pSY15.1—carrying very similar but not identical fragments from the *Su(Ste)* locus (4); plasmid YDm12 from the *Y* chromosome—carrying mobile element 1360 (10) and shown to be inserted in a *Ste*-homologous sequence [the 1360 element has also been described as hoppel (11)]; and the cosmid pgt18—carrying a *Y* chromosome fragment containing the copia-like element MDG1 and two copies of *Ste*-homologous sequences (5). The overview of their structures is presented in Figure 1 and the alignment of the sequences is presented in Figures 2 and 3.

Figure 1 shows the tandem arrangement of the *Su(Ste)* 2.8 kb repeat (4). The repeat unit consists of a *Ste*-homologous region, a *Y*-specific sequence, and an inserted 1360 mobile element. The *X*-linked *Ste* genes were shown to encode an open reading frame (ORF) sharing 38% similarity with the β-subunit of casein kinase II (4). The three short exons of the *Ste* genes are marked in Figures 1 and 2. As diagrammed in Figure 1, the region of homology in the *X* and *Y* repeats are flanked by unrelated sequences, so the 3′ region of the *X*-linked *Ste* genes carry polyadenylation signals that are missing in the *Y*-linked *Su(Ste)*-repeats. A detailed comparison of the *Ste*-like sequences from

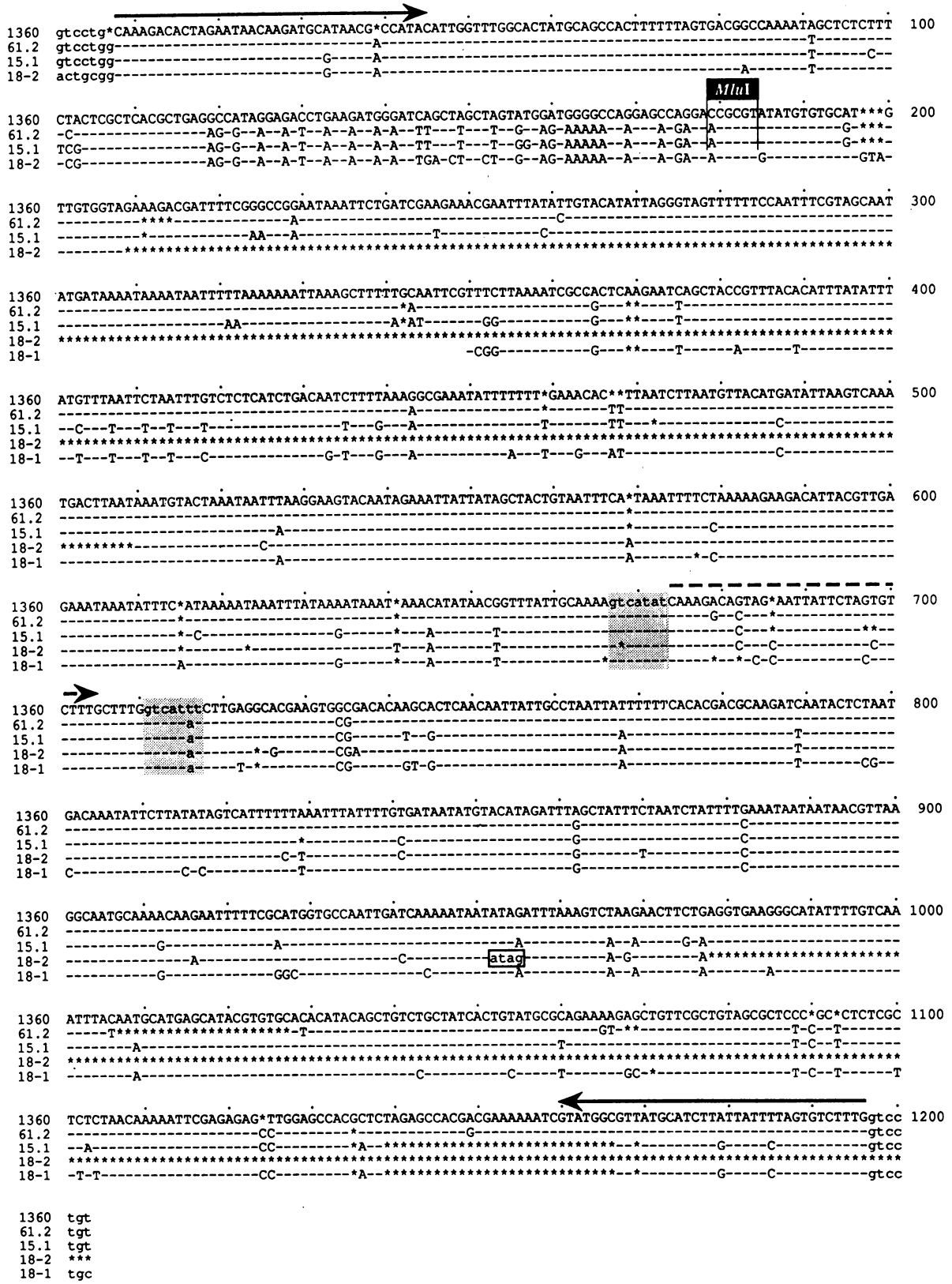


Figure 3. Comparison of the 1360 element sequence of YDm12 to the other four *Su(Ste)* sequences. The sequence marked 1360 is that of YDm12.The sequences of pSY61.2 and pSY15.1 have been permuted in order to align them with the intact 1360 sequence. The 7 bp duplication generated by the insertion of 1360 is indicated by the lower case letters at the beginning and end of each sequence. The solid arrows mark the 37 bp inverted repeats of the 1360 mobile element. The dashed arrow shows a region of internal homology to the inverted repeat sequence. The shaded areas mark the putative target duplication that may be a vestige of a previous mobile element insertion (see text). The boxed sequence is the 4 bp target that is the site for insertion of the MDG1 mobile element into the 18-2 segment. The *Mlu*I site used to clone the *Su(Ste)* fragments of pSY61.2 and pSY15.1 is marked. References for the sequences are in the Figure 2 legend.

the *X* and *Y* chromosomes (Figure 2) shows that the *Y* repeats are probably pseudogenes. All *Y* variants conserve an unaltered RNA start site, but the initiator codon has been changed in the 15.1 and 18-2 variants. All *Y* variants have the same G to C transversion that alters the canonic AG dinucleotide in the 3′ acceptor site of the first *Ste* intron. Variants 15.1 and 18-2 also show alteration of the 3′ acceptor site for the second *Ste* intron. All *Y*-linked sequences contain several stop codons that interrupt the *Ste* ORF. Finally, the 18-2 variant has an extensive deletion that eliminates much of the *Ste* ORF.

In the *Y*-linked *Su(Ste)* repeats, the *Ste*-region is attached to an unrelated AT-rich segment that we have called the *Y*-specific sequence. Based on examination of the sequence, there is no homology or structure that suggests how these two sequences were joined.

The *Su(Ste)* repeats contain a third element—the mobile 1360 element flanked by inverted repeats (10). The location of the 1360 element suggests that the *Ste*-region and *Y*-specific region were joined first, followed by insertion of the 1360 element and amplification of the entire structure [*Su(Ste)* repeat unit]. Variant 18-2 has two deletions in the 1360 element, one of which includes the right end of the element and 30 bp of the *Ste*-region (Figures 1,2). Figure 3 shows a sequence comparison of the 1360 element from YDm12 with the other four *Su(Ste)* repeats. These sequences show that the 1360 element is bounded by inverted repeats of 37 bp and results in a 7 bp target site duplication upon insertion. Within the central portion of the 1360 element, there is a short region of homology that may be a vestige of a similar mobile element insertion. This insertion shows homology (31 nucleotides) to both inverted repeats flanked by a 7 bp duplication. The duplication of 5 nucleotides (CTTTG) at the right end of the vestigial insertion may be a result of slippage during replication.

The 18-2 *Su(Ste)* repeat from pgt18 was further modified by insertion of the copia-like element MDG1 (12) within the 1360 element. This 18-2 variant also contains an additional insertion in the *Y*-specific region. *In situ* hybridization with this insertion suggests that it may be yet another mobile element (data not shown).

DISCUSSION

The sequence comparison of Figure 2 strongly suggests that the *Y*-linked *Su(Ste)* repeats are pseudogenes. On the other hand, by examining a consensus of the five *Su(Ste)* sequences, it is possible to propose an alternative splicing scheme that could result in an intact ORF. This putative construct consists of exon 1 from 174 to 214, intron 1 from 215 to 305, exon 2 from 306 to 651, intron 2 from 652 to 1497, and exon 3 from 1498 to the polyadenylation signal corresponding to position 13−18 in the next repeat. There are a number of problems, though, with this proposed RNA structure. First, it is only the consensus of the five *Su(Ste)* sequences that exhibits an intact ORF. Each individual sequence has at least one frameshift or stop codon. Thus, it must be proposed that a functional *Su(Ste)* repeat has not yet been cloned and sequenced. Second, the 5′ donor sequence of intron 2 (CG/GTGAAG) and especially the 3′ acceptor sequence of intron 1 [$(AAC)_nAG/G$] do not match well with the major consensus donor (AG/GTAAGT) and acceptor [(T or C)$_n$NCAG/G] sequences (13,14). Finally, the proposed *Su(Ste)* ORF would encode a protein quite different from the *Ste* protein. Given the recent evolutionary origin of these sequences in Drosophila (3,4), it is difficult to imagine a selection mechanism that would result

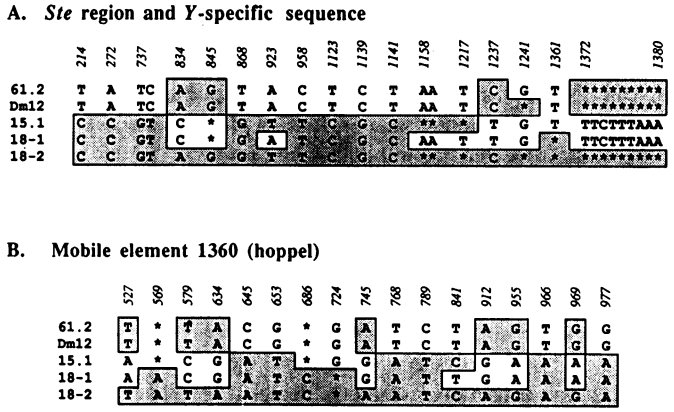


Figure 4. Sequence comparison of *Su(Ste)* repeats detecting recombination events. Shaded areas denote identity with the 18-2 variant. Asterisks represent deletions. The italicized numbers indicate nucleotide positions from Figure 2 (A) or Figure 3 (B).

in two such different proteins from the same underlying DNA sequence. If any *Su(Ste)* repeat encodes a functional protein, this must be shown by direct cDNA cloning and sequencing.

The presented data indicate the following steps in the evolution of *Su(Ste)* repeats from ancestral *X*-linked *Ste* genes:

1. Fusion of the *Ste* sequence and an unrelated AT-rich sequence of the *Y* chromosome using an undefined mechanism of 'illegitimate recombination.' Perhaps, this type of recombination is a property of heterochromatic sequences. This view is supported by the data on telomere healing in *D. melanogaster*, where putative heterochromatic sequences do not require a homology to recombine with a broken end of a chromosome (15).
2. Insertion of mobile element 1360 (hoppel) into the 5′ region of the *Ste* sequence.
3. Amplification of the whole construct carrying *Ste*-like pseudogenes (or genes) flanked by AT-rich sequences.
4. Divergence of *Su(Ste)* repeats by point mutations, deletions, and mobile element insertions.

Comparison of the five sequence variants reveals the occurrence of recombinational events among the variants. Figure 4 shows the cases of point mutations and short deletions (insertions) shared by two variants as compared to the other three. A close similarity is shared by the Dm12 and 61.2 variants, the only difference being a G nucleotide deletion in Dm12 at position 1241. The similarity of the 15.1 and 18-1 sequences is obvious, although slightly less than for the Dm12-61.2 pair. Figure 4a shows that the 18-2 sequence may be considered as composed of different segments shared with either Dm12 or 15.1, with one segment in common with 18-1. A similar pattern of recombination events is revealed by the sequences in the body of the 1360 mobile element (Figure 4b). The boundaries of segments exchanged is impossible to define because of the extended regions of perfect homology. The size of these segments seems to be on the order of tens (e.g., from positions 737 to 868) to hundreds of nucleotides. There is the possibility of clustering as recombination events are observed in the 3′ part of the *Ste*-region, the *Y*-specific region, and the 1360 element, but not in the 5′ part of the *Ste*-region.

These patterns may be explained as a result of multiple unequal mitotic exchanges between sister chromatids or by gene

conversion. Both processes have been shown to affect the rDNA tandem repeats. Unequal exchange is involved in the *X*-linked rDNA magnification phenomenon in *D. melanogaster* (reviewed in 16). Biased gene conversion has recently been demonstrated during the homogenization of rDNA copies in lizards (17). We suggest the revealed recombination events (unequal exchange or gene conversion) represent the first documented case of their functioning in the evolution of heterochromatic non-rDNA tandemly organized sequences. These recombination acts may maintain a high level of diversification of heterochromatic sequences where meiotic recombination is prohibited.

The indication that the *Su(Ste)* repeats evolved from *X*-linked *Stellate* fits well with the speculation of Hurst (18) that *X*-linked *Stellate* was originally a meiotic drive gene. He proposes that the *Stellate* gene product favored transmission of the *X* chromosome relative to the *Y* chromosome, resulting in expansion of the *Ste*-carrying *X* chromosome in the population. The *Su(Ste)* sequences arose after the invasion of the *Stellate* gene and acted in some manner to suppress production of *Ste* product, and thus increased transmission of the *Su(Ste)*-carrying *Y* compared with the wild-type *Y* in such a population. An 'arms race' between the meiotic driving *Ste* genes and the suppressing *Su(Ste)* sequences is a plausible explanation for how related sequences became amplified on both the *X* and *Y* chromosomes. One complication is that *Ste*-homologous sequences are found only on the *Y* chromosome in *Drosophila simulans* (3). The meiotic drive hypothesis requires that *Ste* invaded the *X* and *Su(Ste)* became established on the *Y* before *D. melanogaster* and *D. simulans* diverged and that *Ste* was lost from the *D. simulans* *X* after divergence.

Recently published data (19) suggest there may be other *Ste*-like tandem repeats linked to the *Su(Ste)* region that could interact with the *X*-linked *Stellate* genes. Also, as discussed above, there may by a *Y*-linked ORF that codes for a *Su(Ste)* protein. Therefore, the *Su(Ste)* region of the *Y* chromosome may be considered a complex cluster carrying different types of repeats. It is yet to be determined which type of these repeats is actually involved in suppression of the *X*-linked *Stellate* genes.

REFERENCES

1. Hardy, R. W., Lindsley, D. L., Livak, K. J., Lewis, B., Siversten, A. L., Joslyn, G. L., Edwards, J. and Bonaccorsi, S. (1984) *Genetics* **107**, 591–610.
2. Lovett, J. A. (1983) Ph.D. thesis, Indiana University, Bloomington, IN.
3. Livak, K. J. (1984) *Genetics* **107**, 611–634.
4. Livak, K. J. (1990) *Genetics* **124**, 303–316.
5. Shevelyov, Yu. Ya., Balakireva, M. D. and Gvozdev, V. A. (1989) *Chromosoma* **98**, 117–122.
6. Maxam, A. M. and Gilbert, W. (1980) In Grossman, L., and Moldave, K. (eds.), Methods in Enzymology. Academic Press, New York, Vol. 65, pp. 499–560.
7. Nurminsky, D. I., Shevelyov, Yu. Ya. and Kalmykova, A. I. (1989) *Bioorganic Chemistry (Russ.)* **15**, 1137–1139.
8. Vieira, J. and Messing, J. (1982) *Gene* **19**, 259–268.
9. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular cloning. A Laboratory Manual. Second Edition. Cold Spring Harbor University Press, Cold Spring Harbor.
10. Kholodilov, N. G., Bolshakov, V. N., Blinov, V. M., Solovyov, V. V. and Zhimulev, I. F. (1988) *Chromosoma* **97**, 247–253.
11. Kurenova, E. V., Leibovitch, B. A., Bass, I. A., Bebikhov, D. V., Pavlova, M N. and Danilevskaya, O. N. (1990) *Genetica (Russ.)* **26**, 1701–1712.
12. Ilyin, Yu. V., Tchurikov, N. A., Ananiev, E. M., Rysdov, A. P., Yenikolopov, G. N., Limborska, S. A., Maleeva, N. E., Gvozdev, V. A. and Georgiev, G. P. (1978) *Cold Spring Harbor Symp.* **42**, 959–969.
13. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
14. Shapiro, M. B. and Senpathy, P. (1987) *Nucleic Acids Res.* **15**, 7155–7174.
15. Biessmann, H., Mason, J. M., Ferry, K., d'Hulst, M., Valgeirsdottir, K., Traverse, K. L. and Pardue, M-L. (1990) *Cell* **61**, 663–673.
16. Hawley, R.S. and Marcus, C.H. (1989) *Annu. Rev. Genet.* **23**, 87–120.
17. Hillis, D.M., Moritz, C., Porter, C.A. and Baker, R.J. (1991). *Science* **251**, 308–310.
18. Hurst. L. D. (1992) *Genetics* **130**, 229–230.
19. Danilevskaya, O. N., Kurenova, E. W., Pavlova, M N., Bebehov, D. V., Link, A. Y., Koga, A., Vellek, A. and Hartl, D. L. (1991) *Chromosoma* **100**, 118–124.