Molecular Evolution of Two Paralogous Tandemly Repeated Heterochromatic Gene Clusters Linked to the X and Y Chromosomes of Drosophila melanogaster

Galina L. Kogan, Vitalii N. Epstein, Alexei A. Aravin, and Vladimir A. Gvozdev

Institute of Molecular Genetics, Moscow, Russia

Here we report the peculiarities of molecular evolution and divergence of paralogous heterochromatic clusters of the testis-expressed X-linked Stellate and Y-linked Su(Ste) tandem repeats. It was suggested that Stellate and Su(Ste) clusters affecting male fertility are the amplified derivatives of the unique euchromatic gene βCK2tes encoding the putative testis-specific β-subunit of protein kinase CK2. The putative Su(Ste)-like evolutionary intermediate was detected on the Y chromosome as an orphon outside of the Su(Ste) cluster. The orphon shows extensive homology to the Su(Ste) repeat, but contains several Stellate-like diagnostic nucleotide substitutions, as well as a 10-bp insertion and a 3’ splice site of the first intron typical of the Stellate unit. The orphon looks like a pseudogene carrying a drastically damaged Su(Ste) open reading frame (ORF). The putative Su(Ste) ORF, as compared with the Stellate one, carries numerous synonymous substitutions leading to the major codon preference. We conclude that Su(Ste) ORFs evolved on the Y chromosome under the pressure of translational selection. Direct sequencing shows that the efficiency of concerted evolution between adjacent repeats is 5–10 times as high in the Stellate heterochromatic cluster on the X chromosome as that in the Y-linked Su(Ste) cluster, judging by the frequencies of nucleotide substitutions and single-nucleotide deletions.

Introduction

The processes of the origins of multigene families in eukaryotic genome, as well as mechanisms supporting identity of divergence of their members, has been a topic of numerous discussions (Liao 1999). Special attention might be given to the homologous (paralogous) repeats which have been shown to be functionally related. An unusual and puzzling example of such repeats is represented by paralogous Stellate and Su(Ste) tandem repeats localized on the X and Y chromosomes, respectively, in the Drosophila melanogaster genome (Hardy et al. 1984; Livak 1990). Both of these repeats contain open reading frames (ORFs) with extensive homology to the β-subunit of protein kinase CK2 (Livak 1990; Kalmykova, Dobritsa, and Gvozdev 1997). The Su(Ste) repeats are considered suppressors of testis-specific Stellate transcription, since their deletion causes Stellate hyperexpression, coupled with male sterility (Hardy et al. 1984). The Stellate repeats are located in the euchromatin of the X chromosome, in the 12E region of polytene chromosome map, as well as in constitutive heterochromatin of the X chromosome (Shevelyov 1992; Palumbo et al. 1994; Tulin et al. 1997). The polymorphism of Stellate repeats and their homogenization in the heterochromatic X-linked cluster has been studied (Tulin et al. 1997). The Stellate and Su(Ste) repeats were related to the special “parasitic system” (Bozzetti et al. 1995) or to the pair of “self-promoting elements” driving a balance of male and female progenies in a population (Hurst 1996; Hurst and Smith 1998). It was shown recently that the unique euchromatic testis-expressed gene βCK2tes, encoding the testis-specific β-subunit of protein kinase CK2, might be considered a precursor of the X-linked Stellate and the Y-linked Su(Ste) repeats (Kalmykova et al. 1997). Here, further study of Su(Ste) repeats and detection of a Su(Ste) orphon allow us to suggest that the Su(Ste) Y-linked repeats and their coding capacities, at least at some periods during their evolution, evolved on a Y chromosome under the pressure of translational selection. We accentuate that the Y-linked Su(Ste) repeats carry at least stretches of ORFs evolved under selection on a level of translation. The origin of the Y chromosome is usually thought to occur as a result of degeneration of the X chromosome (Charlsworth 1996). In contrast, our observations might support a view that at least some segments of the Y chromosomes of Drosophila evolved acquiring repeats capable of affecting fertility of individuals (Hackstein et al. 1996). Here, we also report some peculiarities of comparative intralocus concerted evolution of paralogous Stellate and Su(Ste) clusters.

Materials and Methods

The bacteriophage P1 library was screened by hybridization to Stellate and a Y-specific Su(Ste) probe to obtain P1E8 and P13H1 phages (Kalmykova, Dobritsa, and Gvozdev 1998). The orphon was subcloned from P1E8 into the PstI site of pTZ19R. To obtain Su(Ste) oligomers, the fragments of phage P13H1 were subcloned into pTZ19R after partial XbaI restriction. The 300-bp BgII/HindIII fragments of clone 72, comprising three tandem Su(Ste) copies, were subcloned into pTZ19R to obtain the 72-1, 72-2, and 72-3 subclones. The orphon and Su(Ste)-carrying plasmid subclones were sequenced on both strands from universal plasmid or Su(Ste)-specific primers by Sanger’s procedures (Sambrook, Fritsch, and Maniatis 1989) using Sequenase, version 2.0, T7 DNA polymerase (U.S. Biochemical).
Figure 1.—Structure of the tandemly repeated Su(Ste) unit and the PstI clone containing the Su(Ste) orphon. The box flanked by arrowheads (inverted repeats) shows the position of the hoppel (1360) transposon. The region homologous to the Stellate unit and the Y-specific part are indicated by open and blackened rectangles, respectively. The black bar indicates the size of sequenced Bg-H fragments in the adjacent Su(Ste) units. The stippled rectangle near the orphon indicates the position of the AT-rich region homologous to the lamin associated repeat. The abbreviations used for restriction endonucleases are as follows: X, XhoI; H, HindIII; Bg, BgII; B, BamHI; RI, EcoRI; P, PstI.

Results
Su(Ste) Orphon

The 6.5-kb PstI fragment with extensive homology to the Su(Ste) unit was isolated by subcloning of the P1E8 Stellate-positive clone (fig. 1). Direct sequencing of the regions including both XhoI sites revealed the presence of anonymous sequence at the 5′ XhoI site and the AT-rich repeat (610) at the 3′ XhoI site, similar to that which was shown to specifically interact with nuclear lamins (Baricheva et al. 1996). This result allows us to consider the Stellate-like sequence an orphon, localized outside of the cluster and flanked by unassigned repeats. The orphon, included in the PstI fragment, was easily detected by Southern analysis as the 6.5-kb PstI fragment in several randomly chosen stocks (not shown) because the regular Stellate and Su(Ste) repeats contain no PstI sites and are revealed in the DNA bands of high molecular weights (exceeding 15 kb). Comparative Southern analysis of DNA from males and females revealed that this Stellate-like fragment is located on the Y chromosome (not shown). The detected orphon contains 52 diagnostic nucleotide substitutions (fig. 2) coinciding with the Su(Ste) sequence, as well as the downstream Y-specific sequence peculiar to the Su(Ste) repeats, starting from position 1029. At the same time, only 12 diagnostic nucleotide positions peculiar to Stellate sequence were detected in the orphon sequence (fig. 2). The orphon contains the remnant (~80 bp, upstream of position 212, including 37 bp of inverted repeat) of transposon 1360 (hoppel), inserted in each tandemly repeated Su(Ste) unit (Balakireva et al. 1992). Thus, as a whole, the orphon is much more related to the Su(Ste) repeats than to the Stellate ones. The orphon also carries two typical diagnostic features of Stellate repeats: the position of the 3′ splice site of the first intron is similar to the Su(Ste) copies, and the 10-bp deletion of the main exon (positions 826–835) peculiar to the Su(Ste) repeat is absent, as in Stellate copies. The nucleotide Stellate-like positions in the orphon sequence are randomly distributed along the orphon sequence, and no clusters of these positions were observed. Thus, the origin of similarities between the orphon and the Stellate repeat seems improbable as a result of ectopic conversion events. Our observations suggest that the orphon, still carrying a set of diagnostic features of the X-linked Stellate repeats, represents the damaged dead derivative of an ancestor of amplified heterochromatic Y-linked Su(Ste) repeats (see below).

ORFs in Stellate and Su(Ste) Repeats

Figure 2 presents a comparison of the orphon sequence with the chosen Stellate and Su(Ste) copies. Disturbances of the Su(Ste) ORF as compared to the Stellate one have been described earlier (Kalmykova, Dobritsa, and Gvozdev 1998). Here we focused on the peculiarities of synonymous substitutions in the course of the divergence of the Stellate and putative Su(Ste) ORFs. Figure 2 presents the Stellate ORF, as well as positions of amino acids corresponding to the silent or replacement nucleotide substitutions in the Su(Ste) sequence. The Stellate ORF in the Su(Ste) repeats is impaired as a result of G-to-C transversion in the end of first intron; the boundary is shifted downstream to canonical AG nucleotides at positions 383–384. It was shown that this Stellate damage in the Su(Ste) ORF may be restored downstream as a result of nucleotide deletions. Thus, because of two single-nucleotide deletions (positions 429 and 523, indicated by vertical arrows in fig. 2), the Stellate-like ORF encompassing conservative putative zinc finger CPX2CX32CPXC (positions 701–791) is restored in Su(Ste)/511 cDNA, downstream of the shifted site of splicing (Kalmykova, Dobritsa, and Gvozdev 1998). We compared the Stellate ORF with the whole putative unaltered ancient Su(Ste) ORFs, considering the damage at the beginning and the end of putative ORF secondary events occurring after the maintenance period of this ORF. The cases of synonymous or nonsynonymous substitutions in the Su(Ste) ORF as compared with those of the Stellate one are marked by positions of amino acids indicated below the Stellate ORF (fig. 2). We accepted that the observed shift of the 3′ site of splicing caused ORF damage, as well as some other damage that occurred later, following the evolution of the Su(Ste) ORF, driven by selection pressure. Actually, comparison of Stellate and overall putative Su(Ste) ORFs revealed 21 synonymous substitutions, 7 similar substitutions, and only 9 replacement mutations. Twelve substitutions among the total 21 synonymous sites caused the usage of the more preferred codons (indicated by bold letters in figure 2) in the Su(Ste) ORF, taking into account the results of codon bias studies in the D. melanogaster genome (Sharp and Matassi 1994; Akashi 1997). On the contrary, only three substitutions (boxed) cause the usage of rare codons. Thus, the putative Su(Ste) ORF is biased to the usage of more preferred codons than the Stellate ORF. The same codon bias was detected in the Su(Ste)OR sequence, showing that translational selection was finished in the course of the Su(Ste) repeat evolution before the change of the 3′ splice sites of the first intron and the appearance of the other damages.
As a whole, pairwise comparison of the Su(Ste) and Stellate sequences reveals traces of translational selection which might be responsible for the maintenance of a definite level of translational efficiency of these repeats. The Su(Ste) ORF containing the conservative zinc finger is maintained, whereas it is damaged in the orphon. The putative conservative zinc finger CPX_CX_{22}CPX is also damaged as a result of a 5-bp deletion (positions 710–714) and elimination of the second Cys residue. Thus, the orphon might be considered a profoundly damaged pseudogene, since numerous nucleotide deletions, as well as G and C insertions, impair the ORF. The initiator ATG codon is transformed to the ATT Ile codon. The 5′ splice site in the second intron is damaged as a result of a T deletion in the canonical GT dinucleotide (fig. 2). The orphon is much more damaged than the Su(Ste) repeats that have been sequenced (Kalmykova et al. 1992; Kalmykova, Dobritas, and Gvozdev 1998). Detection of the orphon accentuates the existence of selective pressure and homogenization maintaining stretches of ORFs in the clustered Su(Ste) repeats.

Concerted Evolution of the Y-Linked Su(Ste) Repeats

The Su(Ste) size variants, as well as Su(Ste) units with similar restriction sites (McKee and Satter 1996; Kalmykova et al. 1997), appear to be clustered in particular subintervals of the locus. The randomly chosen Su(Ste) repeats differ at 6.4% of nucleotide size, as compared with 2.5% of divergence for homologous Stellate repeats (McKee and Satter [1996]; calculated from the previously reported data of Balakireva et al. 1992 and Shevelyov [1992]). To directly evaluate the efficiency of concerted evolution of Su(Ste) repeats, the fragments of adjacent clustered Su(Ste) repeats were sequenced. The presence of a driver effect of concerted evolution was shown, since all three repeats (72-1, 72-2, and 72-3) contain six diagnostic nucleotide substitutions, as well as three shared single-nucleotide deletions, along the sequenced regions (fig. 2). The earlier randomly chosen and sequenced Su(Ste) copies or cDNA clones do not contain these substitutions or deletions (Balakireva et al. 1992; Kalmykova, Dobritsa, and Gvozdev 1998). Results of the pairwise sequence comparison of three adjacent Su(Ste) repeats show three to four substitutions per 330 nt, that is, more than 1% of divergence. At the same time, intralocus divergence of the adjacent heterochromatic Stellate repeat does not exceed 0.1%–0.2% (Tulin et al. 1997). Thus, the efficiency of homogenization of Su(Ste) repeats is 5–10 times as low as that of Stellate repeats.

Discussion

The Stellate repeats and their suppressors, Su(Ste) repeats, were considered an evolved parasitic system capable of maintaining itself (Bozetti et al. 1995). This system was thought by others to contain self-promoting genetic elements (Hurst 1996), in which the original Stellate gene affected DNA packing of the X chromosome and its successful transmission to progeny. It was proposed that Stellate repeats act as a driver causing preferential transmission of the X chromosome, thus biasing the sex ratio toward females. The evolved Su(Ste) repeats on the Y chromosome were proposed to act in a dose-dependent fashion (Hurst 1996), causing a balance between transmission of the X chromosome and Y chromosome efficiency. The secondary origin of Su(Ste) as a suppressor of driver effects exerted by Stellate was suggested (Hurst 1996). However, Stellate repeats do not exert any driver effect (Palumbo et al. 1994; L. Robbins, personal communication). Alternatively, the Y-linked Su(Ste) repeats can represent an immediate derivative of euchromatic βCK2es-like gene, and possibly one that is more ancient than the X-like Stellate repeats. The Stellate-like sequence is present only on the Y chromosome of Drosophila simulans (Livak 1984). We propose the presence of the Y-linked ancestors of the Su(Ste)–like repeats before the split of the melanogaster and simulans lineages. Actually, the rough estimation (Kalmykova et al. 1997) of the period of divergence of Stellate and Su(Ste) repeats (6 Myr) exceeds the time (2 Myr) elapsed since the split of the melanogaster and simulans lineages (Russo, Takezaki, and Nei 1995). In this report, evidences was presented that Su(Ste) repeats underwent translation selection for the more preferred codons, whereas codon bias for the Stellate units declined. This latter relaxation of selection was demonstrated for D. melanogaster since its split from D. simulans (Akashi 1997).

We detected the Y-linked Su(Ste) orphon detached from the main cluster of tandemly repeated units. The orphon sequence combines diagnostic traits of both types of diverged Stellate and Su(Ste) paralogous repeats. Thus, detection of this “molecular fossil” on the Y chromosome allows us to suggest that the Su(Ste) repeated cluster really evolved on the Y chromosome. The orphon carries the same evidence of translational selection as the Su(Ste) repeats, although this orphon repeat was drastically damaged, escaping the pressure of concerted evolution which has continued to be operative in the Su(Ste) cluster, at least in the region encoding the zinc finger motif of protein kinase CK2. The putative ancestor on the Y chromosome of both the X-linked Stellate and the Y-linked Su(Ste) repeats might be represented by the close ancient relative of the orphon sequence, comprising the Stellate-like 3′ splice site of the first intron, the inserted transposon hoppel (1360), and the ORF still subjected to the pressure of translational selection. We propose that this Y-linked repeated ancestor engendered the contemporary Stellate repeats, as well as Su(Ste) ones. These latter repeats contained partially damaged ORFs, possibly acquiring a new function as suppressors of Stellate repeats when a period of translational selection was transformed to episodic changes and damage to the Su(Ste) ORFs.

We want to accentuate the probability of translational selection on a Y chromosome, which might be opposed to the commonly accepted and established view of genetic erosion and degeneration of the Y chromosomes (Charlesworth 1996). The Su(Ste) repeats conserved stretches of Stellate ORFs and the possibility of
their translation was discussed (Kalmykova, Dobritsa, and Gvozdev 1997). In contrast, the orphan shows numerous damages to this ORF, including the mutation of the Cys residue in the region of the conservative putative zinc finger responsible for dimerization of protein kinase CK2 subunits (Chantalat et al. 1999). Comparative zinc finger responsible for dimerization of protein sequence are indicated by bold figures. The evolved rare codons are boxed. Stellate-like diagnostic nucleotide positions in the orphon are underlined. Two vertical arrows indicate sites of single-nucleotide deletions at position 429 in Su(Ste) sequence and a hyphen means that the base is missing. R designates A or G. The GT of splice donors (D) and the AG of splice acceptors (A) in the Su(Ste) transscripts are indicated by bold italics. A horizontal arrow marks the inverted repeat (37 bp) of the hoppel transposon. The predicted amino acid sequence of the Su(Ste) protein is shown above the DNA sequence; amino acids of the putative of Su(Ste) ORF at the sites of nonsynonymous and synonymous nucleotide substitutions are shown below. Residues in a conservative cystein-rich potential metal-binding motif (PX3CX22CPXC) are indicated by bold letters. The positions of codons in the Stellate sequence where nucleotide substitutions cause the usage of preferred codons in the Su(Ste) sequence are indicated by bold figures. The evolved rare codons are boxed. Stellate-like diagnostic nucleotide positions in the orphon are underlined. Two vertical arrows indicate sites of single-nucleotide deletions at position 429 in Su(Ste) cDNA511 (Kalmykova, Dobritsa, and Gvozdev 1997) and position 523 in Su(Ste) genomic sequences. Sequences 72-1, 72-2, and 72-3 correspond to the adjacent three Su(Ste) repeats; sites of shared diagnostic nucleotide substitutions and deletions are shaded. The sequences of Su(Ste)OR, 72-1, 72-2, and 72-3 have been deposited in GenBank under accession numbers AF173960, AF173957, AF173958, and AF173959, respectively.

The role of gene conversion or unequal crossing over between repeats in the Su(Ste) (Balakireva et al. 1992) and Stellate clusters (Tulin et al. 1997) was deduced earlier and might be considered as a force driving concerted evolution of these repeats. Here we have shown by direct sequencing of adjacent Su(Ste) repeats, as well as by their comparison to the orphan, that concerted evolution drives clustered Su(Ste) sequences. Nevertheless, the level of divergence of the Y-linked Su(Ste) repeats is 5–10 times as high as than that of the Stellate repeats. Interestingly, X-linked rDNA spacers are significantly more similar to each other than are Y-linked rDNA spacers (Williams et al. 1987). These differences can reflect distinct efficiencies of recombination events responsible for concerted evolution of repeats on the X and Y chromosomes. The single spontaneous events causing divergence of the Su(Ste) repeats may entail the further decrease of recombination efficiency. It was thought that polymorphism between tandemly repeated genes suppresses recombination, thereby preventing sequence homogenization of the gene family (Parniske et al. 1997). Actually, it was shown that conversion in D. melanogaster was highly sensitive to single-base mismatches within the homologous region (Nassif and Engels 1993). Thus, Stellate repeats on the X chromosome are suggested to be more susceptible to recombination events than more diverged Su(Ste) repeats on the Y chromosome. However, the differences in recombination frequencies among the X- and Y-linked repeats may be explained by unknown profound disparities of the intimate mechanisms of recombination, rather than by mere differences in repeat divergence.

Acknowledgments

We thank Y. Shevelyov for critical comments and V. E. Alatortsev for improvement of the text. We also thank two anonymous reviewers, through whose comments the manuscript was considerably improved. This work was supported by grants from the Russian Foundation of Basic Research (98-04-49107 and 99-04-48561) and the Russian program Frontiers in Genetics (99-1-069).

LITERATURE CITED


PIERRE CAPY, reviewing editor

Accepted January 5, 2000