Heterochromatic *Stellate* Gene Cluster in *Drosophila melanogaster*: Structure and Molecular Evolution

Alexei V. Tulin, Galina L. Kogan, Dominik Filipp, Maria D. Balakireva and Vladimir A. Gvozdev

Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia

Manuscript received October 9, 1996
Accepted for publication January 31, 1997

ABSTRACT

The 30-kb cluster comprising close to 20 copies of tandemly repeated *Stellate* genes was localized in the distal heterochromatin of the *X* chromosome. Of 10 sequenced genes, nine contain undamaged open reading frames with extensive similarity to protein kinase CK2 β-subunit; one gene is interrupted by an insertion. The heterochromatic array of *Stellate* repeats is divided into three regions by a 4.5-kb DNA segment of unknown origin and a retrotransposon insertion: the A region (~14 *Stellate* genes), the adjacent B region (approximately three *Stellate* genes), and the C region (about four *Stellate* genes). The sequencing of *Stellate* copies located along the discontinuous cluster revealed a complex pattern of diversification. The lowest level of divergence was detected in nearby *Stellate* repeats. The marginal copies of the A region, truncated or interrupted by an insertion, escaped homogenization and demonstrated high levels of divergence. Comparison of copies in the B and C regions, which are separated by a retrotransposon insertion, revealed a high level of diversification. These observations suggest that homogenization takes place in the *Stellate* cluster, but that inserted sequences may impede this process.

THE concept of molecular drive (concerted evolution) for tandemly repeated genes is widely accepted and discussed (DOVER 1982; HARTL and CLARK 1989; LINARES *et al.* 1994; SCHLÖTTERER and TAUTZ 1994); however, the exchange mechanisms that are involved in the turnover of repeating units remain poorly understood. Direct studies of the molecular structure of tandem repeats and polymorphisms within a gene cluster (*e.g.*, ribosomal and histones genes) are limited and currently insufficient to identify which mechanism of recombination is responsible for the homogenization of a gene cluster (DOVER and TAUTZ 1986; LINARES *et al.* 1994). One of the factors characterizing the evolution and structure of tandem repeats is the pattern of spread of variant sequences in a cluster. Studies of the extent of nucleotide sequence identity along a cluster and patterns of polymorphism with respect to interruption in an array may help to unravel the peculiarities of the evolution of that array.

Herein we describe the structure and diversification of the tandemly repeated, heterochromatic *Stellate* genes. The clustered *Stellate* genes, proposed to encode the β-subunit of a protein kinase-like protein (LIVAK 1990; PALUMBO *et al.* 1994), were first localized to the 12E region of the *X* chromosome. These genes were shown to be actively expressed in testes of *XO* males lacking a *Y* chromosome. The heterochromatic variants of *Stellate* genes had been detected earlier as components of amplified, scrambled structures built from middle repetitive DNA (SHEVELYOV 1992; NURMINSKY *et al.* 1994). The abbreviation SCLR was used to designate these composite structures comprised of shortened heterochromatic variants of *Stellate* genes, *copia*-like elements, LINEs, and rDNA fragments. The *Stellate* sequences represented <10% of the SCLR DNA. Herein we report the structure of the cluster of tandemly repeated heterochromatic *Stellate* genes and the sequences of 10 *Stellate* units. There is an inverse relationship between the efficiency of repeat homogenization and the distances between the units. Hindrances to homogenization near the sites of non-*Stellate* DNA insertions are considered. The heterochromatic *Stellate* cluster has been located in a region of distal *X*-linked heterochromatin, similar to SCLR repeats. We describe the first example of a cluster with tandemly repeated, protein-encoding units in constitutive heterochromatin.

Corresponding author: Vladimir A. Gvozdev, Institute of Molecular Genetics, Kurchatov Square, Moscow 123182, Russia.
E-mail: vgvozdev@glas.apc.org

MATERIALS AND METHODS

Drosophila strains: The *D. melanogaster* strains with deletions that were used to localize the heterochromatic DNA fragment are described elsewhere (LINDSLEY and ZIMM 1992). Interspecific *D. simulans*/*D. melanogaster* hybrid females carrying *D. melanogaster* deletions were obtained by crossing *D. simulans* (stock 6a) virgins to $In(1)sc^{4L}sc^{8R}$, $y\ sc^{4L}sc^{8R}w/\ Y^{+}\ Y$ and $Df(1)C10/\ Y^{mal+}$ males.

Standard procedures: Standard procedures for isolation of plasmid and phage DNA, subcloning into the EMBL4 and pTZ19R vectors, agarose gel electrophoresis of DNA and blotting to HybondN membranes (Amersham Corp.) were according to SAMBROOK *et al.* (1989). DNA probes were ^{32}P-labeled by oligo-priming (FEINBERG and VOGELSTEIN 1983). Restriction enzyme digestions were performed according to the manufacturer's instructions.

Genomic DNA samples: Up to 100 flies were homogenized in 1 ml of 0.1 M Tris-HCl (pH 9.1)/50 mM EDTA/0.2% Triton X-100 (TET-buffer) at 0° and centrifuged at 5000 rpm for 5 min. The pellet was resuspended in 1 ml of TET. Proteinase K (to 20 μg/ml) and sarkosyl (to 2%) were added, and after 2 hr of incubation at 37°; the homogenate was sequentially extracted with phenol, phenol/chloroform/isoamyl alcohol (50:49:1) and chloroform. DNA was precipitated from 0.2 M CH_3COONa by 2.5 vol of ethanol and stored at −25°.

Cloning of *Stellate* DNA: Clone DY8 was selected from a partial *D. melanogaster* (Batumy L) genomic DNA library, which was prepared (KOGAN *et al.* 1991) by ligating size-fractionated *Not*I digested total DNA from adult flies into the single *Not*I site of the YAC5 vector (BURKE *et al.* 1987).

*Sau*3AI partially digested and size-fractionated genomic DNA fragments of the clone DY8 were subcloned into the *Bam*HI site of the λ-EMBL4 vector and λ clones ph34, ph14, ph3 and ph1 were selected by hybridization of plaques to ^{32}P-labeled *Stellate* DNA. Recombinant phages were propagated in *Escherichia coli* K12 strain K802 and purified by CsCl-density gradient ultracentrifugation according to published procedures (SAMBROOK *et al.* 1989).

*Bgl*II DNA fragments containing *Stellate* repeats from the λ clones were subcloned into pTZ19R. Plasmid DNA was propagated in *E. coli* K12 strain TG1 or XL1 blue strains. Cells were made competent as described (HANAHAN 1983).

DNA sequencing: All *Stellate*-carrying plasmid subclones were sequenced by Sanger's procedures (SAMBROOK *et al.* 1989) using Sequenase version 2.0 T7 DNA polymerase (U.S. Biochemical) on both strands from universal plasmid primers or *Stellate*-specific primers. The analysis of DNA sequences was performed using the DNASIS 5.00 and PCGENE 3.30 programs.

RESULTS

The *Stellate* repeat cluster is localized in the distal portion of the *X*-linked heterochromatin: Southern hybridization shows that the DY8 clone comprises the major fraction of the *X*-linked heterochromatic *Ste* repeats represented by 1150-bp *Cfo*I fragments (SHEVELYOV 1992). The *Y*-linked, 800-bp, cross-hybridizable *Su(Ste)* repeats (BALAKIREVA *et al.* 1992) and the 950-bp *Cfo*I euchromatic repeats (LIVAK 1990) are absent from this clone (Figure 1a).

The pattern of genomic *Eco*RI fragments hybridized to the *Ste* probe is shown in Figure 1b. The 1.3- and 0.8-kb *Eco*RI fragments that originate from SCLR segments are represented by the heterochromatic *Stellate* unit interrupted by an *MDG*1 retrotransposon insertion (NURMINSKY *et al.* 1994). These fragments were detected in all the stocks studied. A comparison of *Ste*-positive *Eco*RI genomic fragments larger than 5 kb revealed a high level of restriction fragment length polymorphism among the stocks (Figure 1). The size of the largest *Stellate* containing *Eco*RI fragment in the DY8 clone is ~35 kb. As is typical for heterochromatic parts of the *D. melanogaster* genome, the largest genomic fragments, as well as the 0.8-, 1.3-, 2.8- and 5-kb ones, were underrepresented in salivary gland DNA. The 9- and 3.2-kb *Stellate* fragments may represent the repeats located in the euchromatic 12E polytene region.

Hybridization of an *Ste* probe to *Xba*I digests of genomic DNA reveals two bands represented by the regular euchromatic (1250 bp) and shortened heterochromatic (1150 bp) units, as well as a variable set of genomic polymorphic *Ste*-carrying fragments (Figure 1c). The DY8 recombinant clone carries the 1150-bp heterochromatic units. The relative content of heterochromatic and euchromatic regular *Ste* repeats varies among stocks. The Batumi L genome preferentially comprises 1150-bp heterochromatic repeats, whereas the sc^4 stock has a substantial contribution from both variants of *Stellate* repeats. Heterochromatic *Ste* repeats were recently localized by *in situ* hybridization to the boundary between the C and D heterochromatic blocks in mitotic chromosomes (PALUMBO *et al.* 1994). To locate the cloned *Ste* cluster, we used interspecific *D. simulans*/*D. melanogaster* hybrids carrying suitable *D. melanogaster* deletions. This approach is based on observations showing the absence of peculiar *D. melanogaster* repeats in the *D. simulans* genome (DOWSETT 1983). The *D. simulans* genome carries no *Stellate* repeats in the *X* chromosome and autosomes (LIVAK 1984). The $Df(1)sc^{4L}sc^{8R}$ heterochromatic deletion eliminates the region between the C and D blocks (LINDSLEY and ZIMM 1992), as well as the *Ste*-hybridizable, *Xba*I-polymorphic genomic fragments; however, the regular *Ste* units are retained (Figure 1c). Thus, the nonregular, repeated *Stellate* copies detected as polymorphic *Xba*I fragments are located in the region proximal to the sc^4 breakpoint (Figure 1c). The interspecific hybrids between *D. simulans* females and *D. melanogaster* males carrying *Df(1)C10* (Figure 1c) are devoid of *Xba*I-generated 1150-bp *Stellate* units and other *Stellate* units bearing heterochromatic *Xba*I fragments of variable size. The cloned *Stellate* cluster may be assigned to cytological region 26h, distal to the D-C boundary (Figure 1, c and d). These observations are in conflict with earlier published data (PALUMBO *et al.* 1994), but may be explained by interstock variation of *Stellate* repeat locations within the distal heterochromatin. We conclude that most of the *Stellate* heterochromatic repeats are located distal to the nucleolus organizer (Figure 1d).

The heterochromatic, tandemly repeated array of *Stellate* genes is discontinuous The 90-kb DY8 fragment was subcloned into the λEMBL4 vector. The resulting cloned fragments (ph34, 14, 3 and 1, Figure 2) allow us to generate a restriction map of the studied region. The organization of *Stellate* genes in tandem arrays was demonstrated using partial *Bgl*II digestion of the DY8 segments, taking into account that a single *Bgl*II recognition site is located within the *Stellate* unit (Figure 2). A Southern blot of *Bgl*II-digested λ clone ph34, probed with a *Stellate Bgl*II fragment, revealed the characteristic 1150-bp "ladder" (Figure 2a) derived from a cluster bearing seven *Stellate* repeats. The adjacent (see below) ph14 fragment includes four different *Stellate* repeats that were subcloned and sequenced. Partial *Bgl*II digestion of the whole DY8 clone enables us to estimate that there are at least 12 repeats organized in a regular, uninterrupted array (data not shown).

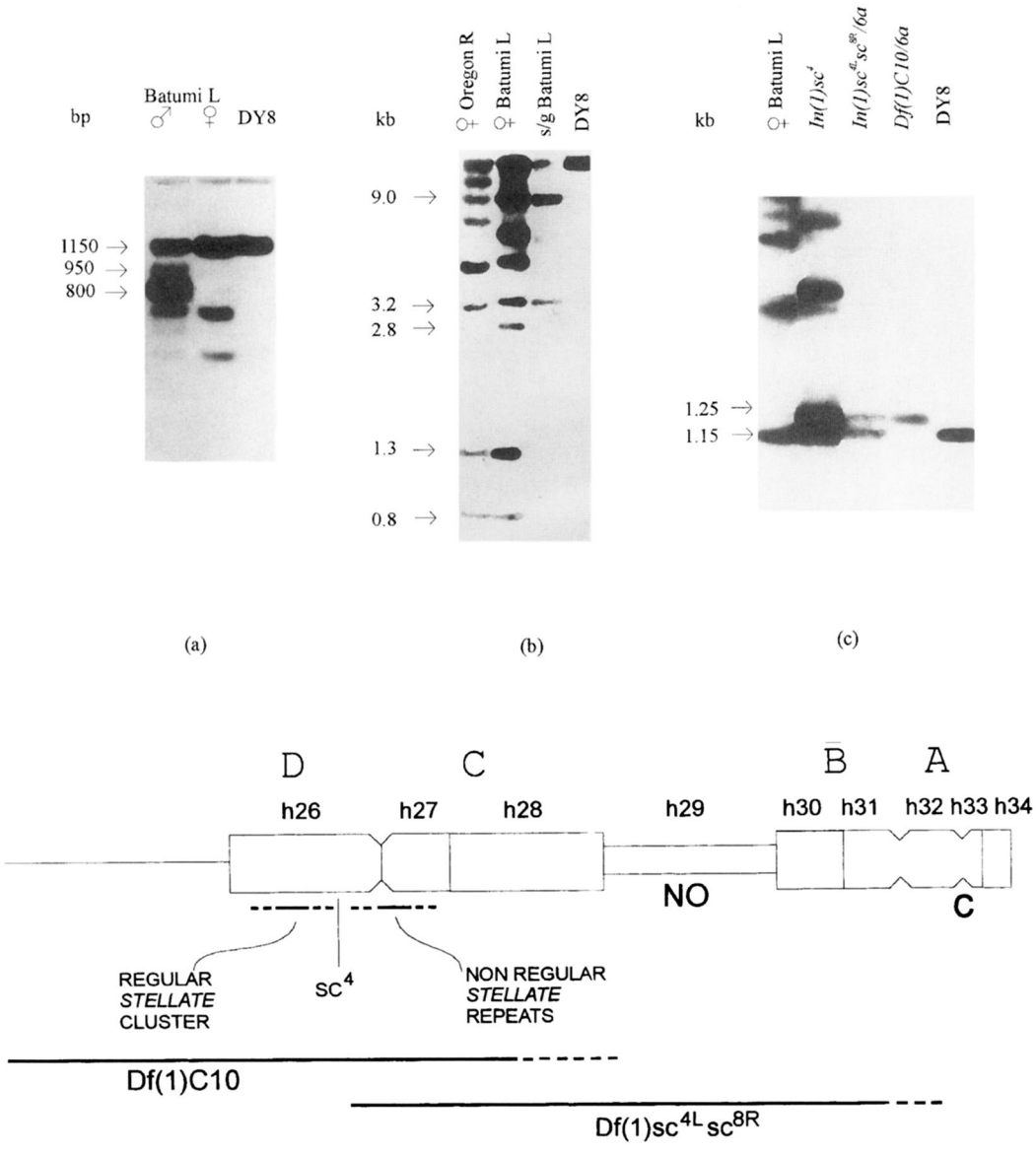


FIGURE 1.—Genomic polymorphism and localization of heterochromatic *Stellate* repeats (a–c). Southern blot analysis of genomic DNA samples and DY8 clone; hybridization with the *Ste* probe. Endonuclease digestion as follows: (a) *Cfo*I, DNA from Batumi males, Batumi females and DY8 clone; (b) *Eco*RI, female DNA from Oregon, Batumi, Batumi salivary gland and DY8; (c) *Xba*I, female DNA from Batumi, *In(1)sc*^{4}, hybrid *Df(1)sc*^{4L}*sc*^{8R}*/6a* and *Df-(1)C10/6a*, and DY8 (6a is *D. simulans*); (d) location of the *Stellate* cluster in distal *X*-linked heterochromatin. NO, nucleolus organizer. The locations of heterochromatic blocks (A–D) according to COOPER (1959) and the h26–h33 segments according to GATTI and PIMPINELLI (see LINDSLEY and ZIMM 1992) are indicated.

The right part of the cluster was reconstructed using overlapping λ phage fragments. The left part of the cluster is represented by the ph34 and ph14 phages. The left edge of the ph34 fragment is represented by a truncated *Stellate* copy (34_2) separated from an 18S *r*DNA fragment by 286 bp of anonymous DNA. It was shown, using specific probes and Southern analysis, that the 19-kb *Pst*I fragment (Figure 2) encompasses the ph34 and ph14 clones (data not shown). The ph34 and ph14 clones are presented as adjacent segments of the *Stellate* cluster. It is unlikely that non-*Stellate* DNA is located between the ph34 and ph14 fragments (Figure 2). Using the 19-kb *Pst*I fragment as a template and two different pairs of *Ste*-specific primers, only *n*-mers (mono-, di- and trimers) of *Stellate* units were revealed during PCR amplification. We failed to detect other PCR products. *Stellate* repeats in the 19-kb *Pst*I fragment extend to 17 kb, taking into account that the edges of the 19-kb *Pst*I fragment comprise 2.2 kb of anonymous non-*Stellate* sequences. The length of the *Stellate* carrying region therefore extends to 17 kb and corresponds to 14 undamaged *Ste* repeats.

The *Stellate* cluster was shown to be discontinuous, being interrupted by a 4.5-kb fragment and a retrotransposon insertion, resulting in target site duplication, in an intergenic spacer (Figure 2). The retrotransposon was identified by sequencing as a GATE element with the same site of insertion previously detected in SCLR structures (NURMINSKY *et al.* 1994). Sequencing of the 4.5-kb insertion and flanking DNA revealed no similarity in EMBL and SwissProt databases, except a fragment of 18S *r*DNA on the left edge of the cluster.

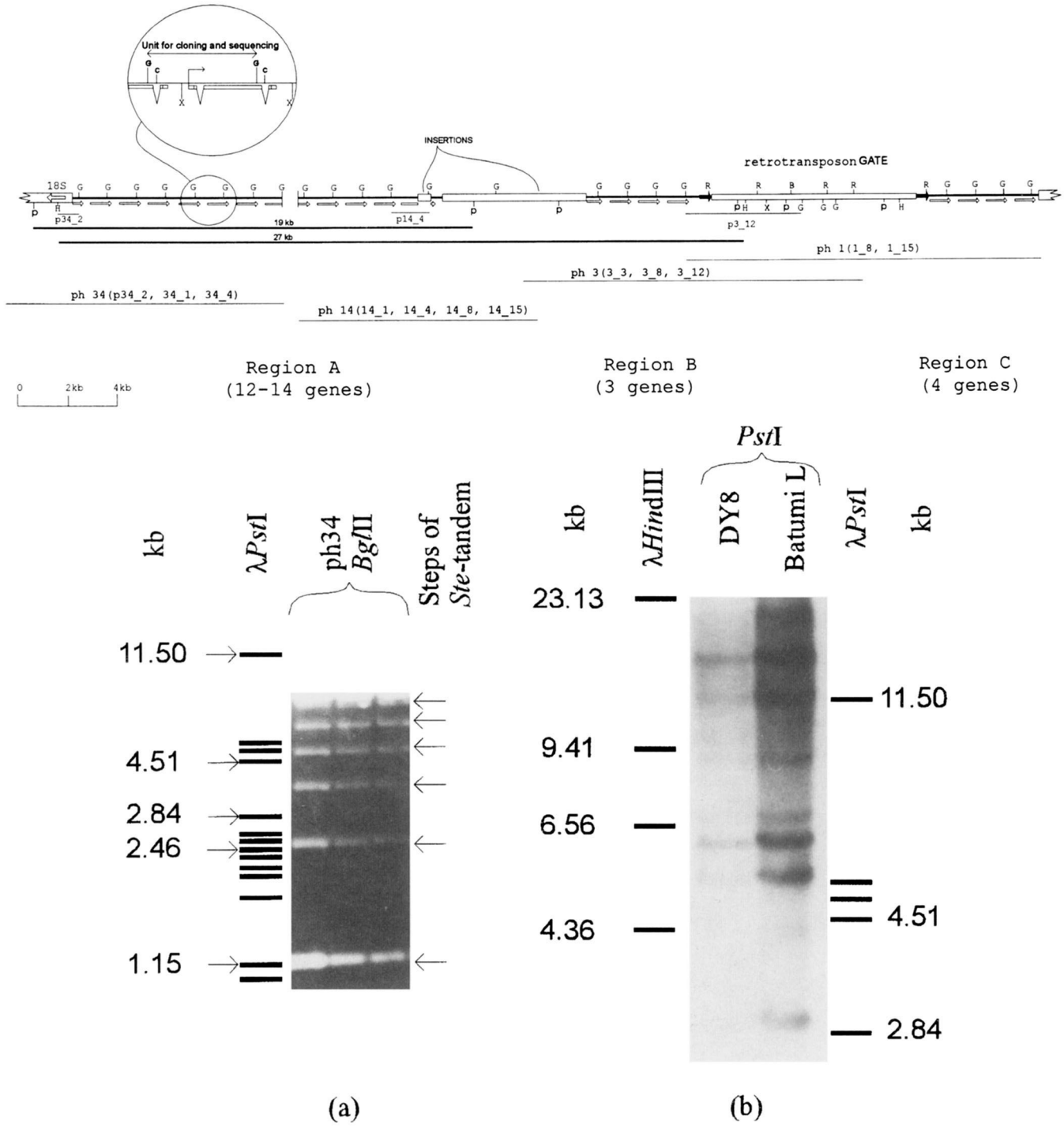


FIGURE 2.—*Top:* Structure of discontinuous *Stellate* cluster comprising 30 kb of the DY8-YAC clone. G, *Bgl*II; R, *Eco*RI; X, *Xba*I; H, *Hin*dIII; P, *Pst*I; B, *Bam*HI. Arrows indicate *Stellate* units; the sizes of λ phages (ph) and plasmid (p) subclones are presented below. The structure of the *Stellate* units is presented in the circle; open regions of the rectangles indicate the noncoding region; the boundaries of cloned and sequenced *Stellate* units are shown. In the main map the open rectangles designate a small insertion in the 14_4 *Stellate* units and a 4.5-kb sequence separating the A and B clusters. Black arrows indicate the LTRs of the inserted retrotransposon. Shaded arrow shows the location of the 18S *r*DNA fragment. *Bottom:* (a) Partial *Bgl*II digestion of the ph34 clone; arrows designate 1.15-kb restriction fragments ladder, ethidium bromide staining. At least seven *Stellate* copies are detected accounting for six steps in the ladder and a vector attached sequenced copy carrying a damaged *Bgl*II site. Similar data concerning the whole DY8 clone and other phage clones are not shown. (b) Detection of 19-kb *Pst*I fragments in the DY8 clone and Batumi L genomic DNA.

The cloned region with the 4.5-kb insertion reflects the genomic organization of the *Stellate* cluster since the large 19-kb *Pst*I fragment (Figure 2) was also detected in genomic DNA (Figure 2b).

Estimation of *Stellate* sequence divergence: The 1150-kb *Bgl*II *Stellate* repeats were subcloned into pTZ19R for sequencing. The resulting fragments could be ascribed with confidence to different repeats if nu-

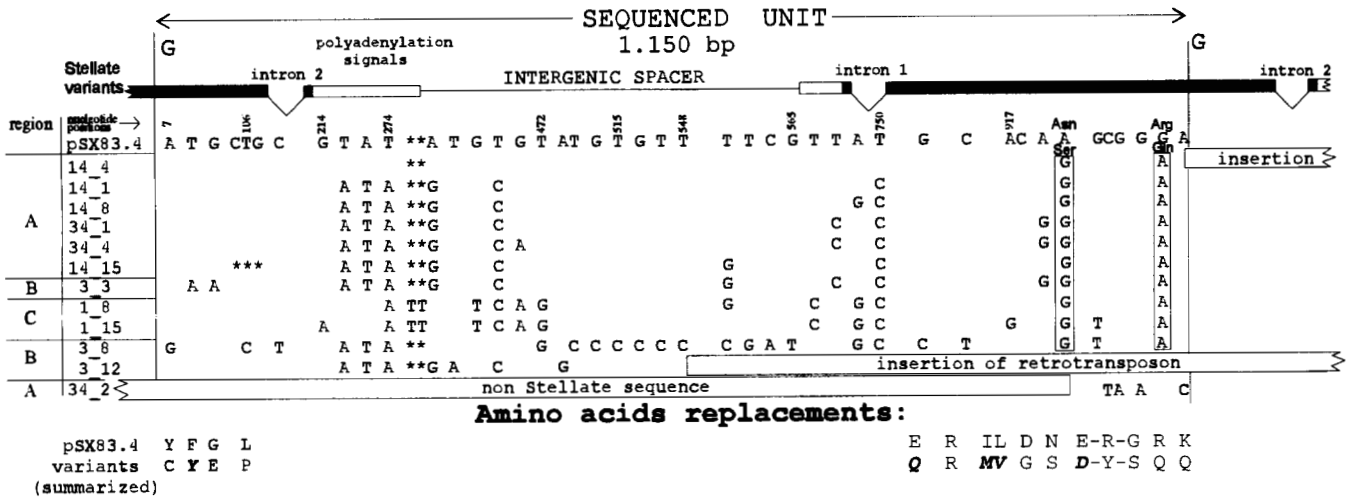


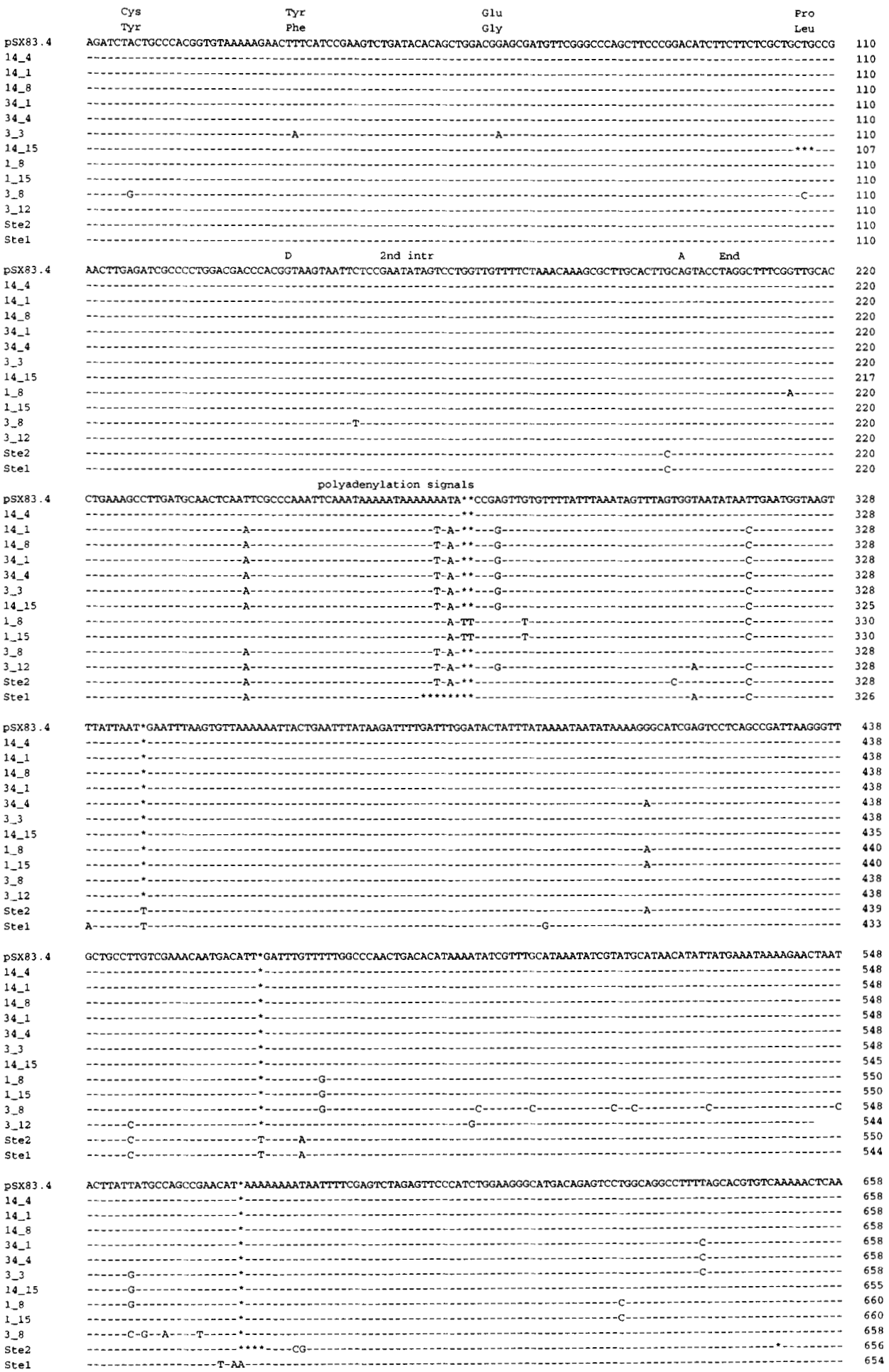
FIGURE 3.—Sequence variations of *Stellate* units. Nucleotide substitutions and amino acid replacements are presented in permuted *Stellate* sequences compatible with the sequencing strategy. Nucleotide positions along the sequence are indicated (out of scale). The *Stellate* sequences are compared to the heterochromatic-like pSX83.4 variant (LIVAK 1990); only those nucleotide positions that were shown to be changed in *Stellate* variants are indicated. * denote deletions. ■, coding regions; □, noncoding regions. The shared nucleotide substitutions resulting in amino acid replacements are boxed. Substitutions resulting in the replacements within the same amino acid exchange group are designated in bold. G, *Bgl*II.

cleotide substitutions (or deletions/insertions) were identified between copies. Four sequenced copies were identical to those previously studied; two from ph34 were identical to the 34_4 copy, one from ph14 was identical to the 14_8 copy, and one from ph3 was identical to the 3_3 copy. The 300–500-bp fragments of partially sequenced copies displaying no differences in nucleotide sequence compared to previously examined repeats were excluded from analysis (five cases for ph34, three cases for ph14 and a single case for ph3). In the case of ph34 we have identified with confidence only two different copies among the seven to eight copies situated in the ph34 clone. We were unable to discriminate the results of recurrent cloning of the same copy from the situation of intrinsic homogeneity of *Stellate* copies in the region A. Thus, the extent of *Stellate* copy identity (homogenization) may be underestimated. Figures 3 and 4 demonstrate divergence between sequenced copies. The sequences were aligned to the previously sequenced *Stellate* variant, pSX83.4 (LIVAK 1990), which represents a shortened heterochromatic repeat, similar to the *Stellate* variant in the SCLR structures (SHEVELYOV 1992). We identified in the A region two nonidentical copies from ph34 and four from ph14. The third ph34 *Stellate* unit (34_2), situated on the left edge of the A region, was shown to be truncated (Figures 3 and 4). Three *Stellate* repeats from the ph3 clone (two undamaged copies and one carrying the retrotransposon insertion) and two copies isolated from the ph1 clone are from the B and C regions, respectively.

Figures 3 and 4 present a compilation of sequence divergences of *Stellate* copies. The sequences of two *Stellate* copies (*Ste*1 and *Ste*2) related to heterochromatic SCLR structures (SHEVELYOV 1992) are also included in Figure 4. A permuted order of sequences is presented due to the cloning and sequencing strategy.

Comparison of five undamaged *Stellate* copies in the A region (two from ph34 and three from ph14) with the pSX83.4 copy revealed six shared nucleotide substitutions in noncoding regions, including the site of the polyadenylation signal, and two substitutions resulting in identical amino acid replacements, Asn to Ser and Arg to Gln (Figures 3 and 4). The majority of these nucleotide positions may be considered as diagnostic for the copies in the cloned *Stellate* cluster. In contrast, the sequence of the marginal 14_4 copy of the A region (interrupted by the short insertion) is much more similar to the pSX83.4 copy, whose genomic location is not known. These similarities are seen in the region of the polyadenylation signal and at positions 316 and 750 in noncoding regions. These observations indicate that at least five copies in the A region were subjected to a homogenization process, but the marginal, interrupted 14_4 and 34_2 copies escaped it. The fragment of the marginal and truncated 34_2 copy contains an Arg codon (position 112, Figure 4), instead of a Gln codon, which is diagnostic of all the other copies in the cloned *Stellate* cluster. The other three substitutions are in the coding region, resulting in amino acid replacements. No such changes in the homologous *Stellate* fragment were detected in the aforementioned five copies in the A region. Thus, both marginal copies, truncated or damaged by insertion, have escaped the putative homogenization process.

Three *Stellate* copies from the B region (ph3 clone) share cluster-specific nucleotide sequences in the region of the polyadenylation signal, as well as in the coding region (diagnostic Ser and Gln codons). The 3_8 copy contains 20 substitutions, including a set of


Cys Tyr Tyr Phe Glu Gly Pro Leu
pSX83.4 AGATCTACTGCCCACGGTGTAAAAAGAACTTTCATCCGAAGTCTGATACACAGCTGGACGGAGCGATGTTCGGGCCCAGCTTCCCGGACATCTTCTTCTCGCTGCTGCCG 110
14_4 110
14_1 110
14_8 110
34_1 110
34_4 110
3_3 110
14_15 107
1_8 110
1_15 110
3_8 110
3_12 110
Ste2 110
Ste1 110
D 2nd intr A End
pSX83.4 AACTTGAGATCGCCCCTGGACGACCCACGGTAAGTAATTCTCCGAATATAGTCCTGGTTGTTTTCTAAACAAAGCGCTTGCACTTGCAGTACCTAGGCTTTCGGTTGCAC 220
14_4 220
14_1 220
14_8 220
34_1 220
34_4 220
3_3 220
14_15 217
1_8 220
1_15 220
3_8 220
3_12 220
Ste2 220
Ste1 220
polyadenylation signals
pSX83.4 CTGAAAGCCTTGATGCAACTCAATTCGCCCAAATTCAAATAAAAATAAAAAAATA**CCGAGTTGTGTTTTATTTAAATAGTTTAGTGGTAATATAATTGAATGGTAAGT 328
14_4 328
14_1 328
14_8 328
34_1 328
34_4 328
3_3 328
14_15 325
1_8 330
1_15 330
3_8 328
3_12 328
Ste2 328
Ste1 326
pSX83.4 TTATTAAT*GAATTTAAGTGTTAAAAAATTACTGAATTTATAAGATTTTGATTTGGATACTATTTATAAAATAATATAAAAGGGCATCGAGTCCTCAGCCGATTAAGGGTT 438
14_4 438
14_1 438
14_8 438
34_1 438
34_4 438
3_3 438
14_15 435
1_8 440
1_15 440
3_8 438
3_12 438
Ste2 439
Ste1 433
pSX83.4 GCTGCCTTGTCGAAACAATGACATT*GATTTGTTTTTGGCCCAACTGACACATAAAATATCGTTTGCATAAATATCGTATGCATAACATATTATGAAATAAAAGAACTAAT 548
14_4 548
14_1 548
14_8 548
34_1 548
34_4 548
3_3 548
14_15 545
1_8 550
1_15 550
3_8 548
3_12 544
Ste2 550
Ste1 544
pSX83.4 ACTTATTATGCCAGCCGAACAT*AAAAAAAATAATTTTCGAGTCTAGAGTTCCCATCTGGAAGGGCATGACAGAGTCCTGGCAGGCCTTTTAGCACGTGTCAAAAACTCAA 658
14_4 658
14_1 658
14_8 658
34_1 658
34_4 658
3_3 658
14_15 655
1_8 660
1_15 660
3_8 658
Ste2 656
Ste1 654

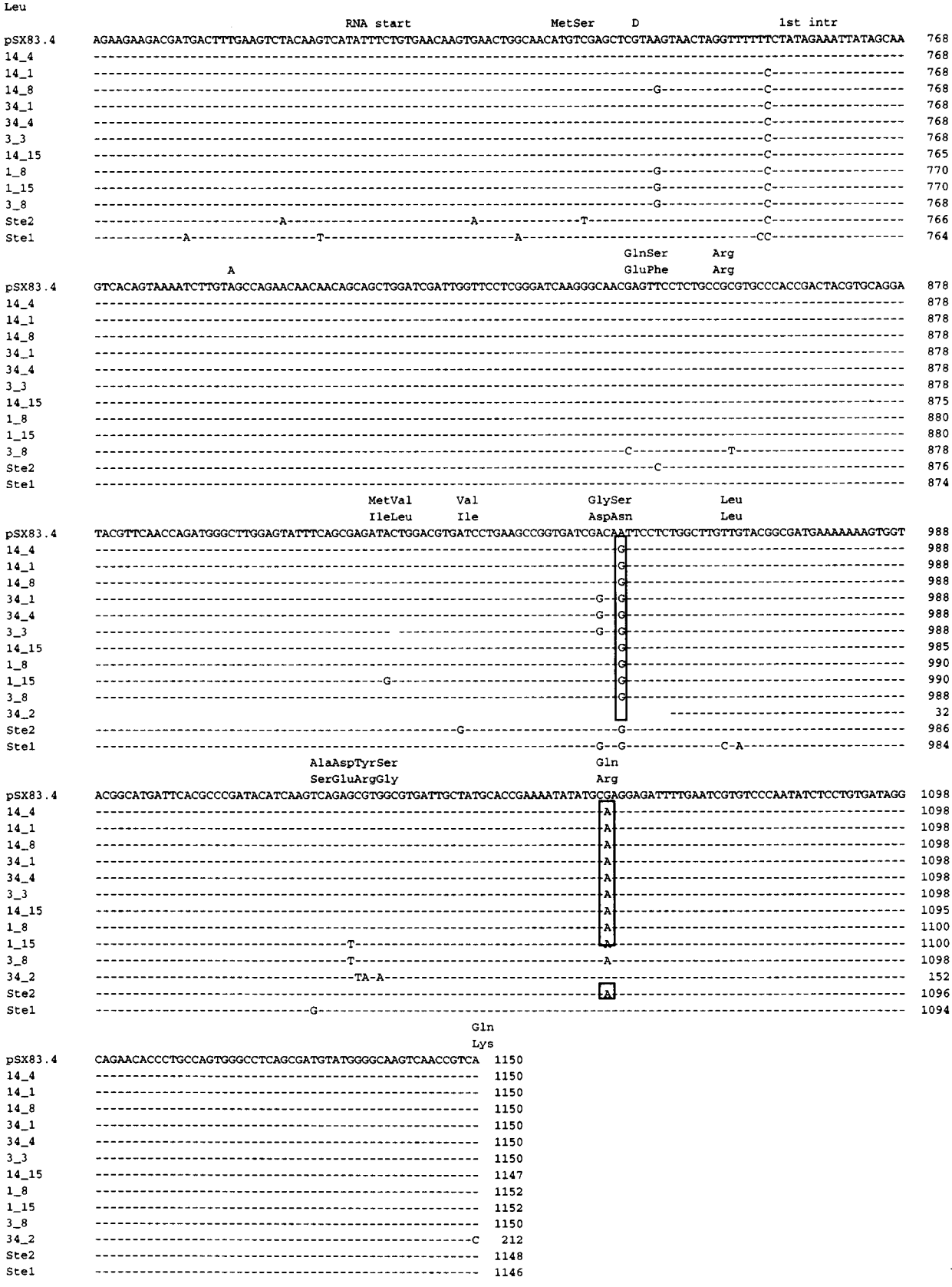


FIGURE 4.—Compilation of heterochromatic *Stellate* sequences. Because of the cloning and sequencing strategy (Figure 2), we present a permuted sequences of the *Stellate* genes starting from the *Bgl*II site. The splice donor (D) and acceptor sites (A) are according to LIVAK (1990). The pSX83.4 and *Ste1*/*Ste2* sequences are from LIVAK (1989) and SHEVELYOV (1992), respectively. A hyphen means that the base is the same as in pSX83.4. Amino acid replacements are indicated above the nucleotide sequences. The nucleotide substitutions peculiar to euchromatic *Stellate* sequences are boxed. The sequences have been deposited in EMBLBANK under the following accession numbers: 1_15: X97132; 1_8: X97133; 14_1: 97134; 14_15: X97135; 14_4: X97136; 14_8: X97138; 3_12: X97139; 3_3: X97140; 3_8: X97141; 34_1: X97142; 34_2: X97143; and 34_4: X97144.

TABLE 1

Estimation of divergence between *Stellate* copies by pairwise comparison

	A region							B region			C region	
	34_1	34_4	34_2[a]	14_1	14_8	14_15	14_4[a]	3_8	3_3	3_12[b]	1_8	1_15
34_1		1	4/212	1	2	3	7	18	3	2/544	10	12
34_4			4/212	2	3	4	8	19	4	2/544	11	13
34_2[a]				4/212	4/212	4/212	4/212	6/212	4/212		4/212	6/212
14_1					1	2	6	17	4	2/544	9	11
14_8						3	7	16	5	2/544	8	10
14_15							8	19	4	3/544	9	13
14_4[a]								19	10	7/544	9	11
3_8									21	7/544	20	20
3_3										4/544	11	15
3_12[b]											7/544	8/544
1_8												4

The numbers of detected nucleotide substitutions, deletions and insertions are indicated per unit (1150 bp); substitutions per 200 and 544 bp length are indicated for the truncated 34_2 and 3_12 copies, respectively.

[a] Copies flanking the A region.

[b] Truncated copy due to retrotransposon insertion in spacer region.

11 that are tightly clustered along a length of 100 bp in the spacer. The significance of the clustering was demonstrated using appropriate statistical treatment (STEPHENS 1985) useful for detecting gene conversion. The extremely low value of the probability of random clustering ($P = 1.78 \times 10^{-6}$) corroborates the supposed role of gene conversion in the origin of the detected substitutions.

Two copies in the C region, localized to the right of the retrotransposon insertion, share a TT insertion and three nucleotide substitutions in the intergenic region. Thus, homogenization between these sequences and the nearest 3_12 truncated copy (separated by a retrotransposon insertion) may be impeded.

Table 1 presents the results of pairwise sequence comparisons of *Stellate* copies where the number of differences per units are indicated. The divergence between the internally located copies (related to fragments cloned in ph34 or ph14) does not exceed a value of one to three differences per repeat (1150 bp). The interrupted, marginal 14_4 copy shows a markedly higher extent of diversity (six to eight differences). The 34_2 marginal copy (Figures 2 and 3), localized on the left boundary of the cluster, contains four substitutions per 212 bp; similar nucleotide stretches were shown to be identical in the other five, neighboring copies (related to the ph34 and ph14 clones). Thus, higher levels of total polymorphism, as well as loss or gain of diagnostic nucleotide positions, were revealed for the repeats located on the edges of the A region. It is worth emphasizing that pairwise comparison of the adjacent *Ste*1 and *Ste*2 copies (see Figure 4), one of which is interrupted by a retrotransposon insertion (NURMINSKY *et al.* 1994), allows us to detect 28 sites of nucleotide sequence polymorphism. The divergence between the 3_3 internal copy in the B region and the A region-related copies seems to be similar to the value of divergence (up to four substitutions per unit) detected between internal copies of the A region (Table 1). Sequencing of half of the *Stellate* 3_12 unit (truncated by the retrotransposon insertion) revealed two nucleotide substitutions when compared to the ph34-related copies. This value of divergence is roughly equivalent to four substitutions per unit, similar to that for internal units of the A region. The same value of four substitutions is found for the C-region copies (1_15 and 1_8).

DISCUSSION

In this paper we describe the structure and molecular diversity of the *Stellate* gene tandem array that encodes a protein with extensive similarity to the β subunit of protein kinase CK2. This cluster was shown to be located in the distal part of the *X*-linked heterochromatin. This is the first example of tandemly repeated, protein-encoding genes with intact open reading frames (ORFs) in a region of constitutive heterochromatin in eukaryotes. We did not detect any random damage to the *Stellate* coding regions, apart from a case of truncation of one copy (34_2). The probability that the heterochromatic *Stellate* repeats are expressed was previously discussed (SHEVELYOV 1992; PALUMBO *et al.* 1994).

The presence of intact ORFs in heterochromatic *Ste* repeats indicates that these units evolved under selective pressure at the translational level. Furthermore, the codon usage bias for the heterochromatic *Stellate* repeats [as well as for a single sequenced euchromatic copy (LIVAK 1990)] is typical of that for highly expressed genes in Drosophila (MORIYAMA and HARTL 1993). In synonymous sites, the A and C/G content in the third position of biased *Stellate* genes was <11% and >40%, respectively. Thus, a strong negative selective constraint against A and in favor of C at synonymous

sites was shown. The degree of bias is indicated by the "scaled chi-squared" and this value is >0.5 for highly biased genes (LI 1987). This value amounts to 0.59 for the *Stellate* genes. However, the *Stellate* cluster looks like a dispensable genomic structure. The sibling species, *D. simulans*, carries no noticeable number of *Stellate* repeats and their benefit to *D. melanogaster* is unclear. Nevertheless, a high codon bias has been observed for several undamaged heterochromatic repeats.

Our observation of molecular similarity among the *Stellate* repeats demonstrates the existence of homogenization (concerted evolution) of the copies in the discontinuous *Stellate* cluster. This conclusion is corroborated by the detected similarities in coding regions and in polyadenylation signal sequences as compared to the pSX83.4 copy that is located elsewhere in heterochromatin. A possibility of homogenization as a result of recombination events is suggested by the presence of the 14_4 copy. This copy preserves several diagnostic nucleotide positions (polyadenylation signal and noncoding regions) similar to pSX83.4, but at the same time contains some positions in the coding region that result in amino acid replacements peculiar to other copies in the studied cluster. Thus, the 14_4 copy may be considered as a recombinant between two diverged *Stellate* copies. Homogenization of this copy may by retarded due to its location on the edge of the A region of a cluster in proximity to an insertion, or due to a short insertion in the coding region of this copy.

Shared polymorphism is more common between adjacent copies than between remote units in this tandem array. Thus, an A → G transition, resulting in an Asp → Gly substitution, is characteristic of both copies carried by the ph34 clone, although these markers are absent in the *Ste* copies from the adjacent ph14 clone. The T → C substitution in the noncoding region of the first exon characterize ph34 related copies but not four other copies in the right half of the A region. Similarly, the TT insertion, as well as other substitutions in the intergenic spacer, are peculiar to both adjacent copies of the ph1 subclone. These observations corroborate findings based on studies of DNA restriction fragment length polymorphism patterns of repeated units in the histone array (COLBY and WILLIAMS 1993), which showed that the identity between repeats decreases with distance.

The *Stellate* cluster is represented by the A–C regions defined by 4.5-kb insertion of unknown origin and a retrotransposon insertion. The marginal copies of the A region, one of which is truncated (34_2) and another that carries an insertion (14_4), were shown to be highly diverged. This observation may be explained by the decreased frequencies of unequal crossing over on the edges of the array or by the alteration of gene conversion events due to the effects of flanking sequences. The high degree of divergence between copies of the B region separated by the retrotransposon (the ph3 and ph1) may be maintained by the insertion acting as an impediment of homogenization. The effects of the flanking sequences on the unexpectedly complex conversion-like events affecting human minisatellites were considered, and it was suggested that the observed patterns of minisatellite polymorphism are controlled by means of elements outside the tandem repeat array (JEFFREYS *et al.* 1994). The patchy nucleotide substitutions in the case of the 3_8 copy are easier to explain by targeted gene conversion between diverged *Ste* copies. The presented data give no unambiguous evidence concerning the mechanisms that homogenize tandem repeats; however, they emphasize that the peculiarities of cluster organization may affect the pattern of molecular evolution.

A high level of codon usage bias in *D. melanogaster* was shown to correlate positively with high rates of recombination in the regions of the affected genes (KLIMAN and HAY 1993). This observation was corroborated by numerous examples. However, heterochromatic *Stellate* repeats are located in a region of distal heterochromatin that is characterized by drastically decreased levels of recombination (ASHBURNER 1989). One may propose the possibility that heterochromatic *Stellate* gene evolution occurred in euchromatin before a transposition to heterochromatin where recombination is highly suppressed. At the same time, specific recombination events (unequal crossing over and gene conversion), hypothesized to be involved in the homogenization of repeats, may generate their separate diversification in heterochromatin. A step of diversification may be followed by selection and a subsequent step of homogenization. The question of how selection can act remains a common and unresolved problem concerning the evolution of tandem repeats. Recombination may be attained by unequal crossing over between sister chromatids that was demonstrated for ribosomal genes situated in the *X*-linked heterochromatin of *D. melanogaster* (ENDOW and ATWOOD 1988). The events of unequal crossing over were also detected for the *X*-linked heterochromatic SCLR repeated structures composed of *Stellate* sequences and mobile genetic elements (NURMINSKY *et al.* 1994). We propose that effective recombination between heterochromatic parts of sister chromatids may be a mechanism of diversification of coding regions of repeats, launching further selective forces to generate the observed level of codon bias.

The authors thank YU. YA. SHEVELYOV and I. A. KRAMEROVA for critical reading of manuscript, V. E. ALATORTSEV for procedure of genomic DNA isolation and V. M. ANDREEV for his help in statistical treatment. We also thank anonymous referees for helpful comments on this manuscript. This research was supported by the grants from Russian Foundation of Basic Researches (93-04-07941 and 96-04-49026), Russian Program Frontiers in Genetics and International Science Foundation (M93000, M93300) to V.A.G.

LITERATURE CITED

ASBURNER, M., 1989 *Drosophila: A Laboratory Handbook.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

BALAKIREVA, M. D., Y. Y. SHEVELYOV, D. I. NURMINSKY, K. J. LIVAK and V. A. GVOZDEV, 1992 Structural organization and diversification of Y-linked sequences comprising *Su(Ste)* genes in *Drosophila melanogaster.* Nucleic Acids Res. **20:** 3731–3736.

BURKE, D. T., G. F. CARLE and M. V. OLSON, 1987 Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science **236:** 806–812.

COLBY, C., and S. M. WILLIAMS, 1993 The distribution and spreading of rare variants in the histone multigene family. Genetics **135:** 127–133.

COOPER, K. W., 1959 Cytogenetic analysis of major heterochromatic elements (especially Xh and Y) in *Drosophila melanogaster* and the theory of "heterochromatin." Chromosoma **10:** 535–88.

DOVER, G. A., 1982 Molecular drive: a cohesive mode of species evolution. Nature **299:** 111–117.

DOVER, G. A., and D. TAUTZ, 1986 Conservation and divergence in multigene families: alternatives to selection and drift. Phil. Trans. R. Soc. Lond. **B 312:** 275–289.

DOWSETT, A. P., 1983 Closely related species of *Drosophila* can contain different libraries of middle repetitive DNA sequences. Chromosoma **88:** 104–108.

ENDOW, S. A., and K. C. ATWOOD, 1988 Magnification: gene amplification by an inducible system of sister chromatid exchange. Trends Genet. **4:** 348–351.

FEINBERG, A. P., and B. VOGELSTEIN, 1983 A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6–13.

HANAHAN, D., 1983 Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:** 557–580.

HARTL, D. L., and A. G. CLARK, 1989 *Principles of Population Genetics,* Ed. 2. Sinauer Associates, Sunderland, MA.

JEFFREYS, A. J., K. TAMAKI, A. MACLEOD, D. G. MONCTON, D. NEIL *et al.,* 1994 Complex gene conversion events in germline mutation at human minisatellites. Nat. Genet. **6:** 136–145.

KLIMAN, R. M., and J. HEY, 1994 The effects of mutation and natural selection on codon bias in the genes of *Drosophila.* Genetics **137:** 1049–1056.

KOGAN, G. L., D. FILIPP, I. P. ARMAN, B. A. LEIBOVICH and E. S. BELYAEVA, 1991 Cloning of *Drosophila melanogaster* genome fragments in *Saccharomyces cerevisiae.* Genetika (Russ) **27:** 1316–1323.

LI, W.-H., 1987 Model of nearly neutral mutations with particular implications for nonrandom usage of synonymous codon. J. Mol. Evol. **24:** 337–345.

LINARES, A. R., T. BOWEN and G. A. DOVER, 1994 Aspects of nonrandom turnover involved in the concerted evolution of intergenic spacers within the ribosomal DNA in *Drosophila melanogaster.* J. Mol. Evol. **39:** 151–159.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster.* Academic Press, New York.

LIVAK, K. J., 1984 Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. Genetics **107:** 611–634.

LIVAK, K. J., 1990 Detailed structure of the *Drosophila melanogaster Stellate* genes and their transcripts. Genetics **124:** 303–316.

MORIYAMA, E. N., and D. L. HARTL, 1993 Codon usage bias and base composition of nuclear genes in *Drosophila.* Genetics **134:** 847–858.

NURMINSKY, D. I., Y. Y. SHEVELYOV, S. V. NUZHDIN and V. A. GVOZDEV, 1994 Structure, molecular evolution and maintenance of extended repeated structures in the *X* heterochromatin of *Drosophila melanogaster.* Chromosoma **103:** 277–285.

PALUMBO, G., S. BONACCORSI, L. G. ROBBINS and S. PIMPINELLI, 1994 Genetic analysis of *Stellate* element of *Drosophila melanogaster.* Genetics **138:** 1181–1197.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual,* Ed. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

SCHLÖTTERER, C., and D. TAUTZ, 1994 Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromasomal exchanges drive concerted evolution. Curr. Biol. **4:** 777–783.

SHEVELYOV, Y. Y., 1992 Copies of a *Stellate* gene variant are located in the *X* heterochromatin of *Drosophila melanogaster* and are probably expressed. Genetics **132:** 1033–1037.

STEPHENS, J. C., 1985 Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. Mol. Biol. Evol. **2:** 539–556.

Communicating editor: M. J. SIMMONS