Nucleic Acids Research, Vol. 20, No. 5 1117–1122

Characterization of a low copy repetitive element S232 involved in the generation of frequent deletions of the distal short arm of the human X chromosome

Xiao-Miao Li[1], Pauline H.Yen[1,2] and Larry J.Shapiro[1,2,3]*
[1]Howard Hughes Medical Institute, Departments of [2]Pediatrics and [3]Biological Chemistry, UCLA School of Medicine, Harbor-UCLA Medical Center, Torrance, CA 90509, USA

Received October 28, 1991; Revised and Accepted January 28, 1992

ABSTRACT

There are several copies of related sequences on the distal short arm of the human X chromosome and the proximal long arm of the Y chromosome which were originally detected by cross hybridization with a genomic DNA clone, CRI-S232. Recombination between two S232-like sequences flanking the steroid sulfatase locus has been shown to cause frequent deletions in the X chromosome short arm, resulting in steroid sulfatase deficiency. We now report the characterization of several S232-like sequences. Restriction mapping and sequence analysis show that each S232 unit contains 5 kb of unique sequence in addition to two elements, RU1 and RU2, composed of a variable number of tandem repeats. RU1 consists of 30 bp repeating units and its length shows minimal variation between individuals. The RU2 elements in the hypervariable S232 loci on the X chromosome consist of repeating sequences which are highly asymmetric, with about 90% purines and no C's on one strand. The X-derived RU2 elements range from 0.6 kb to over 23 kb among different individuals, accounting entirely for the observed polymorphism at the S232 loci. Although the repeating units of the RU2 elements in the non-polymorphic S232 loci on the Y chromosome share high sequence homology with those on the X chromosome, they exhibit much higher intrarepeat sequence variation. S232 homologous sequences are found in great apes, old world and new world monkeys. In chimpanzees and gorillas the S232-like sequences are polymorphic in length.

INTRODUCTION

CRI-S232 is an anonymous human genomic clone originally isolated by Knowlton et al (1) which detects a set of highly polymorphic restriction fragments on the distal short arm of the X chromosome, as well as monomorphic fragments on the proximal long arm of the Y chromosome. Linkage studies showed that the S232 sequences on the X chromosome usually segregate as a single locus, the DXS278 locus, although crossover among the S232 hybridizing fragments was observed at a frequency of 1.4%. Clones containing S232-like sequences have been isolated from both the X and the Y chromosomes (2–4), and the organization of the S232 loci on the X chromosome has been elucidated using adjacent unique sequences. It appears that there are four copies of S232-like sequences on Xp22.3, distal to the DXS143 (dic56) locus. These S232 loci are spread over 2.5 Mb, with one distal and three proximal to the steroid sulfatase locus. The STS locus, which encodes the microsomal enzyme steroid sulfatase, is characterized by a very high frequency of deletions. STS deficiency in humans results in X-linked ichthyosis, a common genetic disease with a prevalence of one in 2000–5000 males (5). About 90% of STS deficient patients studied have deleted their entire STS gene, which spans over 146 kb (6–10). The breakpoints in the majority of these deletions lie within two flanking S232 sequences, suggesting that frequent deletions at the STS locus are caused by recombination between these S232 sequences (4). We now report the characterization and analysis of clones containing sequences from five S232 loci, three on the X chromosome and two on the Y chromosome. Our results show that each S232-like sequence contains two VNTR (variable-number tandem repeat) elements in addition to unique sequences. The VNTR sequence in the hypervariable S232 loci on the X chromosome and the constant S232 loci on the Y chromosomes were compared. In addition, the presence of S232-like sequences in primates was also studied.

MATERIALS AND METHODS

Clones containing S232-like sequences

The original clone CRI-S232 was obtained from Collaborative Research Inc. (Redford, MA). CRI-S232 was used as a probe to isolate several clones containing S232-like sequences from a human cosmid library (4). The Eco R1 fragments that hybridize with CRI-S232 were subsequently subcloned in pUC19 to give clones CS2B, CS12F, CS15A, CS17A and CS21A. CS17A, CS21A and CS2B are derived from the S232A, S232B and S232C loci on the X chromosome, respectively, while CS15A and

* To whom correspondence should be addressed

CS12F are derived from the S232D and S232E loci on the Y chromosome, respectively (4). CS1A is another S232-containing clone isolated from a human genomic library in EMBL3 (Clontech HL1006). It was later found to be derive from the S232C locus.

Southern blotting and DNA sequencing

Isolation of high molecular weight DNA from human leukocytes and Southern blotting of DNA were carried out as described previously (11). M13 subcloning, single and double strand DNA sequencing were carried out according to standard procedures (12) using Sequenase DNA sequencing kits purchased from United States Biochemical Corporation (Cleveland, Ohio). The sequencing primers are PR3: 5'-CTGTCTGGCTAACTGTGA-CC; PR4: 5'-ATCTTGCTAGCACATACACC; PR6: 5'-CAG-CACGAGCTGCCGCC; and PR9: 5'-AACTTAGTCGCTG-CTCTCGG.

MboII partial digestion

BglII fragments of CRI-S232 were end labeled by filling in using DNA polymerase I Klenow fragment (12) followed by digestion with Xmn I. The 1.1 kb fragment containing the RU2 region was gel purified and subjected to MboII partial digestion according to Jeffreys et al (13) with some modification. Briefly, about $1X10^4$ cpm of the end labeled fragment was digested with 0.5 to 2 units of MboII at 37°C in a 10 ul reaction. Aliquots containing one fifth of the reaction mixture were removed at 20 second intervals and added to the same tube containing 10ul of 0.5M EDTA. The products were then analyzed by gel electrophoresis on a 5% polyacrylamide gel (acrylamide:Bis = 29:1).

RESULTS

Sequence organization of CRI-S232

CRI-S232 contains a 7 kb Eco RI fragment which can be divided into four regions by restriction enzymes XmnI, Bgl II and Sst I as shown in Figure 1. No cross-hybridization was observed between these regions (data not shown). To determine which segment of CRI-S232 accounts for the length variation observed at the various S232 loci, Genomic DNA from eight to nine unrelated individuals was Southern blotted with probes a, b, c and d from these four regions of CRI-S232. To separate the various regions, the DNA was digested with restriction enzymes XmnI plus EcoRI for hybridization with probe a, with XmnI plus BglII for hybridization with probes b and c, and with Sst I plus XmnI for hybridization with probe d. As shown in Figure 1A and 1D, probes a and d detect only monomorphic fragments. Restriction fragment length polymorphism was detected by probe c (Figure 1C), though variation in the fragment lengths is rather small. Probe b detected restriction fragments ranging from 0.6 kb to more than 23 kb, and no two individuals have the same hybridization pattern (Figure 1B). Variation in region b therefore accounts for most of the observed polymorphism at the S232 loci.

Restriction fragments of CRI-S232 containing regions a, b+c and d were subcloned in M13 for DNA sequencing. While regions a and d were easily cloned, we were unable to recover clones containing the 2.0 kb Xmn I + Sst I fragment covering regions b+c. The inserts in 16 clones isolated containing regions b+c were all smaller than 1 kb, suggesting that these regions contain unstable sequences. DNA sequencing data reveals that both regions contain tandem repetitive sequences, designated as RU2 and RU1, respectively, separated by a stretch of 160 bp of unique sequence. The structures of RU1 and RU2 are described below. About 260 bp of region a and 1 kb of region d have been sequenced and no tandemly repeated sequences were found. CRI-S232 therefore contains two VNTR elements bracketed by 2.6 kb and 2.4 kb of unique sequences at either end.

Comparison of various S232-like sequences

Although Southern hybridization experiments of human genomic DNA with CRI-S232 suggest that there are likely to be 8 copies of S232-like sequences in the human genome, we have isolated only five of them from a cosmid library (4). All of the Eco R1 fragments containing the S232-like sequences have been subcloned in pUC19. CS17A, CS21A and CS19C are derived from the S232A, S232B and S232C loci on the X chromosome whereas CS15A and CS12F are derived from the S232D and S232E loci on the Y chromosome. Another clone containing S232-like sequences was isolated from a phage library and its S232-containing Eco R1 fragment was subcloned in pUC19 to give clone CS1A. The inserts of CRI-S232, CS1A and CS19C are of the same size, and DNA sequencing data described below indicates that these clones are derived from the same locus. Since these clones were isolated from different libraries constructed by three different companies, they very likely represent the same locus in three distinct individuals. We therefore have clones containing sequences from five S232 loci in one individual, as well as clones containing three alleles of the S232C locus for comparison.

The region to the left of the hypervariable region RU2 in all the S232 clones was sequenced using a sequencing primer PR4 (Fig. 1), and the results are shown in Figure 2. As mentioned above CRI-S232, CS19C and CS1A are identical over 276 base pairs sequenced, indicating that they are derived from the same S232C locus. Sequence similarity between S232C and the other

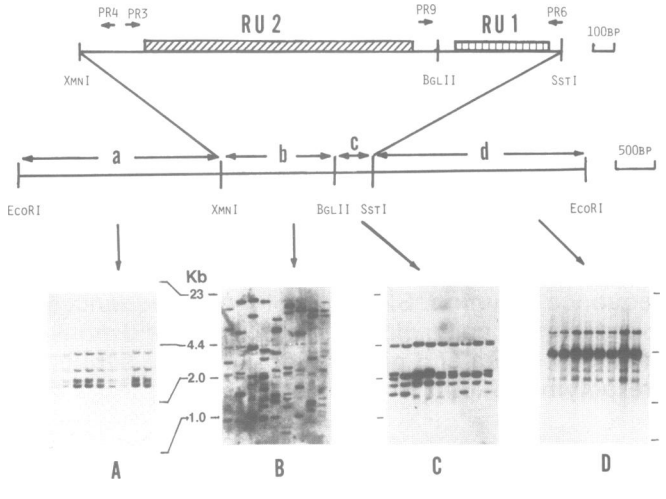

Figure 1. Sequence organization of CRI-S232. Top: Restriction map of CRI-S232. The human insert in CRI-S232 is divided into four regions by various restriction enzymes. Regions b and C contain the RU2 and RU1 elements, respectively, which contain VNTR sequences. DNA sequencing primers and their 5'-3' orientations are indicated with horizontal arrows. Bottom: Southern hybridization of genomic DNA from 8–9 unrelated individuals with fragments containing the four regions of CRI-S232. The fragments and restriction enzymes used are A: fragment a, XmnI + EcoRI; B: fragment b, XmnI + BglII; C: fragment c, XmnI + BglII; and D: fragment d, SstI + XmnI.

two loci on the X chromosome, S232A (CS17A) and S232B (CS21A), are 93% and 90% respectively, whereas S232A and S232B share 91% sequence similarity. It is interesting to point out that S232A and S232B are the only two loci which have been found to be involved in recombination events resulting in STS deficiency. On the basis of sequence homology, we cannot see why S232C would not be involved in a similar recombination process. The fact that S232A and S232B are in the same orientation may be important in explaining why the majority of the deletions occur only between S232A and S232B. DNA sequences of the two Y loci, S232D (CS15A) and S232E (CS12F) are almost identical up to nucleotide 185, including a 6 bp deletion as compared to the X sequences. This region shares 90% similarity with CRI-S232. While S232E continues to show homology with CRI-S232 beyond nucleotide 185, S232D diverges from both of them.

The RU1 elements in the S232 clones were sequenced using primers PR6 and PR9. Most of them consist of a head to tail tandem array of a 30 bp repeating unit shown in Figure 3A. Five basic repeating sequences, designated as motifs A-E, were observed which vary only at nucleotide positions 16, 19 and 24. In motif D all three positions have G's while in motif A, B and C only two of them are G's. In addition, variations of motifs A and B containing base substitutions at other positions were found at the S232B locus, and they are designated as A′ and B′,

```
                                                                      40
LOCUS       CLONE                                                     '
            CRI-S232  TTAGGTTTCA TTGTTAAAAT ACGTTTGCGG TTATAGAAAC
(X)S232C    19C, 1A   .......... .......... .......... ..........
(X)S232A    17A       .......... .......... ..C...C... ..........
(X)S232B    21A                  .......... ..C...??.. ..........
(Y)S232E    12F             ..... .......... ..C...??CT CA...T....
(Y)S232D    15A                ... .......... ..C...??CT CA...T....

                                                                      90
                                                                      '
CRI-S232    CGATGGAAGA ATCGTTTCCA TCTCTTTCCT TTCAGCGTCC CCTCCCTACA
(X)S232C    .......... .......... .......... .......... ..........
(X)S232A    T......... .......... .......... .......... ..........
(X)S232B    .....A.... .......... .......... .......... ..........
(Y)S232E    T.....G.C. ..G....... .......... .....G.... ..........
(Y)S232D    ?.....G.C. ..G....... .......... .....G.... ..........

                                                                     140
                                                                      '
CRI-S232    CTCTAATATA ATTAAGAAGA AAAATTCAAA GTAGAGCAAA ATCTGTCTAC
(X)S232C    .......... .......... .......... .......... ..........
(X)S232A    A......... .......... .......... A......... ....A.....
(X)S232B    AC........ T......... .......... ......A... ....A.....
(Y)S232E    A......... .......... .......... .......... ....A.....
(Y)S232D    A......... .......... .......... .......... ....A.....

                                                                     190
                                                                      '
CRI-S232    TTGAAAAGAG TCTTTTTATT TTTGCGACTG AGGAGACAAA AAGTATATTT
(X)S232C    .......... .......... .......... .......... ..........
(X)S232A    ....T..A.. C......... .....A.... .A........ .....C....
(X)S232B    .......A.. C........A ..C..A.... .......... .....C....
(Y)S232E    .......A.. C......*** ***..A.... .......... .*...C....
(Y)S232D    .......A.. C......*** ***..A.... .........G .....GAAGC

                                                                     240
                                                                      '
CRI-S232    CATATTTCCA TTCATTAGAA ATACACAGGT CTTTTCTGAA TATAGTAGAA
(X)S232C    .......... .......... .......... .......... ..........
(X)S232A    .......... .A........ .......... .......AT. ..........
(X)S232B    .......... .G.G...... .......... .C.....AT. ...***....
(Y)S232E    ........*. .AT....... .G.......
(Y)S232D    TTGCTA

                                          270
                                           '
CRI-S232    TTTACTATAT AGAAATATAC ATAAATATAT GTAAAT
(X)S232C    .......... .......... .......... ......
(X)S232A    ....G..... .C.......T G.......
(X)S232B    ......G... .C.......T .G..CC.... ......
```

Figure 2. DNA sequence comparison among the S232 clones of the region to the left of the RU2 element, as shown in Figure 1. The sequencing direction is from right to left using primer PR4. Identical nucleotides are indicated with . and base deletions are indicated with *. (X) and (Y) in front of the loci indicate their chromosomal locations.

respectively. The RU1 motif organization is shown in Figure 3B for the various S232 clones. Clones CRI-S232, CS1A and CS19C, which represent three alleles of the S232C locus, have different motif organization. They contain 11, 14 and 12 repeating units, respectively. It is noted that the first 6 motifs of CRI-S232 and CS1A are identical while the last 8 motifs of CS1A and CS19C are identical. Therefore CS1A could have been generated by crossing over between CRI-S232 and CS19C. Alternatively, CRI-S232 and CS19C could have been generated from CS1A by deletion of motifs 7−9 and motifs 2−3, respectively. Elucidation of the mechanism(s) that generates these variations awaits family studies of cases in which new alleles of the S232 loci have been generated. CS17A from the S232A locus contains 8 motifs identical to the first 8 motifs of CRI-S232, whereas CS21A from the S232B locus contains two complete motifs and the first 4 bp of the repeating unit. The RU1 elements of the S232D and S232E loci on the Y chromosome are identical, and both contain only one repeating unit in addition to the first 6 bp of the repeat. It is noted that all of the RU1 elements start with motif A, suggesting that it is the original motif from which other motifs are derived.

The RU2 elements of the S232 clones were sequenced initially using the Sanger dideoxy chain termination method with primers PR3. Unreadable sequences were later determined using the Maxam-Gilbert chemical cleavage method. Even so only limited sequence data from the ends of the RU2 elements were obtained due to difficulties caused by high GC content and unusual structure of this region. The RU2 elements of the S232 loci on the X chromosome consist of tandem arrays of a repeating unit which in itself contains variable number of a GGGA (or TCCC on the other strand) tetranucleotide repeat (Figure 4A) The organization of the RU2 motifs differs at the S232 loci. The orders of the limited RU2 motifs sequenced in clones CRI-S232 (S232C), CS17A (S232A) and CS21A (S232B) are 3533333,

```
A:

RU1 MOTIF TYPE                                              INDICATION

CTGAGTCAGGAGAGCGAGCTGGAGGAACCA                                 |A|
..............G..G....A......                                 |B|
..............C..G....G......                                 |C|
..............G..G....G......                                 |D|
..............G..A....A......                                 |E|

............GC.G..C....G......                                 |A'|
..............G..G....A....G.                                 |B'|


B:

MOTIF ORGANIZATION

                LOCUS     CLONE
X CHROMOSOME    S232C:    CRI-S232  |A|B|C|B|C|D|C|E|C|E|E|
                                1A  |A|B|C|B|C|D|B|C|B|C|E|C|E|E|
                               19C      |A|B|C|D|B|C|B|C|E|C|E|E|

                S232A:          17A  |A|B|C|B|C|D|C|E|

                S232B:          21A  |A'||B'||

Y CHROMOSOME    S232D:          15A  |A||

                S232E:          12F  |A||
```

Figure 3. Comparison of RU1 repeating sequences among the S232 loci. The sequencing direction is from right to left, using primer PR6 (see Figure 1). A. The repeating sequences in the RU1 elements. There are seven different sequences which are designated as the A-E, A′ and B′ motif, respectively. B. Sequence organization of the RU1 repeats in the S232 clones. The RU1 elements contain head to tail tandem arrays of the motifs. Clones 21A, 15A and 12F contain incomplete motifs at the ends, as indicated with ||.

4433, 34434, respectively, with the numbers indicating the copy numbers of the tetranucleotide repeats in the motif. The X-encoded RU2 repeating unit is rather asymmetric in that it contains about 90% pyrimidines and no G's on one strand and 90% purines and no C's on the other strand. There is an Mbo II restriction site in the repeating unit which allows us to study interrepeat length variation over a larger range using the method of partial digestion (13). The XmnI + Bgl II fragment containing the RU2 element in CRI-S232 was end labeled at the Bgl II site and partially digested with Mbo II as shown in Figure 4B. The continuous ladder represents about 25 repetitive motifs. Irregularity in the spacing indicates length variation between the repeating units. It is also noticed that the intensities of the bands are not the same, suggesting sequence or structural difference in some of the repeating units which impedes MboII digestion. In the human genome the XmnI + Bgl II fragments containing the RU2 sequences range from 0.6 kb to over 23 kb (Figure 1B). Since only 0.4 kb of these fragments is contributed by flanking unique sequence, the copy numbers of the RU2 repeating units in these fragments therefore range from 6 to 800 copies.

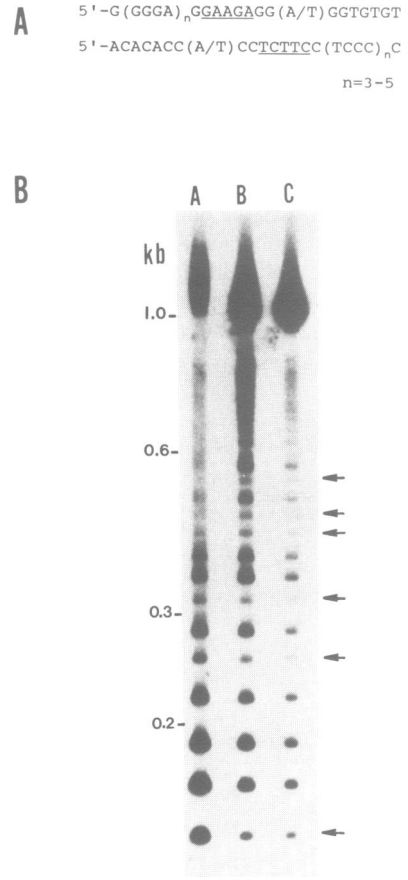

Figure 4. Sequence analysis of the RU2 element in CRI-S232. A: The RU2 repeating units. The sequences of both strands are shown in the 5'-3' direction to facilitate comparison with other published VNTR sequences, and to maintain consistency with other sequences shown in Figure 5. The RU2 repeating unit contains three to five copies of a tetranucleotide GGGA/TCCC repeat. The recognition sequences for the restriction enzyme MboII are underlined. B. MboII partial digestion of the XmnI + BglII fragment containing the RU2 element in CRI-S232. The fragment was labeled at the BglII end and partially digested with (A) 2 units, (B) 1 unit and (C) 0.5 units of MboII. Arrows point to bands with reduced intensity.

DNA sequences of the RU2 region and the segment between RU1 and RU2 in CRI-S232 are shown in Figure 5. As mentioned above, during subcloning of the CRI- S232 RU2 region in M13 we have isolated several clones containing inserts greatly reduced in size. Three of these clones were sequenced and their deletion breakpoints are indicated in Figure 5. Clone 3.2 has deleted all the RU2 repeating sequences whereas clone 3.4 and 3.6 have retained 2 and 3 copies of the repeating units, respectively.

The S232 loci on the X and the Y chromosomes differ greatly in their polymorphism, with the X loci hypervariable and the Y loci constant. In addition, the X loci are unstable on subcloning while the Y loci are not. To understand the mechanisms underlying such differences, sequences from the two Y derived clones and that of CRI-S232 are compared (Figure 5). Sequences of the S232D and S232E loci on the Y chromosome are identical over 470 base pairs determined, and they share 89% similarity with that of CRI-S232. Two significant differences are observed in the Y sequences as compared to the X sequences. First, there are many base substitutions in the repeating units so that none of the units have identical sequences. Second, there are G's in the 'no-G-strand'. These differences may contribute to the stability of the Y sequences.

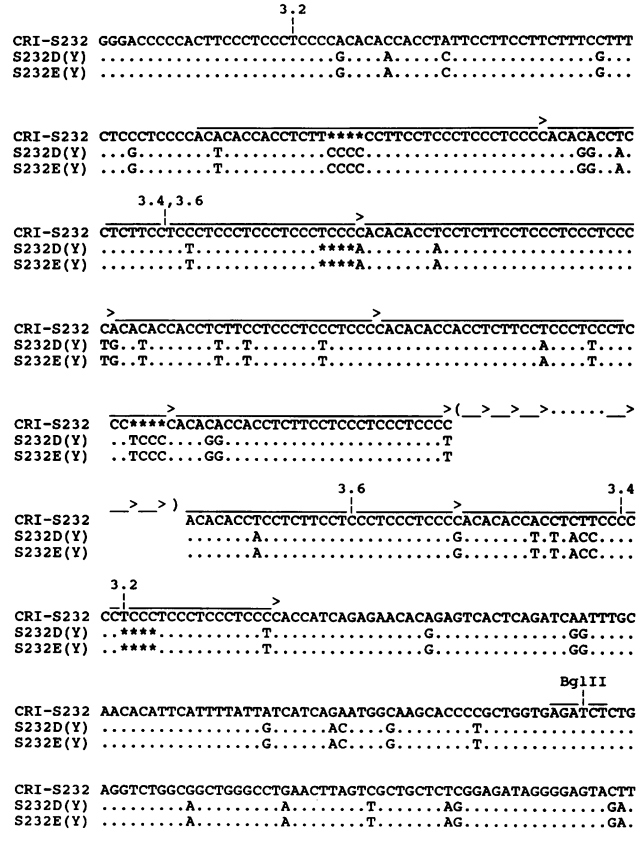

Figure 5. Comparison of the RU2 elements and the 3' flanking sequences in the S232 clones from the Y chromosome with that in CRI-S232. The sequences are determined using primer PR3 and the direction is from left to right as shown in Figure 1. Identical nucleotides are indicated with . and base deletions are indicated with *. The RU2 repeating units are indicated with __>. The deletion breakpoints in three M13 subclones, 3.2, 3.4 and 3.6, are indicated above the sequences. The location of the BglII site which separates the RU1 and RU2 elements is also indicated.

The presence of S232-like sequences in primates

The evolutionary conservation of the S232 sequences in other mammals, especially primates, was investigated. Genomic DNA was isolated from male and female great apes including chimpanzees (Pan troglodytes), gorillas (Gorilla gorilla) and orangutans (Pongo pygmaeus); old world monkeys including moustached guenon (Cercopithecus cephus), barbary macaque (Macaca sylvanus) and gelada baboon (Theropithecus gelada); and new world monkey including woolly monkey (Lagothix lagothricha) and spider monkey (Ateles geoffroyi). In addition, DNAs from lemur, horse and mouse were also included in the study. These DNAs were digested with Eco RI and Southern blotted with the CRI-S232 probe. While DNAs from lemur, horse and mouse failed to hybridize with the probe (data not shown), S232-like sequences are present in all the primates tested (Figure 6). As animals at a progressively greater evolutionary distance from humans are studied, there seems to be a decrease in sequence similarity as indicated by weaker signals, as well as a decrease in the copy number as indicated by the decreased number of hybridizing bands. To address the question of whether the S232 sequences in other primates are also polymorphic, DNA samples were obtained from six chimpanzees (three males and three females) and five gorillas (three males and two females) from three zoos in the United States. Even though some of these animals may be probably blood related, their hybridization patterns are clearly different, indicating that the S232-like sequences in chimpanzees and gorillas are also polymorphic. In the gorillas, some cross hybridizing fragments were present only in males, suggesting that the S232 sequences are present on the gorilla Y chromosome.

DISCUSSION

In this report we describe our studies of the S232-like low copy repetitive sequences on the human sex chromosomes. Two VNTR-containing elements, RU1 and RU2, have been identified in the S232 repeats. The lengths of these elements are either hypervariable or constant depending on their chromosomal locations. The repeating units in the RU1 element are 30 bp long and heterogeneous in nucleotide sequences, differing at three or more positions. Their copy numbers vary slightly among the various S232 loci in one individual, ranging from one copy at the S232D and S232E loci on the Y chromosome to over 10 copies at the S232C locus on the X chromosome. Small variation among different alleles was observed at the S232C locus. Although the RU2 elements in the hypervariable S232 loci on the X chromosome and the non-variable S232 loci on the Y chromosome show a high degree of sequence similarity, their properties are quite different. The X-linked RU2 elements show only a low level of intrarepeat sequence variation, and the repeat copy numbers vary drastically among the various loci, ranging from 6 to 800 copies. On the other hand the repeating units in the Y-linked RU2 elements show greater intrarepeat sequence variation, and their copy numbers are presumably the same among individuals as suggested from the stability of the Y loci. Such correlations between intrarepeat sequence homogeneity and allelic variability have also been found at other VNTRs where loci showing greater levels of allelic variation are usually more homogeneous in the sequences of their repeating units (14).

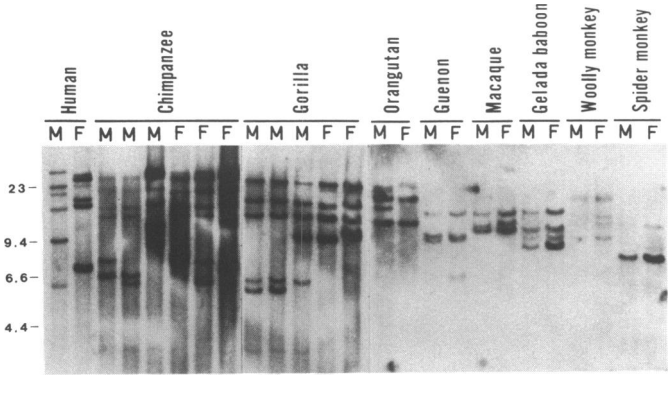

Figure 6. Conservation of S232 sequences in primates. Genomic DNA from male (M) and female (F) primates was digested with EcoRI and Southern blotted with CRI-S232. All of the primates contain S232 homologous sequences, and restriction fragment length polymorphism was observed in chimpanzees and gorillas.

The common features of VNTR sequences are high GC content and strand asymmetry (15). The RU1 repeating units and the X-linked RU2's have GC contents of 60% and 65% respectively. While RU1 has 70% purines on one strand, the X-linked RU2's has about 90% purines. Beside these two features, RU1 and RU2 share no significant homology with other VNTR sequences. About 78% similarity was found between nucleotides 3−16 of RU2 (the G strand in Figure 4) and the 'polycore' minisatellite GGAGGTGGGCAGGXG (14), but this can be attributed to the high G content of these sequences. One interesting feature of the RU2 repeats is the highly asymmetric distribution of purines and pyrimidines between the two strands of the DNA. Homopurine-homopyrimidine tracts, $(dTdC)_n$:$(dAdG)_n$, in DNA have been show to undergo structural transitions to the H-form conformation. In H-form DNA one half of the $(dAdG)_n$ repeats pair with one half of the $(dTdC)_n$ repeats through Watson-Crick pairing and to the other half of the protonated $(dTdC)_n^+$ through Hoogsteen pairing, forming both triple-stranded and single-stranded regions (16−20). The single-stranded regions are sensitive to cleavage by S1 nuclease (16,17). The H-form can also arise from DNA containing homopurine-homopyrimidine mirror repeats (18). Such sequences are present in the RU2 element between two repeats as underlined: GGGGAGGGAG-GGAGGGAGGAAGAGG AGGTGTGTGGGGAGGGAGG-GAGGGAGGAAGAGGAGGTGTGT. Whether H-form DNA is present at the S232 loci is not known, but can be investigated by studies such as S1 sensitivity of the S232 sequences.

It has been suggested that some VNTR sequences are recombinogenic (15). VNTR sequences have been found at a meiotic recombinational hotspot in the mouse major histocompatibility locus (21), and some VNTR sequences share high similarity with the recombination-promoting sequence chi in E. coli (14). In addition, human minisatellite consensus sequences have been found at breakpoints of several oncogene related translocations (22), and more recently the consensus sequence has been shown to stimulate homologous recombination over 10 fold (23). In our previous studies we found that the majority of STS deficient patients have deletions of their entire STS genes caused by recombination between two flanking S232A and S232B loci. This frequent recombination may not be due solely to the long homology (over 7 kb) shared by these loci. It may very well be promoted by the VNTR sequences, especially the RU2 elements. We have cloned and characterized the junction fragments containing the deletion breakpoints of three STS

deficient males (24). Further DNA sequencing of the breakpoints will determine whether the crossover occurred within the VNTR or the flanking sequences.

The molecular mechanisms that generate variability at the VNTR loci are not known. It was originally thought that the length variation arose from unequal crossover between alleles during meiotic or mitotic recombination (25). However this mechanism was rejected in several recent studies on new mutations at various VNTR loci. It was found that new alleles at VNTR loci retain flanking markers on both sides from only one of the parental alleles, contrary to what would have been expected from unequal crossover between two alleles (26–28). Jeffreys et al (13) studied new alleles in sperm and blood cells using the technique of single molecule minisatellite variant repeat mapping. They found that germline and somatic mutations generating new VNTR alleles in single molecules seldom arise by unequal crossover between alleles. Other mechanisms, such as unequal crossover between sister chromatids and replication slippage may be involved. Since meiotic recombination does not appear to be involved in the length alterations of VNTR loci, the lack of pairing of the Y chromosome long arm where the S232 loci are located during male meiosis cannot explain the differences in the variability of the S232 loci on the X and the Y chromosome. The dissimilarities in the sequences of the X- and Y-linked RU2 repeats may therefore be related to the variation in the apparent mutation rates of the S232 loci. Two proteins that bind specifically to hypervariable minisatellite DNA have been identified recently (29). Their possible role in promoting recombination and generating new alleles remains to be elucidated.

In addition the S232 sequences, VNTR's have been found at several loci in the pseudoautosomal region (30, 31, 32) of the sex chromosomes, and at the DXS255 locus in the proximal short arm of the X chromosome (33). It is not known whether any of these sequences promotes recombination or are involved in the generation of deletions or chromosomal rearrangements.

ACKNOWLEDGMENTS

We thank Katherine Neiswanger for primate DNA's, Carey Johnson, Jay Ellison and Henry Lin for useful input and discussion, and Nelva Hitt for help in preparing the manuscript. The work was supported by funds from the Howard Hughes Medical Institute, grant HD12178 from NIH and grant 1-639 from the March of Dimes.

REFERENCES

1. Knowlton, R.G., Nelson, C.A., Brown, V.A., Page, D.C., and Donis-Keller, H. (1989). Nucleic Acids Res. 17:423–437.
2. Ballabio, A., Bardoni, B., Guioli, S., Basler, E. and Camerino, G. (1990). Genomics 8:263–270.
3. Schnur, R.E., Knowlton, R.G., Musarella, M.A., Muenke, M. and Nussbaum, R.L. (1990). Genomics 8:255–262.
4. Yen, P.H., Li, X.M., Tsai, S.P., Johnson, C., Mohandas, T. and Shapiro, L.J. (1990). Cell 61:603–610.
5. Shapiro, L.J. (1985). Advances in Hum. Genet. 14:331–381.
6. Ballabio, A., Carozzo, R., Parerti, G., Gil, A., Zollo, M., Persico, M.G., Gillard, E., Affara, N., Yates, J., Ferguson-Smith, M.A., Frants, R.R., Eriksso, A.W. and Andria, G. (1989). Genomics 4:36–40.
7. Bonifas, J.M., Morley, B.J., Oakey, R.E., Kan, Y.W. and Epstein, E.H. Jr. (1987). Proc. Natl. Acad. Sci. USA 84:9248–9251.
8. Conary, J.T., Lorkowski, G., Schmidt, B., Pohlmann, R., Nagel, G., Meyer, H.E., Krentlr, C., Tully, J., Hasilik, A. and Van Figura, K. (1987). Biochem. Biophys. Res. Commun. 144:1010–1017.
9. Gillard, E.F., Affara, N.A., Yates, J.R.W., Goudie, D.R., Lambert, J., Aitken, D.A., and Ferguson-Smith, M.A. (1987). Nucleic Acids Res. 15:3977–3985.
10. Shapiro, L.J., Yen, P., Pomerantz, D., Martin, E., Rolewic, L., and Mohandas, T. (1989). Proc. Natl. Acad. Sci. USA 86: 8477–8481.
11. Yen, P.H., Allen, E., Marsh, B., Mohandas, T., Wang, N., Taggart, R.T. and Shapiro, L.J. (1987). Cell 49:443–454.
12. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press.
13. Jeffreys, A.J., Neumann, R. and Wilson, V. (1990). Cell 60:473–485.
14. Jeffreys, A.J., Wilson, V., and Thein, S.L. (1985). Nature 314:67–73.
15. Jarman, A.P. and Wells,R.A. (1989). Trends Genet. 5:367–371.
16. Pulleyblank, D.E., Haniford, D.B. and Morgan, A.R. (1985). Cell 42:271–280.
17. Evans, T. and Efstratiadis, A. (1986). J. Bio. Chem. 261:14771–14780.
18. Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A. and Frank-Kamenetskii, M.D. (1987). Nature 330:495–497.
19. Voloshin, O.N., Mirkin, S.M., Lyamichev, V.I., Belotserkovskii, B.P. and Frank-Kamenetskii, M.D. (1988). Nature 333:475–476.
20. Htun, H. and Dahlberg, J.E. (1988). Science 241:1791–1796.
21. Steinmetz, M., Uematsu, Y. and Fischer Lindahl, K. (1987). Trends Genet. 3:7–10.
22. Krowczynska, A.M., Rudders, R.A. and Krontiris, T.G. (1990). Nucleic Acids Res. 18:1121–1127.
23. Wahls, W.P., Wallace, L.J. and Moore, P.D. (1990). Cell 60:95–103.
24. Johnson, C.L., Li, X.-M., Yen, P.H. and Shapiro, L.J. (1990). Am. J. Hum. Genet. 47:A222.
25. Jeffreys, A.J., Royle, N.J., Wilson , V. and Wong, et al. (1988). Nature 332:278–281.
26. Wolff, R.K., Nakamura, Y. and White, R. (1988). Genomics 3:347–351.
27. Wolff, R.K., Plaetke, R., Jeffreys, A.J. and White, R. (1989). Genomics 5:382–384.
28. Morral, N., Nunes, V., Casals, T. and Estivill, X. (1991). Genomics 10:692–698.
29. Wahls, W.P, Swenson, G. and Moore, P.D. (1991). Nucleic Acids Res. 19:3269–3274.
30. Simmler, M.C., Johnson, C., Petit, C., Rouyer, F., Vergnaud, G. and Weissenbach, J. (1987). EMBO J. 6:963–969.
31. Page, D.C.,Bieker, K., Brown, L.G., Hinton, S., Leppert, M., Lalouel, J.-M., Lathrop, M., Nystrom-Lahti, M., de la Chapelle, A. and White, R. (1987). Genomics 1:243–256.
32. Inglehearn, C.F. and Cook, H.J. (1990). Nucleic Acids Res. 18:471–476.
33. Fraser, N.J., Boyd, Y. and Craig, I. (1989). Genomics 5:144–148.