Volume 17 Number 7 1989 Nucleic Acids Research

Definition of a second dimeric subfamily of human α satellite DNA

James D.Thompson, James E.Sylvester[1]*, Iris Laudien Gonzalez[1]*, Carl C.Costanzi and David Gillespie

Department of Neoplastic Diseases, Hahnemann University, Philadelphia, PA 19102 and [1]Department of Human Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA

Received November 16, 1988; Revised and Accepted February 27, 1989 EMBL accession nos+

ABSTRACT

We describe a new human subfamily of alpha satellite DNA. The restriction endonuclease XbaI cleaves this subfamily into a collection of fragments which are heterogeneous with respect to size. We compared the sequences of 6 clones from four different XbaI size classes. Clones from a single size class were not necessarily more related than clones from different classes. Clones from different size classes were found to produce almost identical hybridization patterns with XbaI-digested human genomic DNA. All clones were found to share a common dimeric repeat organization, with dimers exhibiting about 84% sequence identities, indicating that the clones evolved from a common progenitor alphoid dimer. We show that that this subfamily, and the EcoRI dimer subfamily originally described by Wu and Manuelidis (1), evolved from different progenitor alphoid dimers, and therefore represent distinct human alphoid subfamilies.

INTRODUCTION

Alpha satellites are a family of repetitive DNAs based on a ~171-bp repeating unit. This satellite was first identified as a rapidly-annealing fraction of the African green monkey genome (2), and has since been found within the genomes of all Old World primates (3,4). Similar to other satellite DNAs, alpha satellites occur as large arrays of direct repeats, called domains. Alpha satellite domains have been found on all human chromosomes and are localized primarily to the centromeric regions (5).

Human alpha satellite domains contain multiple levels, or hierarchies, of repeat organizations. Domains are comprised of multimeric repeating units, termed higher-order repeats (6,7,8), which may consist of as few as two alphoid monomers (9), or as many as sixteen (7) or more monomers. Higher-order repeating

units within a single domain are very similar in sequence (>98% sequence identity). Higher order repeat structures have been found to vary, however, among the different human chromosomes. It may therefore be possible to identify each human chromosome based on the repeat structure of its alpha satellite component (10).

The sequence diversity among the human alpha satellite fragments that have been cloned suggests that many different subfamilies of alpha satellite DNA are present within the human genome (11). Subfamilies have been differentiated on the basis of restriction site periodicity, cross-hybridization, repeat organization, nucleotide sequence identity, and the presence of particular sequence variations (10-13,23). Alexandrov et al. (23) have determined by in situ hybridization that at least three suprachromosomal subfamilies exist within the human genome which are present on distinct chromosomal subsets. The two human subfamilies that have been most extensively characterized at the sequence level are the EcoRI dimer subfamily originally described by Wu and Manuelidis (1), referred to as suprachromosomal family 1 by Alexandrov et al., and the pentameric subfamily characterized by Waye and Willard (14), referred to as suprachromosomal family 3 by by Alexandrov et al. (23).

The EcoRI dimer subfamily has a dimeric repeat organization, with dimers generally containing a single recognition sequence for the restriction endonuclease EcoRI. Randomly cloned dimers were found to exhibit from about 82% to 99% sequence identities (12). This subfamily was found to be most prevalent on chromosomes 1, 3, 7, 10, and 19 (5), as well as on chromosomes 5, 6, 12, and 16 (23).

Alphoid domains belonging to the pentameric subfamily have been characterized on chromosomes 1, 11, 17, and X (14). Although these domains evolved from a common progenitor alphoid pentamer, they exhibit different higher-order repeat structures. The higher-order repeating units within a given domain are $\geq$ 98% identical in sequence, while the units from different chromosomes are about 83% similar. The differences in repeat organizations and sequence identities indicate that these domains have evolved independently following their divergence from the ancestral

pentameric unit (14).

Jorgensen et al. (15) have cloned tetrameric EcoRI fragments from chromosomes 13 and 21 which are 99% similar at the sequence level. These data indicate that alpha satellite DNA on these two acrocentric chromosomes may interact. This is in contrast to domains of the pentameric subfamily described above, which appear to have evolved independently following their divergence from the ancestral pentamer.

A single chromosome can harbor domains belonging to more than one alpha satellite subfamily. Waye et al. (8) have demonstrated that at least two distinct alpha satellite domains are present on chromosome 7. Restriction analysis showed that these satellites have different repeat organizations, while nucleotide sequence analysis revealed that that one domain belongs to the EcoRI dimer subfamily, while the other belongs to an as-yet undefined subfamily (11).

The genetic mechanisms responsible for generating satellite DNA, for dispersing satellites to other chromosomes, and for producing the hierarchical repeat organizations, are unknown. Processes such as unequal homologous recombination, gene conversion, and excision/reintegration have been implicated (16).

In this report we describe a new human alpha satellite subfamily, which we refer to as the SX subfamily in this report ("S" and "X" refer to the restriction endonucleases StuI and XbaI which cleave frequently within this subfamily, see Fig.1). Gray et al. (17) cloned two tetrameric alpha satellite fragments from XbaI-digested human genomic DNA. The sequences of the tetramers were found to be 98% identical, and to be comprised of two dimers exhibiting 84% sequence identity. These tetramers produced complex hybridization patterns with XbaI-digested human DNA. They hybridized not only to fragments 4n in length, but also to fragments 2n, 5n, 6n, 8n, 10n, and 11n in length, where n represents the length of an alphoid monomer. In order to examine how these these various XbaI fragments are related, we cloned and sequenced two dimers, another tetramer, and a pentamer from a human XbaI genomic library, and an 8-mer XbaI fragment from human chromosome 21-specific DNA. We found that our clones produced hybridization patterns with XbaI-restricted human DNA which were

almost identical to the patterns produced with the previously-cloned tetramers. Sequence analyses showed that all of the clones are comprised of alphoid dimers exhibiting about 84% sequence identities, indicating that they evolved from a common progenitor dimer unit. We therefore define the SX subfamily as satellites evolving from this progenitor dimer. We show that the SX subfamily and the EcoRI dimer subfamily evolved from different progenitor dimers.

METHODS

Filter hybridizations

Restriction digests, agarose gel electrophoresis, and Southern transfers were performed as described (18). Plasmid DNA was radiolabeled by nick translation using ^{32}P- or ^{35}S-labeled dCTP and nick translation reagents purchased from Amersham. Filter hybridizations were performed at 42°C in 50% formamide, 6x SSC, 0.5% SDS, 0.05 M $NaPO_4$ (pH 6.8) containing 10^6 cpm of probe/ml of solution. Filters were washed as described in the text.

Cloning of human XbaI satellite fragments

Human genomic DNA was extracted from spleen tissue of a male CML patient as described (3), and 1 ug was digested with XbaI (Boehringer-Mannheim), cloned into the XbaI site of pUC13 plasmid (Promega) and maintained in E. coli strain JM83 by general methods previously described (18). Recombinants were identified by growing transformants in the presence of "X-gal" (Promega) as described by the manufacturer. Twenty-four recombinants were picked at random and the sizes of the inserted human DNA determined by agarose gel electrophoresis. Twelve of these recombinants had inserts which co-migrated with repeat bands in XbaI-digested human DNA. These clones were radiolabeled and hybridized to transfers of human DNA digested with either XbaI, KpnI, or EcoRI as described above. Filters were washed twice with 20x SSC at 42°C for 30 min and radioautographed. Four of the clones hybridized to alpha satellite-specific bands. These were SX1 and SX4 (alphoid dimers), SX5 (a tetramer), and SX15 (a pentamer).

XBA21 was cloned from a size-selected XbaI digest of a

hamster-human somatic cell hybrid line, 153 E9A, kindly provided by M. Van Keuron and D. Patterson. A partially enriched library was made in pUC12 and screened with a previously isolated alphoid clone, XBA14 (to be described elsewhere).

Nucleotide sequencing

Nucleotide sequences of the cloned XbaI satellite fragments were determined from each end of double-stranded DNA templates as described (19,20). Sequencing reagents and ^{35}S-labeled dATP were purchased from Amersham. Reverse sequencing primer was purchased from Boehringer-Mannheim. Sequences were determined either unambiguously for one strand, or for both strands in instances where base identities were unclear.

In the case of XBA21, the 1365-bp fragment was subcloned into M13 phage vectors mp18 and mp19 and overlapping subclones were generated using the single-strand deletion technique of Dale et al. (21) using reagents purchased in kit form from International Biotechnologies, Inc.

RESULTS

Hybridization patterns of cloned SX-family satellite fragments

Gray et al. (17) cloned two XbaI alphoid tetramers, pE1 and pF5, which hybridized to human XbaI fragments 2n, 4n, 5n, 6, 8n, 10n, and 11n in length, where "n" represents the length of an alphoid monomer. In order to examine the relationship between the various XbaI fragments, we isolated two dimers (SX1 and SX4), a tetramer (SX5), and a pentamer (SX15), from a library of XbaI fragments of total human genomic DNA (METHODS). The top of Figure 1 shows a comparison of the patterns produced with the dimer, SX1, the pentamer, SX15, and the tetramer, pE1, when used as probes in hybridizations with human genomic DNA digested with restriction endonucleases known to cleave frequently within human alpha satellite DNA. The hybridization patterns produced with the dimer and the pentamer were nearly indistinguishable from that of the tetramer, pE1. These results indicate that these clones originated from the same alphoid subfamily, which we refer to as the SX subfamily. The majority of the hybridization signal with EcoRI, BamHI, HindIII, PvuII, PstI and KpnI digested human DNA, occurred with fragments too large to be resolved in these

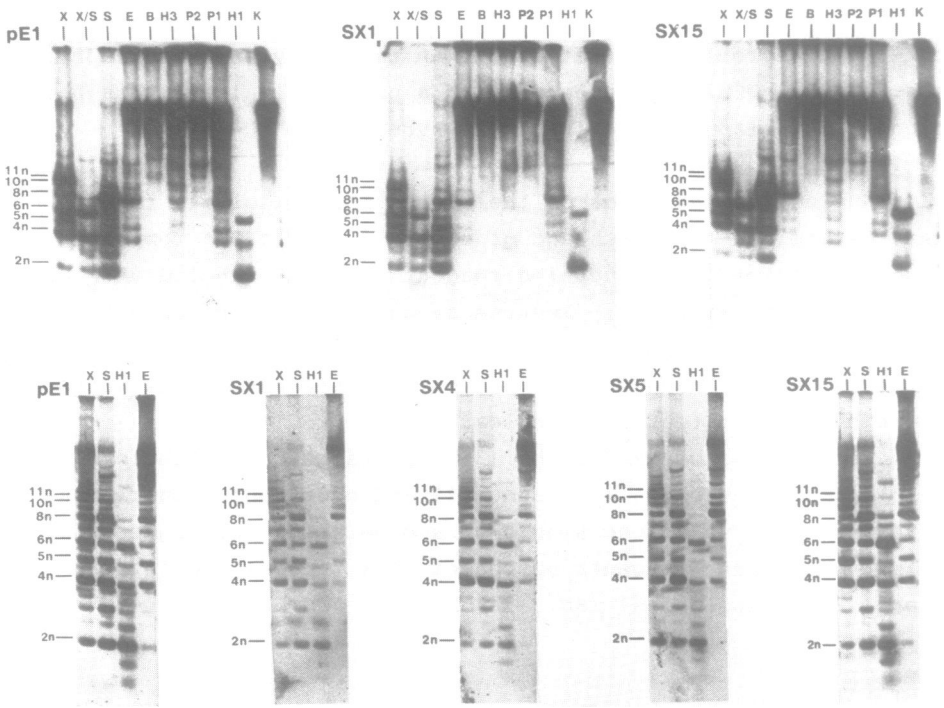

Figure 1. Hybridization patterns of SX satellite clones. Top: Human genomic DNA was cleaved with the restriction enzymes indicated and 1 ug was loaded per lane, electrophoresed through 1% agarose gels, transferred to nylon membranes (Nytran, Schleicher and Schuell), and hybridized with ^{32}P-labeled pE1, SX1, or SX15 plasmid DNA as described in METHODS. Filters were washed twice in 20x SSC at 42°C for 30 min, then washed in 0.1x SSC at 65°C for 15 min. Time of radioautography was 16 hr. Bottom: Human DNA was cleaved with the restriction enzymes indicated and 2 ug was loaded per lane and electrophoresed through a 1.5% agarose gel, transferred to a nylon membrane, and hybridized with ^{35}S-labeled pE1, SX1, SX4, SX5, or SX15 plasmid DNA (2 x 10^{6} cpm per ml). Filters were washed as above. Time of radioautography was 7 days. X, XbaI; S/X, StuI and XbaI double digest; S, StuI; E, EcoRI; B, BamHI; H3, HindIII; P2, PvuII; P1, PstI; H1, HinfI; K, KpnI.

gels. These results indicate that recognition sequences for these enzymes occur infrequently within SX subfamily satellite clusters.

The majority of the hybridization signal within the XbaI, StuI, and HinfI digests, occurred with fragments ranging in size from 0.17 kb to about 2 kb, with very little signal occurring

with higher molecular weight fragments. The patterns produced with these endonucleases are shown more clearly in the bottom of Figure 1, where the clones were labeled with ^{35}S in order to improve resolution in the radioautograms. Strong hybridization signals were produced with multiple bands in each of these digests. While the hybridization patterns of the various clones were similar, the signal intensities to specific bands varied slightly among the cloned fragments. For example, the pentamer, SX15, produced a stronger hybridization signal to XbaI fragments 1530 bp in length (band 9n in Figure 1) relative to the other clones. It is likely that these slight differences reflect variations in repeat organization within the different SX satellite domains from which the clones originated.

Sequence analyses

Figure 2 shows the nucleotide sequences of the clones relative to the human alpha satellite consensus sequence derived by Waye and Willard (11). Included in this figure are sequences from an 8-mer XbaI fragment cloned from chromosome 21-specific DNA (METHODS). Table 1 shows the results of sequence comparisons between the monomers. Alternating monomers were found to have sequence identities of about 84%, while adjacent monomers were found to exhibit about 73% identity. This result shows that all of the clones are comprised of dimers exhibiting about 84% identities. These results were similar for our clones, as well as for the clone pE1 (Table 1 and ref. 17), suggesting that the clones reported here, and the tetramers cloned by Gray et al., evolved from a common ancestral alphoid dimer.

pE1 and pF5 were found to have very similar nucleotide sequences (98% similarity) (17). The tetramer reported here, SX5, only showed about 85% sequence identity with pE1 (Table 1). Also, the sequences of the two dimers (SX1 and SX4) were only about 83% similar. These results show that XbaI fragments in the same size class are not necessarily more related than fragments from different size classes. In fact, the two monomers comprising the dimer, SX1, were more similar to the first two monomers of the tetramer, pE1 (92% identity), than to other monomers in Table 1. Similarly, sequences within the dimer clone, SX4, and the 8-mer clone, XBA21, were 97% identical, and sequences within

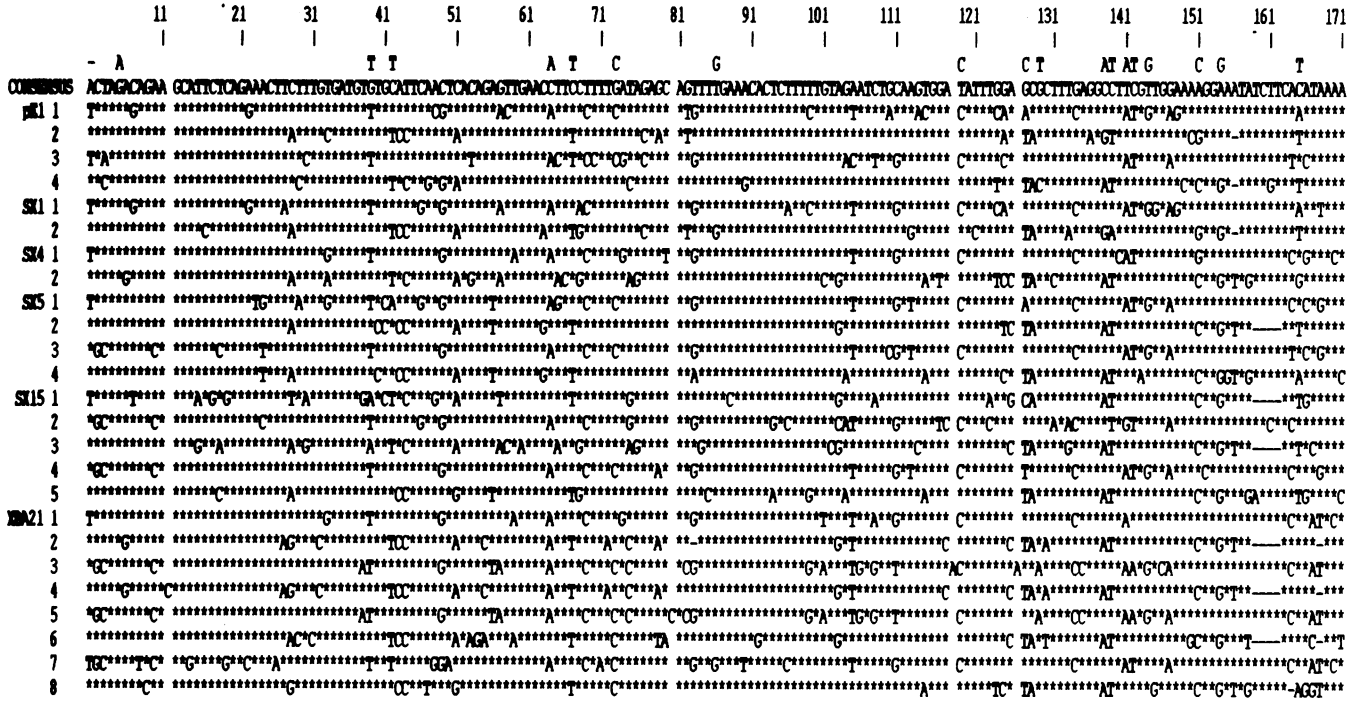

Figure 2. Nucleotide sequences of SX-subfamily clones. The sequences are shown relative to a human alphoid consensus sequence (11). Shown above the consensus are the next-most prevalent identities at positions where the base assigned did not meet the criteria used during derivation of the consensus (11). Asterisks indicate identity with the consensus, dashes indicate spaces inserted during alignment. Spaces were inserted in the consensus to accommodate extra positions present within individual clones.

Table 1. Percent sequence similarity among alphoid monomers from SX-subfamily clones

	pE1			SX1		SX4		SX5				SX15					XBA21							
	2	3	4	1	2	1	2	1	2	3	4	1	2	3	4	5	1	2	3	4	5	6	7	8
pE1 1	74	81	74	91	72	84	70	84	71	85	72	68	78	68	85	71	84	68	82	68	83	65	81	74
2		76	86	73	92	77	82	73	87	74	85	78	71	79	77	85	78	83	70	84	71	82	72	86
3			74	82	74	87	70	84	73	87	75	70	81	67	85	74	87	70	80	70	81	68	82	77
4				74	85	77	84	74	85	76	84	80	75	81	78	84	78	83	74	84	75	81	74	87
SX1 1					71	85	70	85	71	85	71	67	81	67	86	71	86	70	81	70	82	64	82	74
2						75	81	71	86	73	85	78	71	80	75	85	76	81	70	82	70	80	71	85
SX4 1							71	88	74	88	74	71	84	69	88	75	96	73	82	73	84	70	86	77
2								68	84	68	84	77	69	84	70	82	73	81	67	81	68	78	65	86
SX5 1									72	90	71	68	82	65	90	71	88	68	83	68	84	66	83	73
2										72	90	85	71	86	74	87	75	89	69	89	70	87	68	87
3											73	67	87	67	95	74	88	70	88	70	88	67	87	75
4												77	71	78	74	90	75	81	69	81	70	80	68	89
SX15 1													67	82	68	79	73	81	65	82	66	80	67	79
2														67	87	71	84	71	82	71	83	65	84	71
3															68	80	70	82	64	83	65	81	64	78
4																74	89	73	89	73	89	70	88	76
5																	75	81	70	81	71	81	68	89
XBA21 1																		74	85	74	86	71	87	77
2																			67	99	68	85	67	82
3																				67	99	63	82	70
4																					68	85	67	83
5																						64	84	71
6																							64	81
7																								70

the tetramer, SX5, and the pentamer, SX15, were about 95% similar (see below). These results suggest that sequences exhibiting high sequence identity are distributed throughout the various size classes of XbaI fragments.

We next compared SX-family monomers with monomers from the EcoRI dimer, and the pentamer, human subfamilies. Monomers from the EcoRI dimer family, which included the consensus sequence determined by Wu and Manuelidis (1), clones pa7d1, pa7t1 (8), and clones R1-107, -124, -117 (12), failed to exhibit an alternating high/low identity pattern with SX-family monomers (73% average identity), but did exhibit this pattern with each other (1,11,12). Furthermore, a consensus SX-subfamily dimer sequence lacked the 28 specific base positions which are characteristic of

the EcoRI subfamily consensus dimer (11), and exhibited at least 16 positions which are unique to the SX-subfamily consensus (work in progress). These results indicate that SX satellites and EcoRI satellites did not evolve from the same progenitor alphoid dimer unit. Therefore, although both satellites have dimeric repeat organizations, they represent distinct alphoid subfamilies. Similar results were obtained with monomers from p17H8 (7), a clone originating from chromosome 17 and belonging to the pentamer subfamily. These results demonstrate that the sequence characteristics exhibited by clones from the SX subfamily, specifically the presence of a dimeric repeat organization with dimers exhibiting about 84% sequence identities with other SX-family dimers, can be used to distinguish SX subfamily satellites from other human alpha satellite subfamilies.

Fine structure analysis of repeat organization within SX satellite clones

The data in Table 1 indicated that the clone XBA21, and the pairs SX4/XBA21, and SX5/SX15, contained regions exhibiting high sequence identities (>95%). We examined these regions more closely using a mismatch mapping method (7,14) where we aligned the homologous and adjoining regions, and analyzed the density of mismatches along the sequences. The boundaries of the homologous regions were apparent as increases in the density of mismatches along the sequences being compared, and could therefore be localized. The results of these analyses are summarized in Figure 3.

Alignment of the first six monomers of XBA21 with the last six revealed the presence of two dimers arranged in tandem (indicated by arrows above XBA21 in Figure 3A), beginning at about position 148 of the first monomer, whose sequences were 99% identical (hatched regions in Figure 3A). The sequences bordering the dimers exhibited about 84% sequence identity. The dimer, SX4, and the first dimer within XBA21, were found to exhibit about 97% sequence identity within the first 200 nucleotides (shaded regions in Figure 3A), and about 81% identity thereafter.

Alignment of the clones SX5 and SX15 led to the identification of an optimal repeat register within these clones, and which appeared to apply to all of the clones except pE1 and SX1.

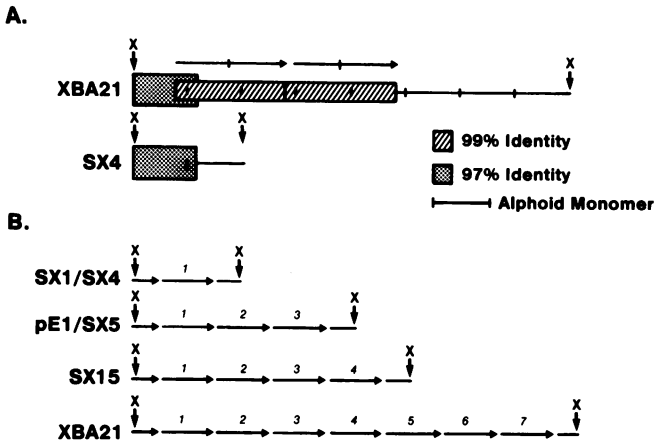

Figure 3. Repeat structure and organization within cloned SX satellite fragments. **A.** Regions exhibiting high sequence identity within clones SX4 and XBA21. **B.** Assignment of an optimal repeat register within SX satellite clones. Numbers above each complete monomer refer indicate the regions used in the comparisons presented in Table 3.

An optimal repeat register is the assignment of the starting position of repetitive units within a tandem array which maximizes the sequence identities among the units (7,14). The identities of the monomers within the clones were consistently found to be greatest when position 78 (relative to the XbaI site, see Figure 2) was chosen as the starting point of alphoid monomers within the clones. Figure 3B illustrates the positions of monomers in this register within the various cloned XbaI fragments. It is interesting that this register does not appear to coincide with the register of the tandem duplication of the dimers in XBA21.

These data indicate that the variation in size of XbaI fragments observed in this satellite subfamily did not arise solely by random mutation. The observation that XbaI fragments from different size classes contain regions having high sequence identities flanked by regions with lower identities, suggests that the diversity in XbaI fragment size can be at least partially accounted for by recombinations between members of this satellite subfamily.

DISCUSSION

We have defined a new subfamily of human alpha satellite DNA, which we refer to as the SX subfamily in this report. Sequence analysis of cloned fragments from this subfamily indicated that this subfamily has evolved from an ancestral alphoid dimer. We therefore define the SX subfamily as alpha satellite DNA evolving from this progenitor alphoid dimer. Sequence comparisons between clones from the SX subfamily and the EcoRI dimer subfamily indicated that these two subfamilies evolved from different progenitor dimers.

Alexandrov et al. (23) have reported a similarity between the in situ hybridization histograms produced with the SX-subfamily clone, pE1 (17), and histograms of the clones they used to define suprachromosomal family 2. It is therefore likely that the sequence data described here, and the hybridization data reported by Alexandrov et al., pertain to the same human alphoid subfamily. If this is the case, then the clones used to define suprachromosomal family 2 should exhibit the same sequence characteristics described above for our clones, and should produce similar hyridization patterns to StuI- and XbaI-digested total human genomic DNA as those presented in Fig. 1.

Sequences which cross-hybridize with human SX-family clones have been found to be abundant in the genomes of chimpanzee, gorilla, and orangutan (17 and CC, Ph.D. thesis, 1987), but not in the genomes of gibbons and lower primates. However, the restriction patterns of SX satellite DNA differ among chimpanzee, gorilla, orangutan, and human. These data indicate that the SX satellite subfamily evolved prior to the divergence of the great apes and humans, but that different repeat organizations have evolved in SX satellite DNA within the genomes of these species subsequent to their divergence.

Processes such as unequal exchange, excision-reintegration, and gene conversion likely play a significant role in the evolution of satellite DNA (16), causing homogenization and co-evolution within and between satellite clusters (22). Defining alphoid subfamilies is the first step towards understanding the evolution of alpha satellite DNA. Subsequent comparison of chromosomal locality, repeat organization, and sequence related-

ness of homologous subfamiles in the human and ape genomes may then lead to the determination of the precise mechanisms by which higher-order repeat organizations arise, the rate at which changes occur and become fixed within satellite domains, and manner in which subfamilies distribute to nonhomologous chromosomes (14).

ACKNOWLEDGEMENTS

The authors would like to thank R.S. Collins and S. Baak for help in maintaining the human XbaI library, R. Akella for help in sequencing, and E. Pequignot for many helpful discussions.

*Present address: Department of Pathology, Hahnemann University, Philadelphia, PA 19102, USA
+EMBL accession nos X14299–X14303 (incl.)

REFERENCES

1. Wu, J.C. and Manuelidis, L. (1980) J. Mol. Biol. **142**,363-386.
2. Maio, J.J. (1971) J. Mol. Biol. **56**, 579-595.
3. Donehower, L. and Gillespie, D. (1979) J. Mol. Biol. **134**, 805-834.
4. Singer, M.F. (1982) Intern. Rev. Cytol. **76**, 67-112.
5. Manuelidis, L. (1978) Chromosoma **66**, 23-32.
6. Willard, H.F., Kirby, K.D. and Sutherland, J. (1983) Nucl. Acids Res. 11, 2017-2033.
7. Waye, J.S. and Willard, H.F. (1986) Mol. Cell. Biol. **6**, 3156-3165.
8. Waye, J.S., England, S.B., and Willard, H.F. (1987) Mol. Cell. Biol. **7**, 349-356.
9. Darling, S.M., Crampton, J.M. and Williamson, R. (1982) J. Mol. Biol. **154**, 51-63.
10. Willard, H.F. (1985) Am. J. Hum. Genet. **37**, 524-532.
11. Waye, J.S. and Willard, H.F. (1987) Nucl. Acids Res. **15**, 7549-7569.
12. Jorgensen, A.L., Bostock, C.J. and Bak, L.A. (1986) J. Mol. Biol. **187**, 185-196.
13. Choo, K.H., Vissel, B., Brown, R., Filby, R.G. and Earle, E. (1988) Nucl. Acids Res. **16**, 1273-1284.
14. Waye, J.S. and Willard, H.F. (1987) J. Mol. Evol. **25**, 207-214.
15. Jorgensen, A.L., Bostock, C.J. and Bak, L.A. (1987) Proc. Natl. Acad. Sci., U.S.A. **84**, 1075-1079.
16. Brutlag, D. (1980) Ann. Rev. Genet. **14**, 121-144.
17. Gray, K.M., White, J.W., Costanzi, C., Gillespie, D., Schroeder, W.T., Calabretta, B. and Saunders, G.F. (1985) Nucl. Acids Res. **13**, 521-535.
18. Maniatis, T., Fritsch , E.F. and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci., U.S.A. **74**, 5463-5467.

20. Chen, E.Y. and Seeburg, P.H. (1985) DNA 4, 165-170.
21. Dale, R.M., McClure, B.A. and Houch, J.P. (1985) Plasmid 13, 31-40.
22. Dover, G.A. (1982) Nature 229, 111-117.
23. Alexandrov, I.A., Mitkevich, S.P. and Yurov, Y.B. (1988) Chromosoma 96, 443-453.