Unequal cross-over is involved in human alpha satellite DNA rearrangements on a border of the satellite domain

Tamara Mashkova[a,*], Nina Oparina[a], Ivan Alexandrov[b], Olga Zinovieva[a], Alina Marusina[a], Yuri Yurov[b], Marie-Helene Lacroix[a], Lev Kisselev[a]

[a] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov str., Moscow 117984, Russia*
[b] *National Research Center of Mental Health, Russian Academy of Medical Sciences, Moscow 113152, Russia*

Received 12 November 1998

Abstract It can be invoked from the theory of tandem repeat homogenization that DNA on a satellite/non-satellite border may carry sequence marks of molecular processes basic to satellite evolution. We have sequenced a continuous 17-kb alpha satellite fragment bordering the non-satellite in human chromosome 21, which is devoid of higher-order repeated structure, contains multiple rearrangements, and exhibits higher divergence of monomers towards the border, indicating the lack of efficient homogenization. Remarkably, monomers have been found with mutually supplementary deletions matching each other as reciprocal products of unequal recombination, which provide evidence for unequal cross-over as a mechanism generating deletions in satellite DNA.

© 1998 Federation of European Biochemical Societies.

Key words: Alpha satellite DNA; Human chromosome 21; Recombination junction; Unequal cross-over; DNA sequencing

1. Introduction

Alpha satellite DNA is located in the centromeric regions of all primate chromosomes and is composed of tandemly repeated divergent monomers (~171 bp) usually organized in higher-order repeat units of length and structure varying for different chromosomes (reviewed in [1,2]). The groups of closely related chromosome-specific subsets form several suprachromosomal families based on divergent single-monomeric, dimeric or pentameric ancestral repeats [3–5]. Each group was mapped to a number of human chromosomes.

Chromosome-specific DNA families exhibit high sequence similarity between higher-order repeat units within a single domain. It is thought that this similarity is maintained by unequal cross-over and/or gene conversion that homogenizes and spreads satellite sequences within and between chromosomes [6–10]. Unequal cross-over and gene conversion are also suggested as putative mechanisms for generating new variants of alpha satellite repeats [11–14].

However, unequivocal evidence in favor of a certain molecular mechanism appeared to be difficult to obtain. Although a case of gene conversion in alpha satellite has been described [13], unequal cross-over, which is supposed to be a major homogenization force, has not been convincingly demonstrated. A main problem is the lack of information about paired reciprocal products matching the schemes of unequal recombination.

Theoretical models predict poor homogenization in satellite domains bordering unrelated sequences [6]. Such regions may serve as 'archives' preserving molecular marks of recombination which would be masked by efficient homogenization in the middle parts of alpha satellite arrays. We looked for evidence of unequal cross-over in alpha satellite DNA adjacent to non-satellite domains in human chromosome 21. Studies of recombination events in this locus and comparisons to the only other reported alpha satellite/non-satellite junction region in human chromosome 7 [15] allowed us to demonstrate a pair of matching recombination products that fit an unequal cross-over mechanism.

2. Materials and methods

2.1. Materials

Restriction enzymes, *Tag* DNA polymerase, DNA polymerase Klenow fragment and T4 DNA ligase were purchased from SibEnzym (Novosibirsk, Russia) and used according to the manufacturer's instructions. *Exo*III and S1 nucleases were from Promega, USA. [α-^{33}P]dATP was produced by the Institute of Nuclear Physics, Obninsk, Russia.

2.2. pIA1 cosmid clone

Isolation of the cosmid clone pIA1 from the chromosome 21-specific cosmid library kindly provided by D.J. Hardy (St. Mary's Medical School, UK) has been described [16]. The library was screened by stepwise hybridization with alpha satellite probe aRI-6 specific for chromosomes 13 and 21, with probes BLUR8 and pHS35 containing Alu repeats, and with probe pHS94 containing a segment of the L1 repeat. A clone which hybridized intensively with all these probes was isolated and called pIA1 [16].

Further analysis (to be reported elsewhere) has shown that the cosmid insert contains a continuous 17 kb stretch of alpha satellite bordering about 20 kb of non-satellite, not highly repetitive DNA. The authenticity of this satellite/non-satellite junction was confirmed by PCR analysis of genomic DNA (not shown).

2.3. Subcloning and sequencing of the cosmid DNA

Fragments of the cosmid clone were *Eco*RI-subcloned into pUC19 or pBluescript SK(−) plasmids, which were used to transform bacterial DH5α cells. Further subcloning of the *Eco*RI fragments was performed using *Hin*dIII, *Xba*I, *Spe*I, *Sca*I, *Pst*I, *Dra*I, *Nco*I, and *Xho*I restriction endonucleases.

Plasmid DNA was sequenced on both strands using the Sanger method [17] with [α-^{33}P]dATP and *Taq*I polymerase or Klenow fragment. Long DNA fragments were sequenced using nested deletions generated in both directions with *Exo*III. Treatment with *Exo*III and S1, ligation of the derivatives, transformation, and screening were performed according to the Promega instructions. The sequences of 11 *Eco*RI fragments were aligned into a single continuous block of about 17 kb in the following order: pE11, pY9, pM3, pZ2, pE52, pA14, pY8, pA17, pJ1, pA3, pA1 (Fig. 1). The junctions of all fragments were verified by partial restriction enzyme digestion and direct cosmid DNA sequencing of junction regions.

*Corresponding author. Fax: (7) (095) 1351405.
E-mail: mashkova@genome.eimb.rssi.ru

0014-5793/98/$19.00 © 1998 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(98)01600-7

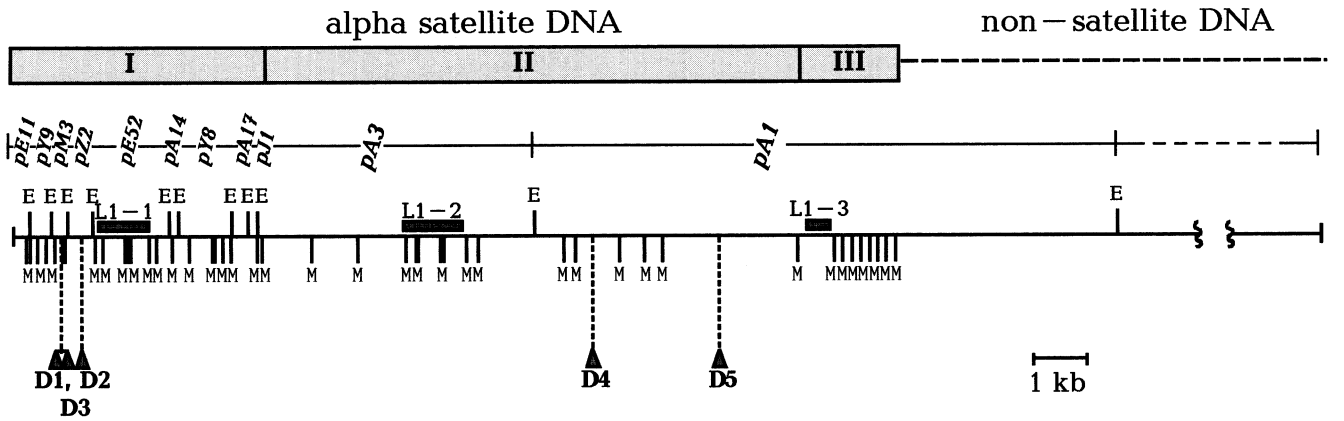


Fig. 1. A partial restriction map of pIA1 alpha satellite DNA adjacent to non-satellite sequences. Subdomains I, II, and III of alpha satellite block are shown at the top. Insertions of L1 elements (L1-1,2,3), *Eco*RI sites (E) and *Eco*RI subclones are marked above the map. Internal deletions (D1–D5) and multiple *Mae*I sites (M) are shown below. *Mae*I sites occur frequently within subdomain I, rarely within the alpha satellite part of subdomain II and are present in each monomer within subdomain III (the interruption of regularity corresponds to the L1-3 insert).

The GenBank accession number for the alpha satellite part of pIA1 DNA is AF105153.

2.4. Computer analysis

Nucleotide sequences were analyzed using VOSTORG (P.S. Morozov and Yu.G. Matushkin, Institute of Cytology and Genetics, Russian Academy of Sciences, Novosibirsk), DNA_SUN (A.A. Mironov, Institute of Genetics and Selection of Industrial Microorganisms, Moscow), and REVN [18] programs. Latest releases of GenBank and EMBL data bases were searched with the NCBI BLAST E-Mail system [19].

Alpha satellite sequence analysis including alignment, derivation of consensus monomers, calculation of consensus identity index and clustering of monomers into subgroups was performed as described [4]. The position in a consensus sequence was considered unambiguous if more than 50% of monomers had the same nucleotide in this position. Position 1 of the monomer was arbitrarily assigned to the *Bam*HI site in the X-specific alpha satellite repeat [20].

Neighbor-joining phylogenetic analysis [21] was carried out using the evolutionary modeling package TREECON [22] with statistical bootstrap optimization.

3. Results

3.1. Characterization of alpha satellite DNA on the border of satellite domain

We determined a complete nucleotide sequence of alpha satellite DNA in cosmid clone pIA1 partially characterized earlier [16]. The sequence encompassed 100 tandem monomers, 88 of which were full-length (~171 bp) and 12 truncated due to integration of L1 fragments or internal deletions. After multiple alignment to maximum homology each monomer was compared to 12 consensus monomers [5] derived for the five known suprachromosomal families (not shown). All monomers were attributed to suprachromosomal family 4 [4] characterized by a single-monomeric organization. Divergence of individual alpha satellite monomers from the consensus M sequence derived for family 4 was 4–16% (9% on average).

3.2. Sequence homogenization in the border region

In order to examine the organization of alpha satellite monomers in more detail, we compared their sequences pairwise. Divergences between full-length monomers vary from 5% to 28% (average 16%). No identical individual monomers were found. We also have not detected any periodically repeated highly similar monomers which shows the absence of higher-order repeated structures in this region.

The pairwise comparison of all monomers revealed a somewhat higher divergence towards the border of the alpha satellite block. The pattern showing these results was too large to be presented in full here. Fig. 2 depicts schematically the comparison of only 44 alpha satellite monomers bordering the non-satellite domain. The average divergence determined for the 22 monomers which are closer to the border is 16.6% versus 13.5% for the more distant 22 monomers. As is evident from Fig. 2, this increase in divergence could not be attributed to the influence of monomer deletions or L1 insertions in the region.

The absence of higher-order repeated structures and a gradient of divergence observed in our 100 monomer stretch indicated that homogenization of alpha satellite repeats near the

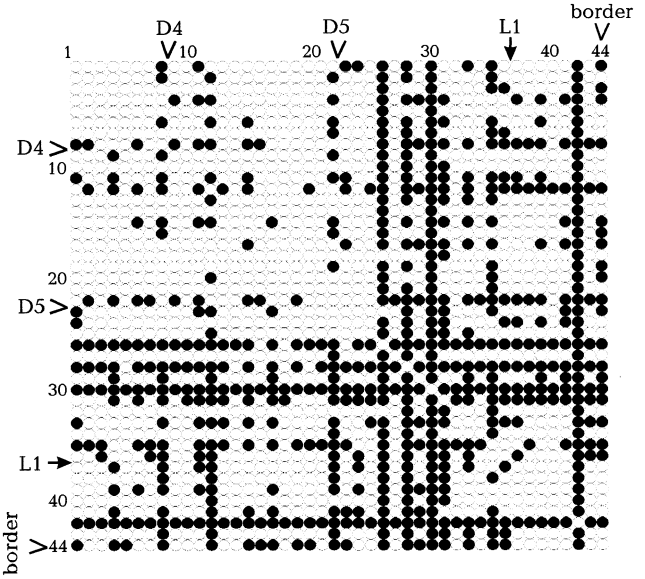
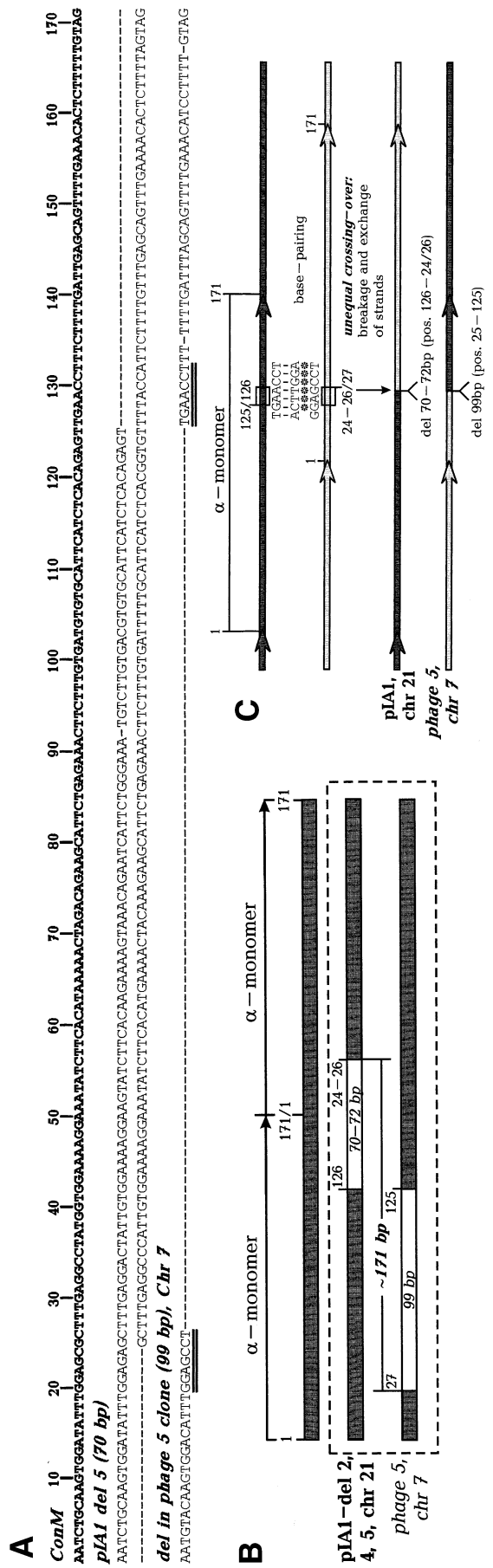


Fig. 2. Pairwise comparison of 44 alpha satellite monomers of pA1 DNA adjacent to the satellite/non-satellite border. Divergences of individual monomers (%) that exceed the average value are designated by black circles; those that are below or equal to the average value by empty circles. D4 and D5 deletions and the truncated L1 repeat interrupting the alpha satellite block are shown on both axes.



junction was limited and this restriction increased towards the border.

3.3. Sequence irregularities in pIA1 DNA

In the alpha satellite block of the pIA1 DNA we revealed five unusually large deletions within monomers interrupting a regular tandem organization. The positions of deletion boundaries were established by maximizing similarity to the consensus sequence of all alpha satellite monomers of clone pIA1 and in uncertain cases by taking into account not only a predominant consensus nucleotide in a particular position, but also other nucleotides frequently occurring in this position in other monomers (Fig. 3A). In some cases it was impossible to define precisely the beginning and the end of the deletions because of the high similarity or identity of sequences on both sides of deletions. Deletions 1 (58 bp) and 2 (72 bp) were found in subclone pM3 (positions 59–116 and 126–26, respectively), deletion 3 (65 bp) in subclone pZ2, positions 1–65 or 11–76 (alternative possible position) and two identical 70 bp deletions (positions 126–24) in subclone pA1 (del 4 and del 5). All these deletions were observed in several independently obtained subclones (2–4) of clone pIA1 and therefore were not cloning artifacts. Deletion 5 is shown in Fig. 3A.

Three truncated L1 elements were found in the alpha satellite part of pIA1 DNA. Assuming random distribution within the genome, L1 elements (50 000–100 000 copies per genome) should be separated by DNA segments of 30–60 kb. The relative abundance of L1 repeats in pIA1 DNA was unusual as the presence of only a few L1 elements was reported for other alpha satellite sequences [23–26]. All three L1 fragments differed by length (368 bp, 890 bp, and 1040 bp) and sequence (divergences of shared parts varied from 6.3 to 8.1%) and corresponded to the 3′ portion of L1 repeat (7–8% divergence with the consensus LINE 1 sequence [27]).

3.4. Recombination events in pIA1 DNA

The left third of satellite block up to monomer 27 contained nine *Eco*RI and multiple *Mae*I sites (Fig. 1). The latter ones occurred frequently also in the right terminal 10 monomers. However, alpha satellite monomers in the middle part had

←

Fig. 3. Deletions within alpha satellite monomers match each other as products of unequal cross-over. A: Nucleotide sequences of monomers with 70 bp (pIA1) and 99 bp (clone 'phage 5', chromosome 7) deletions. The consensus M alpha satellite 171 bp monomer derived for suprachromosomal family 4 is shown at the top. The short homologous sequences which can facilitate base pairing between misaligned strands are underlined. B: Schematic representation of mutually supplementary deletions. Intact regions are denoted by black boxes, deleted regions by empty boxes. The sizes of the deletions are indicated. The positions of the deletion ends are shown above the monomers. Almost identical deletions (del 2, 4, 5) are depicted collectively as 70–72 bp deletion with the ends in positions 126 and 24–26. The total length of the 70–72 bp and 99 bp deletions is equal to one monomer. They supplement each other as paired products of unequal cross-over. C: Hypothetical scheme of unequal cross-over producing 70–72 bp and 99 bp deletions. The nucleotide sequences of short homologous regions present in misaligned alpha satellite monomers and separated by 99 bp are shown. Possible pairing including a non-canonical TG pair is indicated by asterisks. Such pairing followed by breakage and strand exchange may lead to formation of the reciprocal recombinant products: with 70–72 bp deletions in the pIA1 DNA monomer and 99 bp deletion in 'phage 5' DNA monomer.

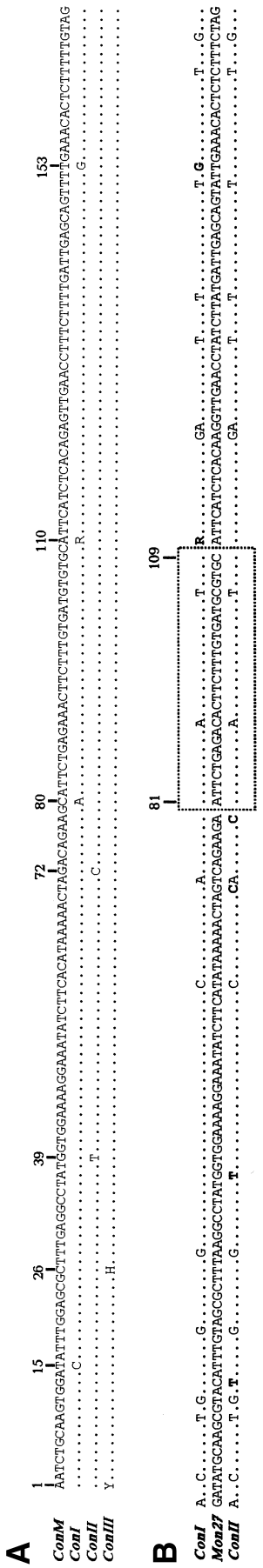


Fig. 4. Consensus monomers of pIA1 alpha satellite subdomains. A: Comparison of consensus sequences derived for subdomains CONI, CONII and CONIII. The consensus alpha satellite 171 bp monomer (M) derived for suprachromosomal family 4 is shown at the top. Only differences between consensus sequences are shown. Dots indicate identical base pairs. Each position was considered unambiguous if more than 50% of monomers had the same nucleotide at that position. R indicates A or G. B: Location of recombination exchange region in composite monomer MON27, the nucleotide sequence of which is shown in full. The location of the recombination window (boxed region) was determined by comparing the MON27 sequence to CONI and CONII consensus sequences. Bold-faced letters indicate consensus nucleotides that match neither the MON27 sequence nor the other consensus monomer. The size of the window is defined by the two nearest diagnostic nucleotides that identify the two parts of the sequences with different consensus monomers.
←

only one *Eco*RI and a few *Mae*I sites. This structure suggested the presence of three types of monomers within the stretch. Detailed analysis had revealed that indeed the alpha satellite block consisted of three distinct subdomains each of which matched its own consensus (Fig. 4A) and possessed a number of diagnostic nucleotide positions which identified the respective subset (Fig. 4 and Table 1). Nucleotide changes accumulated in certain diagnostic positions led to formation of *Eco*RI and *Mae*I sites (C→A in position 80 and C→G in position 72, respectively).

Most likely these three subsets of suprachromosomal family 4 monomers had once evolved independently being separated by some kind of intervening domains and were later joined by recombination between repeats misaligned by a whole number of monomers. We determined the putative recombination junctions in monomers 27 and 90 which divided the alpha satellite block into the three subdomains. The recombination window (see legend to Fig. 4B) can be located in monomer 27 to an interval of 30 bp (positions 81–109). The ends of monomer 27 flanking this window may be assigned to subdomains I and II by comparison with the respective consensus sequences. In monomer 90 (not shown) the window within which the recombination junction could be located was rather wide (positions 72–171). The large window size in this case was due to a higher similarity of monomeric types between which recombination had occurred.

4. Discussion

4.1. Less efficient homogenization of alpha satellite DNA on the edge of the satellite domain

Long centromeric alpha satellite arrays are composed of tandemly organized highly homologous higher-order repeat units. The alpha satellite DNA on the border of the satellite array described here has a number of notable features: the lack of higher-order repeated structures, single-monomeric organization and relative enrichment in L1 insertions and recombination junctions. Sequence comparisons of alpha satellite monomers clearly show the increase in divergence towards the border with non-satellite sequences as was anticipated in the homogenization model [6] and demonstrated for the short tandem repeats in the ribosomal DNA spacers of wheat and maize [28,29]. Here it is revealed for the first time in a centromeric satellite sequence.

A high rate of homogenization requires frequent recombination events which lead to fixation or extinction of mutations

Table 1
Comparison of nucleotide frequencies (%) at diagnostic positions of the different alpha satellite subfamilies

positions / consensus / nucleotide frequences, %	1	15	25	26	39	68	72	80	110	127	130	153
ConM	*A*	*T*	*G*	*C*	*G*	*A*	*G*	*C*	*A*	*G*	*C*	*T*
pIA1-I	· 80	C 81	GA 62/38	· 81	· 68	AT 52/43	· 72	AC 55/41	AG 48/39	GA 61/39	CA 57/26	GT 57/36
pIA1-II	· 89	· 96	GAC 54/20/18	· 86	TC 63/30	· 96	C 84	· 86	· 93	· 91	CA 63/24	TG 69/27
pIA1-III	TC 38/38	· 100	· 100	ATC 38/25/25	· 75	· 100	· 100	· 88	· 75	· 88	· 71	· 100
pTRA2	· 74	· 96	· 65	· 78	· 65	· 90	· 95	· 81	· 83	· 83	CA 62/30	· 90
pTRA4	· 87	· 100	GAC 45/29/23	· 83	· 67	· 80	CG 47/40	· 90	· 80	· 79	CA 62/30	· 90
HUMASA	· 100	· 100	· 100	· 100	· 100	AT 67/33	GA 67/33	AC 67/33	ATG 33/33/33	AGC 50/25/25	CA 75/25	TG 67/33
HUMASB	· 100	· 100	· 100	· 75	GA 50/50	· 100	· 100	· 100	· 100	· 100	· 75	TG 50/50
HUMASD	· 90	C 82	GA 60/25	· 91	· 82	AT 58/42	· 90	AC 45/45	AG 57/29	GA 50/45	CA 73/23	TG 57/21
HSASTLTD	· 100	· 100	N	· 67	· 67	AC 67/33	· 100	· 57	· 71	· 100	CA 71/29	G 100
HSASTLTM	· 57	· 83	N	· 83	· 86	· 86	· 86	· 86	AT 71/29	· 67	A 60	GT 50/50
HSASTLTN	AT 67/33	· 100	GA 67/33	CT 67/33	GA 67/33	· 100	· 100	· 100	· 100	· 100	· 100	· 100

Nucleotide frequencies (%) have been determined at the diagnostic positions of the three pIA1 subdomains and several alpha satellite subfamilies of suprachromosomal family 4 from chromosomes 21 (pTRA-2,4 and HUMAS-A,B,D) and 7 (HSASTLT-D,M,N). ConM indicates consensus monomer derived for suprachromosomal family 4. Nucleotides identical to those in corresponding consensus M positions are shown by dots. A few ambiguous consensus positions at which two or three dominant bases occur which were of diagnostic value have been added to those shown in Fig. 4A. N represents positions at which all four bases occur. The most related alpha satellite subsets which share obvious diagnostic positions are marked in a similar manner.

(including more complex ones such as deletions or rearrangements). Hence, recombination maintains the regular structure of satellite domains. On the edge of a satellite domain a given sequence segment has less chance to undergo recombination because it has a vast number of similar sequences only on one side. Consequently, a theoretical prediction confirmed in this work is that the border sequences should be poorly homogenized. On the other hand we demonstrate here that the restricted homogenization in this region made it possible to identify the marks of rare recombination events of another kind, the ones disrupting the regular structure of alpha satellite domains. Normally, most of these marks would be homogenized to extinction and lost. Paradoxically, the area of restricted recombination is the only place where recombination

can be actually detected at the sequence level. In other loci it can mostly be deduced from the homogeneity of repeated units. Thus, border domains are likely to be areas where actual recombination events can be traced and their mechanisms and parameters established.

4.2. Evidence for unequal cross-over

The alpha satellite pIA1 DNA contains a number of rearrangements including five unusually large deletions (Fig. 2) which may have arisen due to unequal cross-over, gene conversion or double-strand gap repair [10,11,30]. Unequal cross-over between misaligned alpha satellites accounts well for the generation and homogenization of tandem repeats [11,14]. When misaligned repeat arrays are shifted by the whole number of monomers, cross-over results in two new sequences: one with a deletion and the other with a duplication of a certain number of monomers. If they are shifted by a fractional number of monomers, cross-over leads to products with altered monomer length. In the latter case unequal cross-over is thought to be initiated by base pairing between short homologous regions present within monomers. Such short imperfect homologies are sufficient to ensure recombination [31]. Numerous direct repeats that may provoke recombination events are common in alpha satellite monomers including monomers of the M type typical of suprachromosomal family 4 to which the sequences described in this work belong.

We observed short homologies flanking the large deletions found in this work in both the consensus M sequence and actual sequences of type M monomers. The 9 bp imperfect direct repeats separated by 58 bp in positions 54–62 and 113–120 of consensus M providing for 8 bp pairing may explain the formation of deletion 1. Direct repeats flanking the 65 bp deletion (del 3) cannot be found in the consensus sequence, but do occur in the individual monomers in the vicinity of this deletion. For example, in the 3rd, 13th, 17th, and 21st monomers of pIA1 two regions of 10 bp imperfect homology spaced by 65 bp in positions 1–10 and 66–76 approximately correspond to the position of deletion 3.

Base pairing of two short highly homologous sequences separated by X bp leads to formation of the X bp long deletion or (171−X) bp long monomer. The second reciprocal product of such a recombination would be an insertion, which may be interpreted as an incomplete monomer with X bp length and a deletion of (171−X) bp. The combined lengths of the mutually supplementing deletions in two strands should be equal to one full-length monomer (171 bp). Both of these recombination products may be either fixed or eliminated by homogenization. While a number of small 3–4 bp deletions have been fixed and become characteristic of certain types of alpha satellite monomers, fixation of large deletions has never been observed in efficiently homogenized alpha satellite domains. Only very rare instances of large single-copy deletions were found [15,32]. The lack of available data did not make it possible to observe the predicted occurrence of both reciprocal products of unequal cross-over and thus verify the mechanism of deletion generation.

Multiple deletions observed in the poorly homogenized pIA1 alpha satellite stretch include two identical 70 bp deletions (positions 126–24) and an almost identical 72 bp deletion (positions 126–26). However, the highly homologous sequences that may be involved in the generation of these deletions are absent in the alpha satellite consensus. Short imperfect homologies can be found in the actual sequences of some individual M monomers highly related to pIA1 alpha satellite in positions 21–27 and 126–132 which approximately correspond to the ends of del 1, 4 and 5 (Fig. 3). Perhaps pairing of such short sequences may facilitate bringing together two DNA molecules. Also, some other factors, for example local chromatin structure, may influence the recombination process.

The deletion size in a product reciprocal to the 70–72 bp deletion should be 99–101 bp (positions 25/27–125). Remarkably, such a 99 bp deletion is present in alpha satellite DNA of clone 'phage 5' spanning the satellite/non-satellite border from human chromosome 7 [15].

These 70–72 bp and 99 bp deletions may possibly have arisen from reciprocal exchange between the misaligned alpha satellite sequences of chromosomes 21 and 7 (Fig. 3C). Alternatively, and more probably, the recombination processes generating such deletions proceed in a similar manner even in non-homologous chromosomes, if they share similar sequences. It should be noted that both deletions were found in border alpha satellite regions, which belong to suprachromosomal family 4 characterized by single-monomeric organization, the lack of higher-order repeated structures and binding sites for centromeric protein CENP-B [33]. Both regions contain multiple L1 insertions and alpha satellite rearrangements. However, the alpha satellite sequences determined in chromosome 7 are short and give insufficient information for more detailed comparisons. The finding of the two deleted alpha satellite monomers which could represent the mutually supplementing recombination products is the first instance of such matching structures known to us. If more such cases are found they will constitute solid proof of the involvement of unequal cross-over in the evolution of satellites.

4.3. Three subsets of family 4 alpha satellite repeats

Alpha satellite pIA1 DNA belongs to suprachromosomal family 4 and contains three related subsets which differ in several diagnostic positions (see Fig. 4A). Earlier, alpha satellite subsets of family 4 pTRA-1,2,4 [32] and pN [34] were revealed in the centromeric region of chromosome 21. We have compared the three subsets found in pIA1 to other alpha satellite sequences of suprachromosomal family 4 using two different approaches. The matching of diagnostic positions was evaluated by calculating the index of consensus identity and the overall similarity was addressed by phylogenetic analysis (see Section 2). Both methods gave similar results. Table 1 illustrates the first approach and shows the frequency of nucleotides at diagnostic positions. Comparison of consensus and actual sequences as well as phylogenetic analysis shows that subset I of pIA1 DNA is more similar to subfamily pN (HUMASD) than to others, subset II to subfamily pTRA-4, and subset III to the border sequence of chromosome 7 (HSASTLTN) which is also identical to consensus monomer M. Thus, the diagnostic positions observed in the three alpha satellite subsets of pIA1 are common in other family 4 domains on human chromosome 21, which is thought to contain several separate stretches of such sequences.

We have shown here that investigation of the complex organization of alpha satellite DNA on the border of the satellite domain where a sequence homogenization process is less efficient makes it possible to find the sequence imprints of

recombination events that otherwise would be hidden by homogenization.

Acknowledgements: This work was supported by the Russian National Human Genome Program and by Russian Foundation for Basic Research (Grant 98-04-49118).

References

[1] Warburton, P.E. and Willard, H.F. (1996) in: Human Genome Evolution (Jackson, M., Strachan, T. and Dover, G., Eds.), pp. 121–145, BIOS Scientific, Oxford.
[2] Murphy, T.D. and Karpen, G.H. (1998) Cell 93, 317–320.
[3] Alexandrov, I.A., Mashkova, T.D., Akopian, T.A., Medvedev, L.I., Kisselev, L.L., Mitkevich, S. and Yurov, Y.B. (1991) Genomics 11, 15–23.
[4] Alexandrov, I.A., Medvedev, L.I., Mashkova, T.D., Kisselev, L.L., Romanova, L.Y. and Yurov, Y.B. (1993) Nucleic Acids Res. 21, 2209–2215.
[5] Romanova, L.Y., Deriagin, G.V., Mashkova, T.D., Tumeneva, I.G., Mushegian, A.R., Kisselev, L.L. and Alexandrov, I.A. (1996) J. Mol. Biol. 261, 334–340.
[6] Smith, G.P. (1976) Science 191, 528–535.
[7] Southern, E.M. (1975) J. Mol. Biol. 94, 51–69.
[8] Dover, G. (1982) Nature 299, 111–117.
[9] Baltimore, D. (1981) Cell 24, 592–594.
[10] Nagulaki, T. and Petes, T.D. (1982) Genetics 100, 315–337.
[11] Warburton, P.E., Waye, J.S. and Willard, H.F. (1993) Mol. Cell. Biol. 13, 6520–6529.
[12] Waye, J.S. and Willard, H.F. (1986) Mol. Cell. Biol. 6, 3156–3165.
[13] Alexandrov, I.A., Mashkova, T.D., Romanova, L.Y., Yurov, Y.B. and Kisselev, L.L. (1993) J. Mol. Biol. 231, 515–520.
[14] Mashkova, T.D., Akopian, T.A., Romanova, L.Y., Mitkevich, P., Yurov, Y.B., Kisselev, L.L. and Alexandrov, I.A. (1994) Gene 140, 211–217.
[15] Wevrick, R., Willard, V.P. and Willard, H.F. (1992) Genomics 14, 912–923.
[16] Mashkova, T.D., Tyumeneva, I.G., Zinov'eva, O.L., Romanova, L.Yu., Jabs, E. and Alexandrov, I.A. (1996) Mol. Biol. (Moscow) Engl. Edn. 30, 617–625.
[17] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[18] Shepelev, V.A. (1991) Biopolimery Kletka (Moscow) 7, 22.
[19] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[20] Waye, J.S. and Willard, H.F. (1985) Nucleic Acids Res. 13, 2731–2743.
[21] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[22] Van de Peer, Y. and De Wachter, R. (1994) Comput. Appl. Biosci. 10, 569–570.
[23] Potter, S.S. and Jones, R.S. (1983) Nucleic Acids Res. 11, 3137–3152.
[24] Jackson, A.S., Mole, S.E. and Ponder, B.A.J. (1992) Nucleic Acids Res. 20, 4781–4787.
[25] Laurent, A.M., Puechberty, J., Prades, C., Gimenez, S. and Roizes, G. (1997) Genomics 46, 127–132.
[26] Lee, C., Wevrick, R., Fisher, R.B., Ferguson-Smith, M.A. and Lin, C.C. (1997) Hum. Genet. 100, 291–304.
[27] Hutchison, C.A., III, Hardies, S.C., Loeb, D.D., Shehee, W.R. and Edgell, M.H. (1989) in: Mobile DNA (Berg, D.E. and Howe, M.M., Eds.), pp. 593–617, American Society for Microbiology, Washington, DC.
[28] Barker, R.F., Harberd, N.P., Jarvis, M.G. and Flavell, R.B. (1988) J. Mol. Biol. 201, 1–17.
[29] McMullen, M.D., Hunter, B., Phillips, R.L. and Rubenstein, I. (1986) Nucleic Acids Res. 14, 4953–4969.
[30] Orr-Weaver, T.L. and Szostak, J.W. (1985) Microbiol. Rev. 49, 33–58.
[31] Rubnitz, J. and Subramani, S. (1984) Mol. Cell. Biol. 4, 2253–2258.
[32] Vissel, B. and Choo, K.H.A. (1992) J. Mol. Evol. 35, 137–146.
[33] Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) J. Cell. Biol. 109, 1963–1973.
[34] Ikeno, M., Masumoto, H. and Okazaki, T. (1994) Hum. Mol. Genet. 3, 1245–1257.