Volume 14 Number 19 1986 Nucleic Acids Research

A human Y-chromosome specific repeated DNA family (DYZ1) consists of a tandem array of pentanucleotides

Yutaka Nakahori, Kounosuke Mitani, Masao Yamada and Yasuo Nakagome

Department of Congenital Abnormalities Research, National Children's Medical Research Center, 3-35-31 Taishido, Setagaya-ku, Tokyo 154, Japan

Received 1 August 1986; Accepted 9 September 1986

ABSTRACT

We have determined the complete nucleotide sequence of a 3564 bp EcoRI fragment which represents a major component of the human Y-chromosome specific repeated DNA family (DYZ1). Sequencing result showed a tandem array of pentanucleotides after five nucleotides were inserted or deleted at four positions. 229 out of the 713 pentanucleotides were TTCCA, and 297 were its single-base substituents. Southern hybridization analyses of male genomic DNAs showed that several endonuclease cleavage sites were located at intervals of 3.56kb in the DYZ1 locus. This indicates that the DYZ1 repeated DNA family evolved and expanded by unequal crossovers which occurred at distances of 3.56kb. As there is a uniformly distributed array of pentanucleotides on this locus, it is not a sequence homology that determines the distance of unequal crossovers. A higher order of chromatin structure may be involved in the determination of distance in unequal crossovers.

INTRODUCTION

The long arm of the human Y chromosome has several characteristic features. No functional genes have been mapped on it (1). It doesn't participate in pairing with any other chromosomes during meiosis. A major part of the long arm is heterochromatic, which fluoresces brightly when stained with quinacrine. The length of the long arm is extremely variable within the normal male population.

A 3.4kb Y-specific repeated DNA family (DYZ1) was first reported by Cooke as a 3.4kb band in HaeIII digests of male genomic DNA (2). It was later shown that this repeated DNA family was a major component of the heterochromatic region of the long arm (3). DNA fragments of this region have been molecularly cloned (4) and used as probes to detect the Y-chromosome, i.e. in cytogenetical studies of Y chromosome translocation (5) and in determination of the male cell fraction of a mosaic (6).

Repeated sequences have different properties from unique sequences and are sometimes fractionated by physical methods. The satellite III fraction

was isolated as having a discrete bouyant density (7). The total DNA of satellite III was shown to hybridize in situ heavily to the long arm of the Y-chromosome and weakly to several autosomes (8). Conversely, the male-specific 3.4 kb HaeIII DNA hybridized to satellite III DNA (3). However, not all the DNA in satellite III originated from the Y-chromosome, because the satellite III fraction was recovered from both male and female DNA. A DNA fragment cloned from the satellite III contained dispersively repeated sequences and was shown to originate from chromosome 1 (9). Three other plasmids which carry satellite III variants were cloned and sequenced (10). All the inserts in the three plasmids had very similar, but not identical, sequences based on tandem repetition of sequences related to the pentanucleotide TTCCA. However, two of the plasmids hybridized to the 3.4 kb HaeIII DNA and the other did not. It is not clear why some of the similar sequences hybridized and others did not. A sequencing analysis of a Y-specific repeated DNA family would be helpful to further our understanding.

To address the question of why the long arm of the Y-chromosome has characteristic features, we determined the complete nucleotide sequence of a DNA fragment which carried one unit of the Y-specific repeated DNA family. Here we report the results of nucleotide sequencing and discuss the evolution mechanism of the repeated DNA family.

METHODS

Molecular cloning. Escherichia coli K12 strains DH1 and HB101 were used as hosts for plasmid construction. General procedures used for bacterial transformation, plasmid DNA isolation, Southern blotting, nick translation and hybridization with radio-labeled probes were described previously (11,12). The restriction enzymes and DNA modifying enzymes used were products of Takara Co. (Kyoto, Japan) and used under conditions recommended by the supplier. HindIII digests of lambda phage DNA were used as a molecular weight standard.

Cloning of pHY10. High molecular weight DNA was prepared by phenol extraction from cultured fibroblasts of a normal male, digested with EcoRI, and fractionated through a 4% polyacrylamide gel by electrophoresis. DNA fragments having an apparent size of 3.4 kb were recovered from the gel by electroelution, purified by passing through a DEAE-Sephacel column, and then ligated to EcoRI-digested pBR325 DNA. After transformation of HB101 cells with the ligation mixture, ampicillin-resistant and chloramphenicol-

sensitive colonies were selected, and subjected to screening by colony hybridization. A probe used in the colony hybridization was a ^{32}P-labeled 3.4kb HaeIII fraction which was isolated from the same DNA by a method similar to that used for the 3.4 kb EcoRI fraction. Colonies which gave a strong signal were selected and the length of the inserts was measured by agarose gel electrophoresis after mini-scale plasmid preparation. A plasmid (pHY10) was chosen of which the insert had the same size as the male-specific HaeIII band.

Isolation of deletion plasmids. The EcoRI fragment carried by pHY10 was subcloned on the EcoRI site of pBR322 resulting in the formation of two plasmids (pHY10R & pHY10L), which carried the inserts in opposite orientation. Both plasmids had one HindIII site in the pBR322 vector sequence. Plasmid DNAs were linearized by HindIII digestion and digested with Bal31 at 30°C for various incubation periods in a high salt buffer (11). The DNA fragments were recircularized with ligase after insertion of HindIII linker (12mer, Takara). Thus, a series of plasmids (pHYR series) were obtained from pHY10R where various deletions started from the right EcoRI site in the restriction map indicated in Fig. 1, while another series of plasmids (pHYL series) were obtained from pHY10L with various deletions starting from the left EcoRI site. After cutting these plasmids with HindIII, the 3'end was labeled with ^{32}P-α-dATP and Klenow enzyme. The DNA was then digested with EcoRI and the insert was recovered by electroelution. Nucleotide sequences were determined by the Maxam-Gilbert method (13). We usually read a sequence ladder between 30 and 300 nucleotides from the labeled end in each plasmid except for the sequences from both EcoRI sites in the original plasmids (pHY10R and pHY10L). Nucleotide sequences were analyzed by a microcomputer using a software package from SDC Co. (Tokyo, Japan).

In situ hybridization. Peripheral blood leukocytes were cultured and

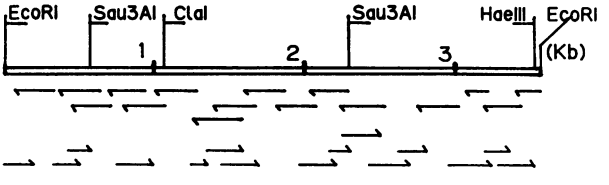

Fig. 1. Restriction map of the pHY10 insert and sequencing strategy. An arrow indicates the direction and region of the sequence which was obtained from a plasmid carrying deletion.

harvested by the conventional methods. After hypotonic treatment and fixation, cells were dispersed to glass slides by the air-dry method. In situ hybridization was carried out according to Harper et al. (14).

RESULTS

Cloning of a copy of the Y-specific repeated DNA family. It is known that the Y-specific repeated DNA family is mostly recovered as 3.4 kb DNA fragments by digestion of male DNA with either EcoRI or HaeIII. To clone a copy of the repeated DNA family, an enriched 3.4 kb EcoRI fraction of male DNA was used as a source material for inserts in ligation, while a 3.4 kb HaeIII fraction was used as a probe for screening as described in Materials and Methods. Although the exact length of the cloned fragment was 3564 bp as described below, we still use '3.4 kb' to indicate the unit size of the repeated DNA family to be consistent with previous reports.

The pHY10 insert was confirmed to be highly specific to the long arm of the Y-chromosome by Southern hybridization (described below) as well as in

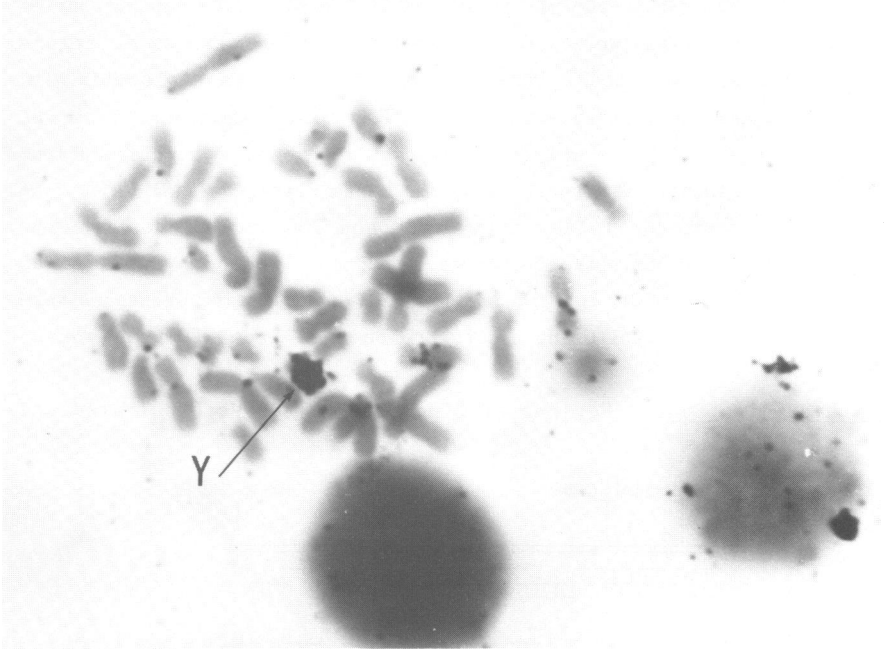

Fig. 2. In situ hybridization of the pHY10 insert to a male chromosome indicating that the insert is specifically hybridized to the long arm of the Y chromosome.

situ hybridization (Fig. 2). To estimate the copy number of the insert, male DNA carrying various lengths of the Y-chromosome were analyzed by Southern hybridization including the purified plasmid DNA as a standard. Based on densitometry of the autoradiogram, smaller and longer Y-chromosomes carried 800-1000 and 3000-5000 copies of the insert, respectively.

Nucleotide sequencing. Preliminary studies revealed that the pHY10 insert was highly repetitive and contained few available restriction enzyme sites for sequencing. Therefore, we decided to take a strategy including a construction of a series of deletion plasmids with Bal31 exonuclease for nucleotide sequencing. Nucleotide sequences were determined from the rejoining site where a synthetic linker was inserted. Plasmid isolation was described in Materials and Methods, and the sequenced region in each plasmid as well as relevant restriction sites are described in Fig. 1.

The complete nucleotide sequences (3564 bp) of the pHY10 insert are shown in Fig. 3. The fragment is apparently composed of a tandem array of pentanucleotides. To prove this, the frequency of nucleotide species which appeared in 5 different positions in the pentanucleotides (5n+m) was counted after a minor adjustment by insertion of spaces at 4 positions (after nucleotide numbers; 902, 1129, 1398 and 1959). Nucleotides of T, T, C, C and A are dominant in the first, second, third, fourth and fifth position of the pentanucleotides, respectively, as shown in Table 1. This is consistent with the fact that TTCCA is the most frequently appearing pentanucleotide in the fragment followed by TTCGA (Table 2). 229 out of the 713 pentanucleotides were TTCCA, and 297 were its single-base substituents.

Comparison of the sequence with other known sequences. Nucleotide sequences of satellite III were reported by analyses of clones PPD9, PPD17, PPD18 (10) and HS3 (9), whose characteristic features are briefly described in the discussion section. The pHY10 insert had homologies with PPD9, PPD17, PPD18 and HS3 at the maximum levels of 80 %, 35%, 74% and 45%, respectively, when whole nucleotide sequences were compared without phase adjustment. The homologies increased to 98 % for PPD9 and PPD19, and to 70-75 % for PPD17 and HS3, when small segments (40 nucleotides) of both sequences were compared. Structures like transposon terminus and core sequences (15) were not detected in the pHY10 insert.

Restriction cleavage sites in DYZ1 DNA family. Computer analyses of the nucleotide sequences indicate that there are 64 Taq I sites and 74 HinfI sites in the pHY10 insert. Only one single-base mutation is enough to

```
TTCCATTCCA TTCCAATCCA TTCCTTTCCT TTCGCTTGCA TTCCATTCTA TTCCCTTCTA CTGCATACAA    70
TTTCACTCCA TTCGTTCCCA TTCCATTCAA TTCCATTCCA TTCAATTCCA TTCCATTTGT TTCCATTCTC   140
TTCGATTCCA TTTCTTTATA TTCCATGCCA TTCGATTCCA TTCTATTGGA TTGCATTACA TACGTGTTCA   210
TTCCATTCCA GACCATTCCA TTTGACTCCA TTCCTTTCGA GCCCTTTCAA TTTGAGTCCA TTCCTTTCCA   280
GTCCATTTCA CTCCAGTCCA TTACTATCCA TTCCATACCA TTCCATCCCA TTCCATTCCA TTCCATTCCA   350
TTCCATTCCA TTGCATTCCA TTCCATTCCA TTCCATTCCA TTCCATTCCA TTGCACTGCA CTCCATTCCA   420
TTACATTCTA CTCTATCTGA GTCGATTTTA TTGCATTAGA TTCTATTCCA TTGGATTACT TTCCATTCGA   490
TTACATTCCA TTCATGTACA TTCCATTCCA GTCAATTACA TTCGAGTTCA TTACATTACA TTCCAGTATA   560
TTCCATTGTA TTCGATCCCA TTCCTTTCAA TTCCATTTCA TTCGACTCCA TTATATTCGA TTCCATTCCA   630
CTCGAATCCA TTCCATTAGA GGACATTCCA TTCCAATGCA TTCCATTCCA TTCCATAGCA TTCCATTGCA   700
TTCGATTCCA TTCCATTTGA TGCCATTCCA TTTGATGCCA TTCCATGACA TTCCATTCCA TTCGAGTCCG   770
TTCCGTTCCA ATTCATTGCA TTCCGTTTCA TGAAATTCGA GTCCTTTCCA GTACATTTCA TTCCAATCCC   840
ATCCAATCCA ATCTACTCCA TTCAATTCCT TTCCATTCCA TTTGATTTGA TTCCATTGAT TT-GATTCCA   909
TTCAGTTTGA TTCCATTCCG TGAAATTTCG TTCCATTCTA TTCTATTACA TAACTTTCCA TTCAATTCCA   979
TTCCATTTCA TTTCAGTCCA TTCGCTTCCT TTCCTTTCGA TTCAATTCCA TTTGATTCCA CTCCATTCTA  1049
TGCAATTTCA TTCCAATCGA TTCAATTCCA TTCGATGACA TTCCTTTCGT TTCCATTCCA TTCGAGTCCA  1119
TTCAATTTGA --GCATTCGT GTCCATTCTA TTCGAGTCCA TTCCATTACC GTCTATTCTA TTCCCTTCCA  1187
TTCCTGTTGA TTCAATTTCA TTCCCTTCCA TTCGATTCCT TTCCATTCGA TTCCATTCCT TTCCATTCCA  1257
TTCCATTCGT TCCCATTCCA TGTGATTTCA TTCCATTCCA GTCCATTATA TTCGAGTCCA CTCCACTCCA  1327
TTCTATTACA TTCAATTCCT TTTGAGTCCG TTCCATAACA CTCCATTCAT TTCGATTCCA TTTCTTGACA  1397
G----TTTTC TTCCATTTTA TTCCATTCCG TTCGATTCCA TTCCATTCGA TTGCATTCCA TTCGAATCCT  1463
TTCCATTCCA TTTCATTCCA TTCCTTTCTA TTCCATTCCA TTTCATTCGA TTTGATTCCA TTCTGTTCTA  1533
TTCCATTCAA TTCTTTTTCA TTCCATTCGA ATCCTTTCTA TTGCAGTCCA TTCCATTCGA GTCCATTCCA  1603
ATCCCTTCCA TTCCATTCCA TTACAGTCCA TTCCAATAGA TTCCATTCCT TTGCCTTCCA TTCGAATCCA  1673
TTCCATTCTA GTCCATTCCA TTTGAGTCAA TTCCATTCCA TTCCATTCTA TTCCTTTCCA ATCCATTCGA  1743
TTCCATTCGA TTCAATTCCA TTTGATTCTC TTTCATTCTA TTTTATTCCA TGCCATTTTA TTGCGTTGCA  1813
TTCCATTCCG TTTGATTCCA GTCCATTCAA GAAAGTTCCA TTCCAGTCCA TTGCTTTCCA GTCCATTCCA  1883
TTCCACTCTT GTCTATTCCA CTCCATTCCT TTCCATTCCA TTCCATACTA TTCCATTCCA TTCCTTTGCA  1953
TTCCGT---- TTCCAATCTA TTCGAGTCCA TTGCATTCCA GTCCAATCCA TTCCATTACA TTCCTTTTGG  2019
TTCCCTGCCA GTCGATTGCA TTGCATACTA GACCATTCCA AACCAGTCCA TTCCATTCTA TTTCAACACT  2089
TTCCATTCCA CTCTGTTCGA GTCCATTCCA TTCCAGTCCA TTTAATTCAA GGGCATTCCA TTCCATTCCA  2159
GTCCATTTCA TGTTATTCCA TTCCATTCAA TTCCATTCCA GATGATTCCA TTCCATTCTA TACCATTGCT  2229
CTCTGTTCCA TTCCATTCCA TCTGTCTCCA TTCCTTTCGT TTCGATTCCT TTCCATTCCA TTCCATTACA  2299
TTTGATCCTA TTTTATTAAA TTGCATTCTA TTCGAGTGAT TTCCCTTCGA GTCCTTTCCA TTCAATTCCA  2369
TTCCATTCTA TTCCATTCCT TTGGATTCCA TTCCATTCCG TTCCGTTCAC ATCAATTCCT TGTGATTCCA  2439
TTACATTCGA TTTCTTGCCA TTCGATTCCA TTCCTTTTGA CTCCATTTCA TTCGATTCCA TTCCATTCCA  2509
TTAATTTCCA TTCCATTCGA GACCTTTCCA TTGCAGTCTT TTCCCTTCGA GTCCATTCCG TTCGATTCCC  2579
TTCCATTCGA TTCCATTCCA TTGGAGTCCG TACCAGTCGA GTCCATTCTA TTCCAGTCCA TTAGTTTCGA  2649
CTCCATTGCA TTCGAGTGCA TTCCATTCCG TGGTTGTCCA TTCCATTCCG TTTGATGCCA TTCCATACGA  2719
TTCCATTCAA TTCGAGACCA TTCTATTCCT GTCCATTCCT TGTGGTTCGA TTCCATTTCA CTCTAGTCCA  2789
TTCCATTCCA TTCAATTCCA TTCGACTCTA TTCCGTTCCA TTCAATTCCA TTCCATTCGA TTCCATTTTT  2859
TTCGAGAACC TTCCATTACA CTCCCTTCCA TTCCAGTGCA TTCCATTCCA GTCTCTTCAG TTCGATTCCA  2929
TTCCATTCGT TTCGATTCCT TTCCATTCCA GCCCATTCCA TTCCATTCCA TTCCTTTCCT TTCCGTTTCA  2999
TTAGATTCCA TTGCATTCGA TTCCATTCAA TTCAATTCCG TGCTATTCAA TTTGATTCAT TTCCATTTAA  3069
TTCCATTCCA TTAGATTCCA TTCCGTACGA TTCCATTCCT TTTGAATCCA TTCCATTGGA GTCCATTCAC  3139
TTCCAGAACA TTCCATTCCA GTCGAATCCA TTCGAGTACA TACCATTAAA GTTCATTACA TTCTAATACA  3209
TTCCATTCCA TTGCATTCCA TTCCATTCCA TTAGATGCCA TTCGATTCCA TTCCATGCCA AATCATTGCA  3279
TTCCTTTCCA TTCCGTTCCT ATCAATTCTA TTCCATTCGA TTTAGTTCGA TTCTATTCAC TTCCATTCCA  3349
TTCGATTCCA GTCCATTGGA GTCAATTCCT TTCGACACCC AGCCTTTCCA GTCAATGATT TTGGATTCCA  3419
TTTTTTTGCA TTCCATTACA TTCTATGACA TTCGATTCCG TTTCATTGCA TTCCATTCCA TACATTTTTA  3489
TTCCATTCGA GACCGTAGCA TTCCACTTTA TTCCAGGCCT GTCCATTACA CTACATTCCC TTCCATTCCA  3559
ATGAA                                                                        3564
```

Fig. 3. Nucleotide sequence of the pHY10 insert. Nucleotide number starts at the middle of the recognition sequence of the left EcoRI site, and ends at the center of the right EcoRI site in Fig. 2. To show the pentanucleotide array clearly, insertions are made at 4 positions, which are indicated by a bar. Note that the insertion of four nucleotides at one point is equivalent to one nucleotide deletion in order to form the pentanucleotide array.

Table 1. Frequency of nucleotides that appeared in the 5n+m position in the pHY10 insert.

	5n + 1	5n + 2	5n + 3	5n + 4	5n	total
G	74	27	47	118	32	298
A	28	27	52	52	583	742
T	585	648	71	65	74	1443
C	27	10	542	478	24	1081
total	714	712	712	713	713	3564

The frequency was counted after phase adjustment by insertion.

generate the recognition sequence of TaqI (TCGA) and HinfI (GANTC) from the tandem array of the TTCCA pentanucleotides. On the other hand, there are a few endonuclease cleavage sites of RsaI (GTAC, at 507, 822, 2610, 3095 and 3176), Sau3AI (GATC, at 575 and 2304), HaeIII (GGCC, at 3526), ClaI (ATCGAT, at 1067) and EcoRI (GAATTC, at both ends, 0 and 3564), and no site of MspI (CCGG) and SphI (GCATGC). A mutation at two bases is necessary to generate the recognition sequence of those restriction enzymes. To see the distribution of endonuclease sites in the DYZ1 repeated DNA family, a male DNA sample was analyzed by Southern hybridization using the pHY10 insert as

Table 2. Frequency of pentanucleotide species that appeared in the pHY10 insert.

Pentanucleotides		Frequency
TTCCA		229
with 1 nucleotide substitution		298
among them	TTCGA	(58)
	TTCCT	(33)
	GTCCA	(31)
	TTCTA	(27)
	TTCAA	(23)
	TTGCA	(22)
	TTTCA	(20)
with 2 nucleotide substitution		139
with 3 nucleotide substitution		36
with 4 nucleotide substitution		8
with 5 nucleotide substitution		1
with insertion		2
total		713

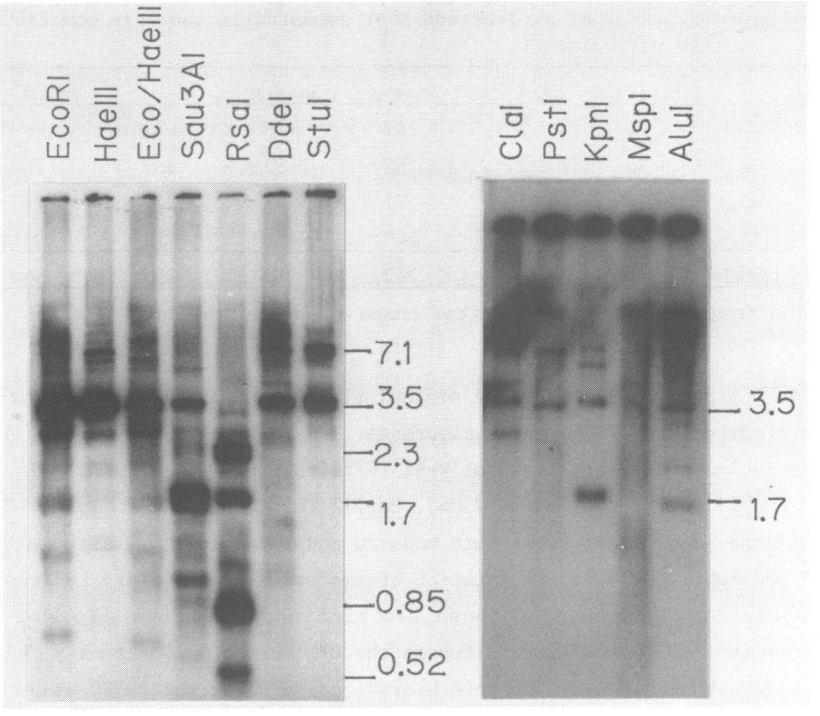

Fig. 4. Southern hybridization analysis of male DNA with the pHY10 insert after digestion with various restriction enzymes. The enzymes used are indicated at the top of each lane. Sizes in kb are indicated on the right.

a probe after digestion with various restriction enzymes (Fig. 4). Prominent bands detected in both EcoRI and HaeIII digests had an apparent size of 3.4 kb as described above, and a band detected in double digests by combination with EcoRI and HaeIII is a bit smaller than 3.4 Kb. This result is consistent with a previous result (16). There is one HaeIII site in the cloned fragment and two EcoRI sites at both ends. The two sites are separated by 38 bp. These results indicate that the unit length of the DYZ1 DNA family is 3.4 kb with one restriction site for both EcoRI and HaeIII in most of the repeated units. They also indicate that most units are ordered by direct joining in a head-to-tail manner. Thus the pHY10 insert represents a complete single unit of the DYZ1 DNA family. Several major bands smaller than 3.4 kb were detected in either TaqI or HinfI digest (data not shown). This is accounted for by the large number of cleavage sites in repeated units as found in the pHY10 insert. Computer searching

showed that most TaqI sites (also HinfI) in the insert were separated from each other by 15, 30, 45, 60, or 75 nucleotides. In the ClaI digest, most radioactivity was detected in a high molecular weight region. This indicates that there are a few ClaI sites in the DYZ1 repeated DNA family, and that the ClaI site which appeared in the pHY10 insert was an exceptional case. However, a small fraction of radioactivity was detected at the 3.4 kb region. This indicates that if there is a ClaI site in a repeated unit of the DYZ1 DNA family, ClaI sites frequently appear in the same position in adjacent repeated units. Minor bands at 3.4 kb were also detected in digests with several other restriction enzymes, some of which did not appear in the cloned pHY10 fragment (i.e. KpnI, PstI, MspI and AluI). These results indicate that once an endonuclease cleavage site is generated by a mutation, it is expanded at intervals of 3.4 kb in the DYZ1 DNA family.

DISCUSSION

A tandem array of pentanucleotides in DYZ1 repeated DNA family. We determined the nucleotide sequence of an EcoRI fragment which represents a complete unit of the DYZ1 repeated DNA family in the human Y chromosome. We isolated a series of plasmids carrying various deletions using Bal31, and then determined the nucleotide sequence in each plasmid from the deletion end point where a synthetic oligonucleotide was inserted. This sequencing strategy has been applied to a variety of genes in combination with the Maxam-Gilbert method or the Sanger method (17,18). This strategy is quite useful for highly repeated DNA sequences as described in this report, because an endonuclease cleavage site is always available for end labeling regardless of the sequence nature, and also because ordering of the partial sequences is not hampered by a similar sequence.

The EcoRI fragment thus determined was 3564 bp long and was composed of 713 iterations of pentanucleotides, among which TTCCA is dominant, followed by its single-base substituents. Southern hybridization analyses by us and others show that male DNA contain 800-5000 copies of the 3.4 kb fragment, and that most of the copies are arranged in a head-to-tail manner (16). Taken together, these results indicate that there are over hundreds of thousands of such pentanucleotides in the DYZ1 locus, most of which are presumably ordered tandemly. However, the results do not exclude the possibility that some other components interrupt the tandem array of pentanucleotides at several sites. Such high repetition of a relatively small unit have been found in telomeres and subtelomeres(19).

Similarities of the nucleotide sequence of the DYZ1 DNA family to that of satellite III. Four plasmids which carry a relatively small segment of satellite III have been cloned and sequenced. All four plasmids had features similar to a pentanucleotide array and the most frequently appearing pentanucleotide species, but there were considerable differences in the perfectness of the array. PPD9 (168-nucleotides) and PPD18 (166-nucleotides) consist of a perfect array of pentanucleotides, of which TTCCA is dominant. In contrast, a pentanucleotide array found in PPD17 was disturbed at 9 positions by the insertion of trinucleotides within 224-nucleotides (10). The longest segment that has been sequenced in HS3 is 132 nucleotides long, where a nanonucleotide ATTC$\begin{matrix}C\\G\end{matrix}$ATTC is reported to be a major component (9). However, we can assume that the HS3 segment also consists of a pentanucleotide array, where the dominant pentanucleotide species is TTCCA with several break points in the pentanucleotide array. These features are consistent with the high degree of homology between PPD9 or PPD18 and the pHY10 insert by direct comparison and the relatively low degree of homology with PPD17 and HS3. Homologies of PPD17 and HS3 with the pHY10 insert considerably increase when their short segments are compared or when the sequences are compared after phase adjustment. The sequences of pHY10, PPD9 and PPD18 are classified as a perfect pentanucleotide array, and the other two are classified as pseudo-pentanucleotide arrays.

Chang et al. reported that PPD9 and PPD18 hybridized to the 3.4 kb HaeIII fragment of male DNA but PPD17 did not (20). If this is the case, the perfectness of the pentanucleotide array is an important factor for the affinity of the two sequences in a hybridization condition, even if the two sequences basically consist of the same repeated unit, TTCCA. We do not know whether PPD9 and PPD18 are actually derived from the Y-chromosome. An enormous number of repeating units have TTCCA pentanucleotides as a major component in the Y-chromosome which makes their chromosomal assignment difficult.

Endonuclease cleavage sites in the pHY10 insert. The cloned pHY10 fragment had many TaqI and HinfI sites, most of which (45 sites) appeared in the overlapping region of their recognition sequences. The recognition sequences of TaqI (TCGA) and HinfI (GANTC) can be generated by a single-base substitution in the tandem array of the pentanucleotide TTCCA. Moreover, if the forth nucleotide C of one of the pentanucleotides is substituted with nucleotide G in the array, TaqI and HinfI sites are simultaneously generated

in the sequence TTCGATTCCA thus formed. In contrast, there are few, if any, cleavage sites in the pHY10 insert for many other restriction enzymes for which more than two base-substitution is necessary to generate its recognition sequence in an array of (TTCCA)s. These results indicate that TTCCA is not only the pentanucleotide most frequently appearing in the sequence, but also the original sequence from which the sequence has been diversified in an evolutionary process.

Unequal crossovers in the DYZ1 repeated DNA family. Hybridization analyses of male genomic DNA samples using the pHY10 insert as a probe revealed endonuclease cleavage sites in the DYZ1 DNA family. A 3.4 kb DNA fragment was detected as a major component and DNA fragments having multiples of 3.4 kb lengths were also detected as minor components in EcoRI and HaeIII digests. A 3.4 kb DNA fragment was also detected as a minor component in digests of several restriction enzymes, which rarely cleave the DYZ1 repeated DNA family. This means that there is a periodicity at an interval of 3.4 kb in the DYZ1 DNA family. It is generally accepted that the tandem repeated DNA family evolved and expanded by unequal crossovers (21). If we assume the DYZ1 repeated DNA family also evolved by unequal crossovers, then the EcoRI and HaeIII sites must have been generated at an early stage and expanded thoughout the DNA family because their sites are fixed in the repeated unit. The existence of minor components having multiples of the 3.4 kb length indicates that elimination of cleavage sites sometimes occurrs in a repeated unit or repeated units adjacent to each other. Rare endonuclease cleavage sites such as ClaI must be generated after most repeated units expand, and the unit carrying such a cleavage site expands, if at all, into only a limited portion of the DNA family. Unequal crossovers must have taken place at distances of 3.4 kb, because the unit length of a repeated DNA family is a consequence of site expansion by an unequal crossover.

It is generally assumed that unequal crossovers take place between homologous sequences of two DNA strands. Since the DYZ1 DNA family essentially consists of a pentanucleotide array, a small segment of the repeated DNA family has a homology to any other portion of the DNA family as far as the pentanucleotide phases match. Nevertheless a periodicity at an interval of 3.4 kb is maintained in the DYZ1 DNA family. Therefore, sequence homology alone is not a factor for determination of the distance of crossovers. A higher order of chromatin structure may exist for measuring distance during crossover. A similar hypothesis was proposed by Donehower

et al. in the α-satellite repeated DNA family (22,23).

Non-functional DNA sequences tend to behave selfishly (24,25) and sometimes form tandem repeated sequences. The DYZ1 DNA family seems to be a typical case of selfish DNA because no functional genes appear to be mapped. The large number of repeated sequences in the DYZ1 locus offer a good opportunity to study the mechanism of unequal crossover.

ACKNOWLEDGEMENT

We would like to thank Dr. M. Sugiura for his valuable discussions and Mrs. M. Kusuda for her technical assistance at the beginning of this work. This work was supported in part by a grant-in-aid from the Ministry of Education, Science and Culture, and from the Ministry of Welfare and Health, Japan.

REFERENCES

1. Human Gene Mapping 8 (1985): Eighth International Workshop on Human Gene Mapping.
2. Cooke,H.J. (1976) Nature 262, 182-186
3. Bostock,C.J., Gosden,J.R. and Mitchell,A.R. (1978) Nature 272, 324-328
4. Lau,Y.F., Huang,J.C., Dozy,A.M. and Kan,Y.W. (1984) Lancet 1, 14-16
5. Lau,Y.F. (1985) Cytogenet.Cell Genet. 39, 184-187
6. De La Chapelle,A., Page,D.C., Brown,L., Kaski,U., Parvinen,T. and Tippett,P.A. (1986) Amer.J.Hum.Genet. 38 330-340.
7. Corneo,G., Ginelli,E. and Polli,E. (1971) Biochim.Biophys.Acta, 247, 528-534
8. Gosden,J.R., Mitchell,A.R., Buckland,R.A., Clayton,R.P. and Evans,H.J. (1975) Exp.Cell.Res. 92, 148-158
9. Cooke,H.J. and Hindley,J. (1979) Nucl.Acids Res. 6, 3177-3197
10. Deininger,P.L., Jolly,D.J., Rubin,C.M., Friedmann,T. and Schmid,C.W. (1981) J.Mol.Biol. 151, 17-33
11. Molecular Cloning (1982) Cold Spring Harbor Laboratory
12. Yamada,M., Lewis,J.A., and Grodzicker,T. (1985) Proc.Natl.Acad.Sci. 82 3567-3571.
13. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol. 65, 499-560
14. Harper,M.E., Ullrich,A. and Saunders,G.F. (1981) Proc.Natl.Acad.Sci. U.S.A. 78, 4458-4460
15. Jeffreys,A.J., Wilson,V. and Thein,S.L. (1985) Nature 314, 67-73
16. Cooke,H.J. and McKay,R.D.G. (1978) Cell 13, 453-460
17. Kost,T.A., Theodorakis,N., Hughes,S.H. (1983) Nucl.Acids Res. 11 8287-8301.
18. Poncz,M., Solowiejczyk,D., Ballantine,M., Schwartz,E. and Surrey,S. (1982) Proc.Natl.Acad.Sci. U.S.A. 79 4298-4302
19. Blackburn,E.H. and Szostak,J.W. (1984) Ann.Rev.Biochem. 53, 163-194
20. Chang,H.C., Jolly,D.J., Biro,E., Jones,O.W. and Friedmann,T. (1983) Amer.J.Hum.Genet. 35 meeting abstract 170A
21. Smith,G.P. (1976) Science 191, 528-535
22. Donehower,L., Furlong,C., Gillespie,D. and Kurnit,D. (1980) Proc.Natl. Acad.Sci. U.S.A. 77, 2129-2133
23. Brutlag,D.L. (1980) Ann.Rev.Genet. 14, 121-144
24. Doolittle,W.F. and Sapienza,C. (1980) Nature 284, 601-603
25. Orgel,L.E. and Crick,F.H.C. (1980) Nature 284, 604-607