Newly identified repeat sequences, derived from human chromosome 21qter, are also localized in the subtelomeric region of particular chromosomes and 2q13, and are conserved in the chimpanzee genome

H.-S. Park[a], M. Nogami[b,1], K. Okumura[b], M. Hattori[a], Y. Sakaki[a,c], A. Fujiyama[a,d,*]

[a] *RIKEN Genomic Sciences Center, c/o Kitasato University, 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan*
[b] *Mie University, Faculty of Bioresources, 1515 Kamihama, Tsu, Mie 514-8507, Japan*
[c] *Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Tokyo 108-8639, Japan*
[d] *National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan*

Received 19 April 2000

Edited by Horst Feldmann

Abstract Subtelomeric regions have been a target of structural and functional studies of human chromosomes. Markers having a defined structure are especially useful to such studies. Here, we report 93 bp tandem repeat sequences found in the subtelomeric region of human chromosome 21q. They were also detected in the telomeric region of several other chromosomes. Interestingly, the repeat was also found in the 2q13 region which is known to be a position of chromosomal fusion, a major difference between the human and chimpanzee karyotypes. To the best of our knowledge, this repetitive sequence is a new member of human subtelomeric interspersed repeats.

Key words: Human chromosome; Telomere; Repeat sequence; Primate; Subtelomeric interspersed repeat

1. Introduction

The chromosomal region adjacent to the telomeric repeat array is commonly referred to as the subtelomeric region, and is known to contain a variety of repeat sequences called telomere-associated sequences, most of which are species-specific or shared only by closely related species [1–4]. In human cells, variable repeat sequences have also been found in the region, and these are collectively referred to as subtelomeric interspersed repeats (STIR) [5]. For example, families of 29 bp, 37 bp, and 61 bp repeats present in TelBam 3.4 and TelSau2.0 sequences that are distributed among several chromosomes [6]. It has been reported that the repeating structure at the ends of chromosomes might function in inter-chromosomal rearrangements during meiosis and mitosis, and might also act as a buffer zone to protect genes near the telomere from truncation caused by low telomerase activity [7,8].

In this study, we report a new STIR sequence composed of a tandemly repeated 93 bp sequence covering a region of about 1 kb, mapped to a location 7.5 kb from the telomeric array of the long arm of human chromosome 21. Fluorescence in situ hybridization (FISH) analysis revealed a specific distribution of the repeat units among chromosomes. To the best of our knowledge, this 93 bp repetitive sequence is a new human STIR. In addition, the repeat was also found in the genomic DNA of other species including chimpanzee.

2. Materials and methods

Construction of a human chromosome 21-specific library will be described elsewhere. Briefly, human chromosome 21 was isolated from a male lymphoblast cell line having a normal karyotype. Crude chromosomes were extracted from M-phase cells and were stained with Hoechst 33258 and chromomycin A3 prior to separation of individual chromosomes by means of a EPICS 750 cell sorter (Coulter, USA). Droplets containing chromosome 21 were collected and the DNA was isolated by the standard procedure. Fosmid vector pKS143 containing the replication origin and genes to control segregation of the F-plasmid and two *cos* sequences was constructed from a derivative of the SuperCos1 plasmid (Stratagene, USA) and a mini-F plasmid. Purified chromosomal DNA was fragmented by mechanical shearing and ligated to the vector before in vitro packaging using the GigaPackXL system (Stratagene, USA). Screening and isolation of the F50F5 telomeric clone and DNA sequencing will be described elsewhere (the accession number of the entire F50F5 is AP001478). For the subcloning of the 93 bp repeat unit, P1-11C1 isolated from a human chromosome-specific P1 library was also used (unpublished results obtained by H.-S. Park).

The mapping position of the 93 bp repeat was determined initially through PCR analysis using sequence-tagged site (STS) primers specific for the S100B gene (GDB: 185164) and the HRMT1L1 gene (NM_001535.1, GenBank). The PCR primer pair Irep93-FOR (5′-AGCCATGAATCAGGCATAAATA) and Irep93-REV (5′-GAAATCGTGTTTGACCAGCT) was used to confirm the number of repetitive sequence units.

Human metaphase spreads were prepared independently from peripheral blood donated by two non-related normal male volunteers, and analyzed by the standard R-banding FISH procedure [9]. Signals detected at the same position on the same chromatid were scored.

Chimpanzee lymphoblast cell lines were a kind gift from Dr. T. Ishida (The University of Tokyo).

3. Results

3.1. Detection and mapping of the 93 bp repeat sequences

To study the structural characteristics of the region adjacent to the telomere of human chromosome 21, we isolated fosmid clone F50F5 and the P1 clone P1-11C1 from the chromosome 21-specific libraries using telomeric repeat and STS

*Corresponding author. Fax: (81)-559-81 6716.
E-mail: afujiyam@lab.nig.ac.jp

[1] Present address: Department of Genetic Information, Division of Molecular Life Science, Tokai University School of Medicine, Isehara, Kanagawa 259-1195, Japan.

Abbreviations: FISH, fluorescence in situ hybridization; STIR, subtelomeric interspersed repeat; STS, sequence-tagged site

0014-5793/00/$20.00
PII: S0014-5793(00)01632-X

markers. From the results of DNA sequencing, we identified 93 bp repeating sequences in which highly homologous but not identical unit sequences were tandemly repeated 10 times (Fig. 1a). The exact position of this 93 bp repeats is 7361–8103 nucleotides from the start point of the telomeric $(TTAGGG)_n$ repeats (Fig. 1b).

3.2. Characterization of the 93 bp repetitive sequences

The nucleotide sequences of the 93 bp repeating units are highly conserved (93–99%). No identical unit is present in our sequence, and the positions of nucleotide substitution seem to be quite limited within the repeating units (Fig. 1). To test if the 93 bp is unique to chromosome 21, we surveyed the currently available nucleic acid databases using the BLAST program at first; however, sequences showing clear similarity were detected in human chromosomes 10, 19, and 22. Interestingly, we also detected several sequences in the human EST database that have similarity to the 93 bp units. However, the 93 bp units do not have a repeating open reading frame. No oligopeptide having a corresponding amino acid sequence was found in the protein databases. Thus it is unlikely that the repeat structure either contains part of a gene cluster or encodes a repeating oligopeptide motif.

Since human genome sequencing is still in its early stages, the above observations led us to experimentally survey homologous sequences in the human genome by the FISH technique using the entire sequence or the repeat structure (Fig. 1b) as a probe. As shown in Fig. 2, we detected strong positive signals at the chromosomal loci 5qter, 7pter, 17qter, 19pter, 19qter, 20pter, 21qter, and 22qter. In addition, several weaker signals

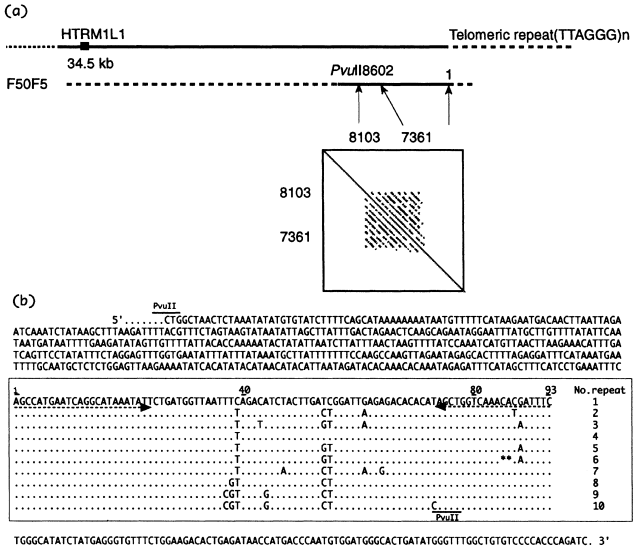

Fig. 1. Position and structure of the 93 bp repeat units. a: Position of the best known telomeric STS marker, HTRM1L1, which is located about 35 kb from the 21q telomere (H.-S. Park, unpublished observation). Fosmid clone F50F5 is the most telomeric clone isolated in this study. Positions of the *Pvu*II site, and the start and end points of the 93 bp repeats are indicated. Nucleotide positions are numbered from the start position of the telomeric repeats. Dot matrix analysis demonstrated the presence of direct repeats between positions 7361 and 8103. b: The 93 bp repeat unit is repeated 10 times between nucleotide positions 7361 and 8103. Nucleotides identical to the first unit are replaced by dots. A single nucleotide gap is represented by an asterisk. Arrowheads represent two sets of primer oligonucleotides, Srep93 and Irep93.

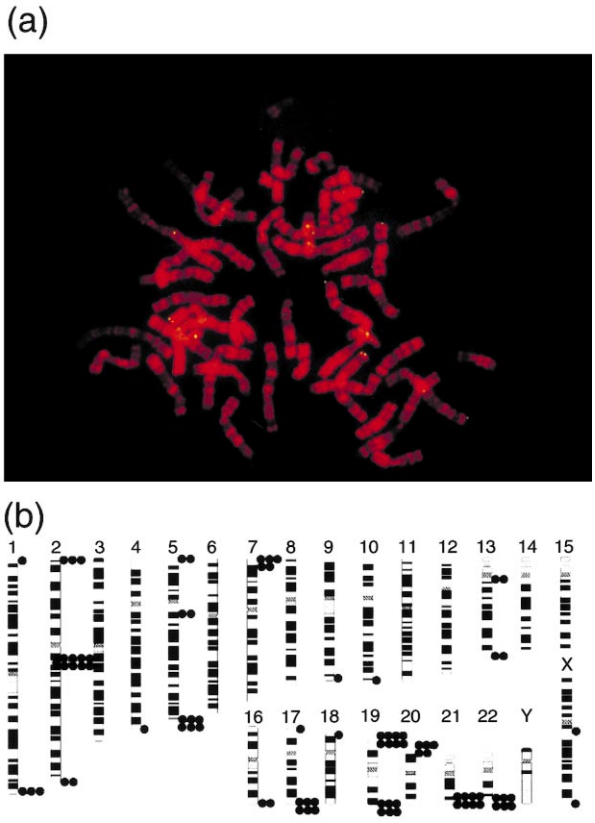

Fig. 2. Distribution of the 93 bp repetitive sequences in the human chromosomes. a: Example of R-banded FISH analysis. Positive signals are shown as bright dots. One chromosome 2 is seen at the upper left corner. (b) Scores of positive signals are displayed to the right of the ideogram of each chromosome. Only signals detected at the same site on both chromatids were counted.

could be detected at the telomeric region of other chromosomes (Fig. 2). The variable signal strength might reflect different similarities or number of repeats between the repeat structures. The 93 bp repeat was detected in the genomic DNA of at least 10 different individuals (data not shown), indicating that the repeat is widely conserved in the human genome.

An interesting finding was that we detected strong hybridization signals at the position of 2q13. Because this position is known to be a remnant of an ancient chromosomal fusion event possibly coupled to the divergence of humans from other primates [10,11], we further tested the existence of the 93 bp repeat using genomic DNA of other species. As shown in Fig. 3, we detected the presence of similar 93 bp repeats in DNA obtained from two different chimpanzees, but no significant number of repeats was detected in mouse DNA using the same PCR primers as with humans (Figs. 1 and 3). Thus, we concluded that the 93 bp repeat is conserved not only in human populations but also among primates and in other species probably in a less prominent form. Since our present resources are limited, detailed analysis is left for future study.

4. Discussion

The results of FISH analysis showed a definite distribution of the 93 bp repeats in the telomeric region of several chromosomes, although their evolutionary relationship is unclear

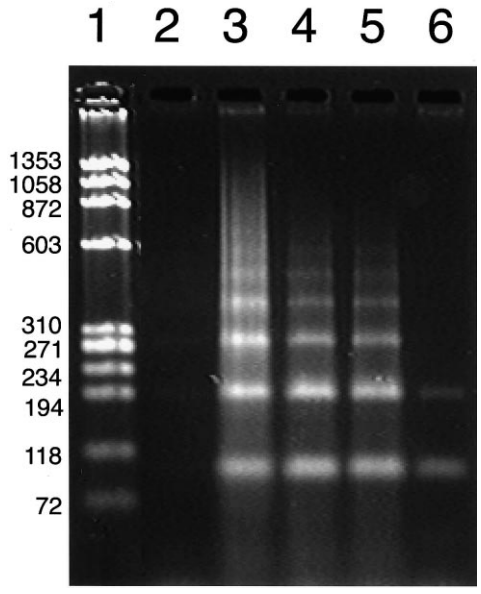

Fig. 3. Detection of the 93 bp repetitive sequences in human, chimpanzee, and mouse genomic DNAs by the PCR amplification method. Lane 3, human normal lymphoblast cell line; lanes 4 and 5, chimpanzee DNAs isolated from two different specimens; lane 6, mouse DNA. PCR reactions were carried out using Irep primers (see Section 2, and the legend to Fig. 1) and 10–20 ng of DNA (lane 2 indicates no DNA control) to obtain the 93 bp ladder. Lane 1, *Hae*III digested ϕX174 DNA (numbers on the left indicate the size of each fragment).

at this time (Fig. 2). This restricted distribution of the 93 bp repeats and sequence of the surrounding area may provide useful clues to study the evolution of human chromosomes; however, this remains a subject for future investigation.

Based on these observations, we regard this 93 bp repeat sequence as a new member of the STIR family. Unlike other STIR sequences found in the human genome [5,12,13], this is the first example of a STIR whose exact chromosomal position is known at the nucleotide level. In the reported chromosome 22 sequence, the 93 bp unit is repeated 14 times with 92–98% sequence similarity ([14]; NT_001128, GenBank). A similar inter-chromosomal distribution has been reported by Brown et al., using the TelBam11 fragment isolated from a YAC clone [6]. It has also been reported that the subtelomeric region of each human chromosome contains a unique combination of STIR sequences, with a unique distribution [6,15]. From this point of view, the STIR sequences have been used as genetic markers for studying the relationships between certain inherited disorders and the genes or sequences at the tips of chromosomes [16–18]. The most telomeric clone, F50F5, may be a useful DNA substrate to study functions of telomerase, telomere binding proteins, or the mechanism of chromosomal stability and other telomere functions.

In the FISH analysis, a strong hybridization signal on chromosome band 2q13 was obtained using the probe specific for the 93 bp repeat sequence (Figs. 1 and 2). This locus is thought to be the point of fusion of ancestral chromosomes found in the chimpanzee karyotype [10], which occurred a few million years ago [19,20]. The telomeric array or repetitive sequences in the subtelomeric region of the chromosomes might be involved in this event [7,8,21]. The results of PCR analysis (Fig. 3) suggest that similar 93 bp repetitive sequences are conserved in the genomes of various species with variable similarities and numbers of repeats.

Although our present knowledge of the molecular evolution of human chromosomes and genome is quite limited, we believe that this kind of repeat sequence will be a good marker in studies of the structure and evolution of the human genome and chromosomes. The entire human genome sequence and results of comparative genomic studies should shed further light on this issue in the near future.

Acknowledgements: This work was supported in part by a grant from the Ministry of Education, Science, and Sports of Japan. The authors are grateful to Dr. Inaho Danjo (Radiation Research Foundation, Hiroshima) and members of RIKEN Genomic Sciences Center for stimulating discussions. We also thank Mss. Mitsuko Endo, Megumi Masujima, and Ayuko Motoyama for superb technical assistance.

References

[1] Horowitz, H., Thorburn, P. and Haber, J.E. (1984) Mol. Cell. Biol. 11, 2509–2517.
[2] Foote, S.J. and Kemp, D.J. (1989) Trends Genet. 5, 337–342.
[3] Krejci, K. and Koch, J. (1999) Genomics 58, 202–206.
[4] Fajkus, J., Kovarik, A., Kralovics, R. and Bezdek, M. (1995) Mol. Gen. Genet. 247, 633–638.
[5] Rouyer, F., de la Chapelle, A., Andersson, M. and Weissenbach, J. (1990) EMBO J. 9, 505–514.
[6] Brown, W.R.A., MacKinnon, P.J., Villasanle, A., Spurr, N., Buckle, V.J. and Dobson, M.J. (1990) Cell 63, 119–132.
[7] Ashley, T. (1994) Hum. Genet. 94, 587–593.
[8] Wintle, R.F., Nygaard, T.G., Herbrick, J.A., Kvaloy, K. and Cox, D.W. (1997) Genomics 40, 409–414.
[9] Takahashi, E., Hori, T., O'Connell, P., Leppert, M. and White, R. (1990) Hum. Genet. 86, 14–16.
[10] Wienberg, J., Jauch, A., Ludecke, H.J., Senger, G., Horsthemke, B., Claussen, U., Cremer, T., Arnold, N. and Lengauer, C. (1990) Chromosome Res. 2, 405–410.
[11] Ijdo, J.W., Baldini, A., Wells, R.A., Ward, D.C. and Reeders, S.T. (1992) Genomics 12, 833–835.
[12] Jurka, J., Walichiewicz, J. and Milosavljevic, A. (1992) J. Mol. Evol. 35, 286–291.
[13] Petit, C., Levilliers, J. and Weissenbach, J. (1990) Proc. Natl. Acad. Sci. USA 87, 3680–3684.
[14] Dunham, I., Shimizu, N. and Roe, B.A. et al. (1999) Nature 402, 489–495.
[15] Cross, S., Lindsey, J., Fantes, J., McKay, S., McGill, N. and Cooke, H. (1990) Nucleic Acids Res. 22, 6649–6657.
[16] MacDonald, M.E., Anderson, M.A., Gilliam, T.C., Tranebjaerg, L., Carpenter, N.J., Magenis, E., Hyden, M.R., Healey, S.T., Bonner, T.I. and Gusella, J.F. (1987) Genomics 1, 29–34.
[17] Wasmuth, J.J., Hewitt, J., Smith, B., Allard, D., Haines, J.L., Skarecky, D., Partlow, R. and Hayden, M.R. (1988) Nature 332, 732–736.
[18] Vocero-Akbani, A., Helms, C., Wang, J.C., Sanjurjo, F.J., Korte-Sarfaty, J., Veile, R.A., Liu, L., Jauch, A., Burgess, A.K., Hing, A.V., Holt, M.S., Ramachandra, S., Whelan, A.J., Anker, R., Ahrent, L., Chen, M., Gavin, M.R., Iannantuoni, K., Morton, S.M., Pandit, S.D., Read, C.M., Steinbrueck, T., Warlick, C., Smoller, D.A. and Donis-Keller, H. (1996) Genomics 36, 492–506.
[19] Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. and Takahata, N. (1995) Proc. Natl. Acad. Sci. USA 92, 532–536.
[20] Goldstein, D.B., Ruiz Linares, A., Cavalli-Sforza, L.L. and Feldman, M.W. (1995) Proc. Natl. Acad. Sci. USA 92, 6723–6727.
[21] Weber, B., Allen, L., Magenis, R.E. and Hyden, M.R. (1991) Cytogenet. Cell Genet. 57, 179–183.