Volume 231, number 1, 130-134 FEB 05767 April 1988

A new human alphoid-like repetitive DNA sequence

Jeffrey L. Doering, Anne E. Burket and Susan L. Carnahan

Department of Biology, Loyola University of Chicago, Chicago, IL 60626, USA

Received 10 February 1988

A new tandemly repetitive sequence family, having the 170 bp basic repeat characteristic of alphoid sequences, has been identified in the human genome. Its organization in the whole genome and on chromosome 21 is different from that of any of the previously described alphoid families. Members of this new family are unusually heterogeneous in sequence, and there are a number of variant sequence classes. Some of the variant classes exist in separate genomic domains, and even on a single chromosome the members of such a class are not significantly intermixed with members of another class.

Alphoid DNA; Human genome; Chromosome 21; Sequence heterogeneity

1. INTRODUCTION

The human genome contains a number of repetitive DNA families [1]. One of the major ones is alphoid DNA, characterized by long tandem arrays of 170 bp units located primarily at centromeres [2,3]. The basic organization of this family was originally described as a repeating 340 bp unit composed of two 170 bp subunits that are 27% divergent in sequence [4]. Two such 340 bp units, differing in sequence by 1%, form a 680 bp unit [4]. Alphoid sequences with either the 340 bp or 680 bp organization compose nearly 2% of the human genome [5]. It has recently become clear that the alphoid repeats are organized into a number of other distinct families with their own particular patterns of genomic organization, some of which are apparently chromosome-specific [6–9]. We have found a new alphoid-like sequence family that exhibits chromosome-specific organization. The members of this family are unusually heterogeneous in sequence with variant sequence classes located in separate genomic domains.

Correspondence address: J.L. Doering, Department of Biology, Loyola University of Chicago, 6525 North Sheridan Road, Chicago, IL 60626, USA

2. MATERIAL AND METHODS

Placental DNA, purified as described [10], was a gift of D. Bieber. DNA from hamster-human hybrid cell line 153-E9A [11], containing human chromosome 21, was a gift from M. Cummings. Our previously described methods [12,13] were used for restriction digests, gel electrophoresis and transfer of DNA to filters. Filter hybridizations, using probes that were ^{32}P-labelled by nick-translation, were done as described [12], except that the most stringent washes were at 60°C in 2×SSC.

For cloning, placental DNA was digested to completion with *Hin*dIII and run on a 2.0% agarose gel. Fragments in the 400–600 bp size range were cut from the gel and eluted by standard techniques [14]. The isolated fragments were then ligated with *Hin*dIII-cut pBR322 and the recombinant molecules were used to transform *Escherichia coli* K12 (HB 101) [15]. Clones containing the repetitive sequence were detected by colony hybridization [16] using total genomic DNA as the probe. Plasmid DNAs were purified as described [14].

3. RESULTS

In searching the human genome for new classes of reiterated DNA, total genomic DNA was digested with various restriction enzymes, fractionated on agarose gels, blotted to filters and probed with total genomic DNA. This procedure, which only detects repetitive sequences, permitted us to identify a previously undescribed 550 bp fragment in *Hin*dIII digests (fig.1). Copies of this fragment were cloned as described in section 2.

Published by Elsevier Science Publishers B.V. (Biomedical Division)

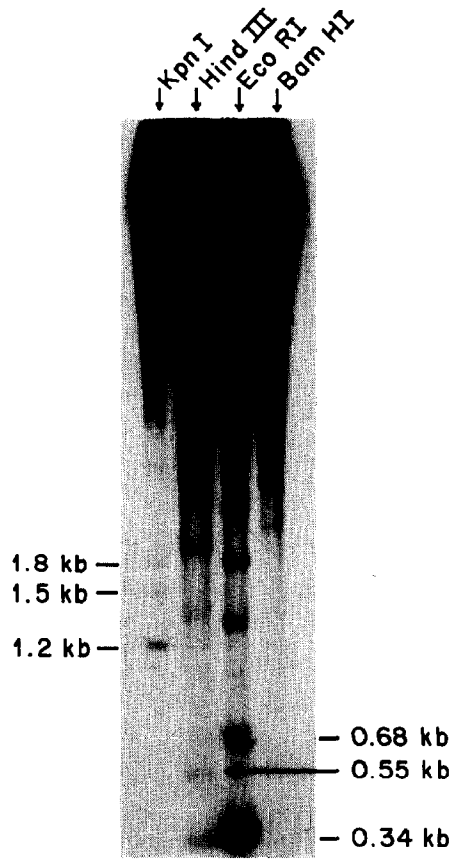

Fig.1. Identifying repetitive sequences in the human genome. Placental DNA was digested with the restriction enzymes indicated, blotted and probed with total genomic DNA (see section 2). The 0.55 kb *Hin*dIII fragments are denoted by the arrow.

Although of uniform size, some of the fragments are part of a dispersed repetitive sequence family (to be described elsewhere) while others are part of a tandemly reiterated family. One clone representing the tandemly organized family (pHH 550) was chosen for detailed study. The restriction map for this clone is shown in fig.2.

pHH 550 was hybridized to blots of total genomic DNA that had been digested to completion with restriction enzymes that do not cut within the cloned sequence (fig.3). These enzymes produce extensive 170 bp multimer series while also leaving significant amounts of high molecular mass DNA. This indicates that the basic repeating unit must have a number of sequence variants. The presence of multimer series also means that a

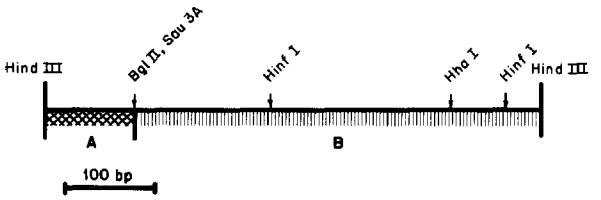

Fig.2. Restriction map of pHH 550. Locations of restriction sites for several enzymes are indicated along with the size scale. Fragments A and B were used as the probes for the blot shown in fig.5. Restriction enzymes that do not cut in this sequence are *Ava*I, *Bam*HI, *Cla*I, *Eco*RI, *Hae*III, *Hpa*II, *Pst*I, *Pvu*II, *Sal*I, *Taq*I, *Xba*I and *Xho*I.

variant restriction site is located at the analogous position in all the repeats in which it occurs.

The 170 bp repeating unit is characteristic of alphoid-like sequences [2,3]. The alphoid nature of pHH 550 can also be seen from the fact that it hybridizes weakly to cloned *Eco*RI 680 bp standard alphoid repeats at moderate stringency (not shown). The restriction map for pHH 550 (fig.2) is distinct from that of previously described alphoid sequences. Moreover, in sequences detected by pHH 550, the cleavage patterns for a variety of restriction enzymes (figs. 3 and 4) are clearly different from those made by the same enzymes in other variant alphoid families [6–9, 17–20]. Thus, pHH 550 appears to represent a new family of alphoid-like sequences.

There is an unusually high degree of sequence variation between members of this alphoid family. The clone pHH 550 represents only one type of variant, as can be seen from the fact that restriction enzymes not cutting in the cloned sequence (fig.3), as well as those which do (fig.4), produce extensive 170 bp multimer series in genomic blots. Thus the basic repeating unit of the pHH 550 family is more heterogeneous in sequence than that of many other alphoid families [7,8,17,19,20]. Sequence variation is further indicated by the presence of fragments in genomic blots which are not integral multiples of 170 bp (figs 3 and 4). The 550 bp *Hin*dIII fragment is in fact one such non-integral variant organization, and is relatively uncommon within the family (fig.3). Since probes from two regions of pHH 550 detect essentially the same sets of 170 bp multimers (fig.5), including fragments smaller than 550 bp, the *Hin*dIII 550 bp variant must itself be part of a tandem array of 170

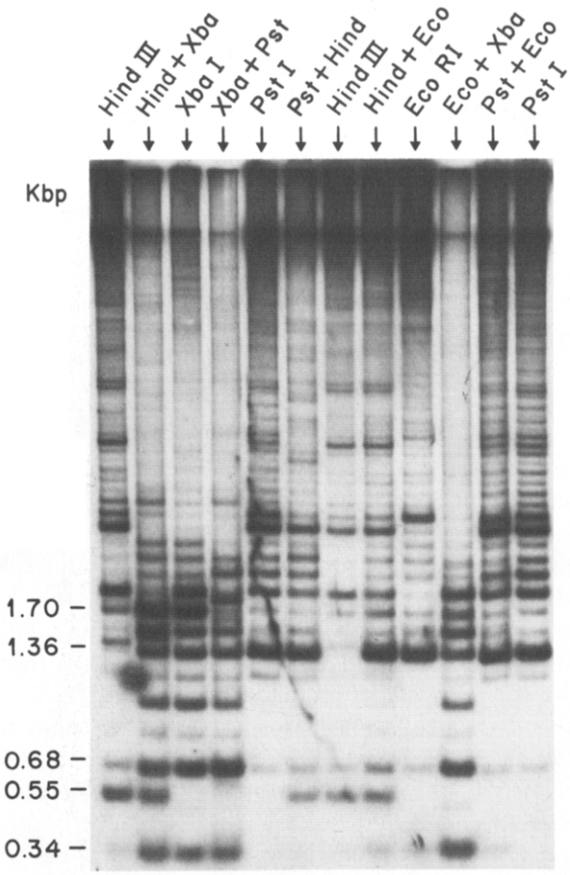

Fig.3. Genomic organization of pHH 550 sequences. Placental DNA was digested with the enzymes indicated, blotted and probed with pHH 550.

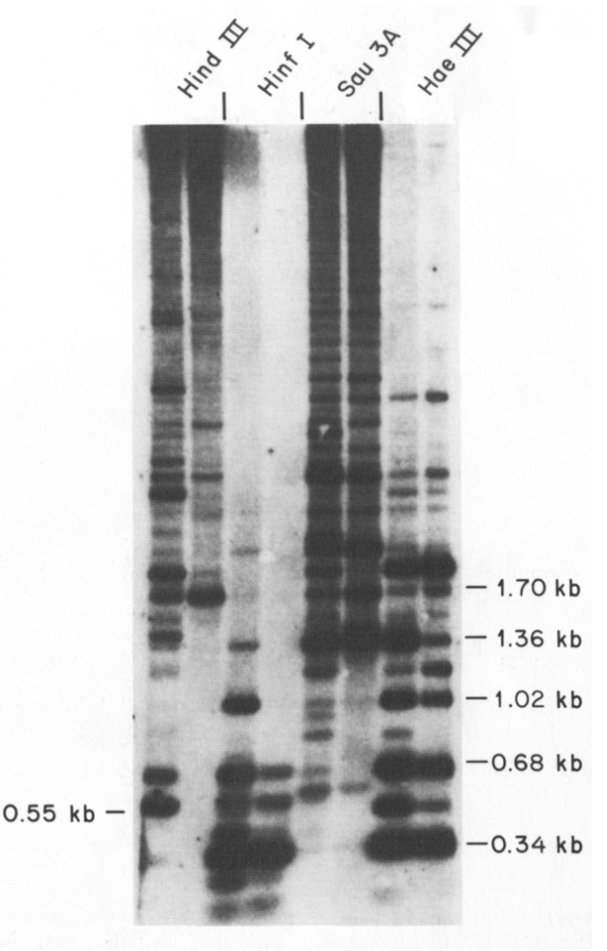

Fig.4. Presence of pHH 550 sequences on chromosome 21. 0.1 μg of placental DNA (left) or 2.0 μg of DNA from hybrid cell line 153-E9A (right) were digested with the enzymes indicated, blotted and probed with pHH 550.

bp repeats. Restriction sites in pHH 550 do not appear with a 170 bp spacing (fig.2), thus further emphasizing the heterogeneity in this family.

At least some of the variant sequences appear not to be randomly distributed throughout the family. If the variants recognized by one restriction enzyme are randomly intermixed in the tandem arrays with variants recognized by other restriction enzymes, then double digests with two such enzymes should result in the preferential disappearance of high molecular mass fragments. This is not the case for several of the double digests in fig.3. Thus, for example, repeats containing a variant *Eco*RI site are not intermingled with repeats containing a variant *Pst*I site. The banding pattern in a double digest is often the sum of the patterns seen in the two single digests. Each restriction enzyme thus appears to define a different class of variant sequences, and members of such classes tend to be located in distinct genomic regions apart from members of other classes.

The organization of the new alphoid family on a single chromosome was examined using DNA from a hamster–human hybrid cell line containing only human chromosome 21 [11]. While sequences homologous to pHH 550 are present on chromosome 21, there are a number of variants present in total genomic DNA which are not on chromosome 21, and also variants that are prominent on chromosome 21 which are rare in the total genome (fig.4). Of special note is the fact that the 550 bp *Hin*dIII fragment is one of the variants not found on chromosome 21. Comparable hybridiza-

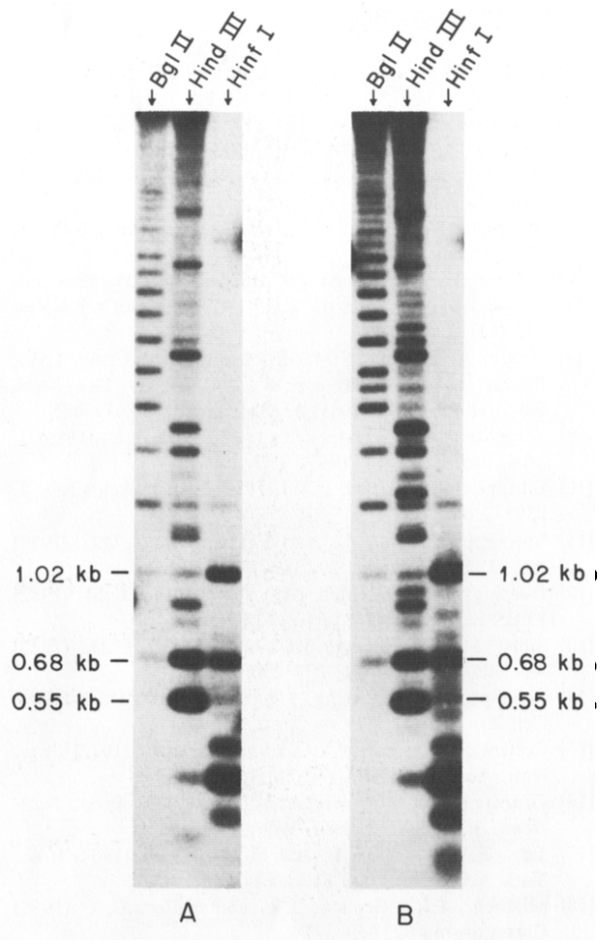

Fig.5. Presence of the full pHH 550 sequence on most variants. Placental DNA was digested with the enzymes indicated, blotted and probed with either the A or B fragment of pHH 550 (fig.2). Only a few of the variants do not contain both portions of the pHH 550 sequence.

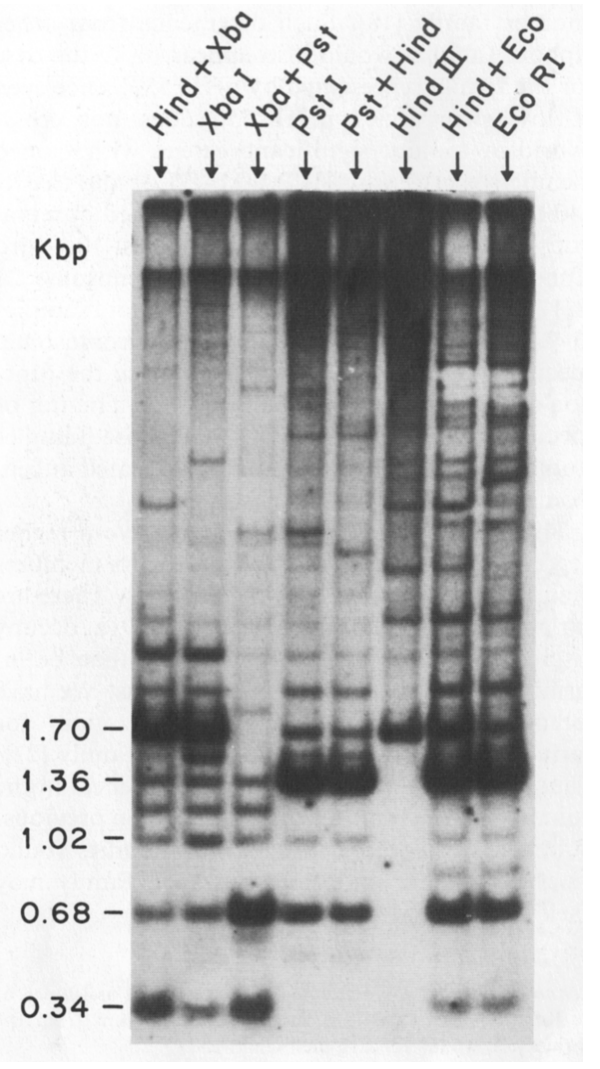

Fig.6. Organization of pHH 550 sequences on chromosome 21. DNA from hybrid cell line 153-E9A was digested with the enzymes indicated, blotted and probed with pHH 550.

tion intensities are seen in fig.4 for both total genomic DNA and 153-E9A DNA, even though 20-times more hybrid cell DNA was loaded on the gel. This result indicates that sequences hybridizing to pHH 550 are present on many other chromosomes in addition to 21, since the DNA content of chromosome 21 represents approximately the same fraction of genomic and 153-E9A DNA [21]. Double restriction digests of 153-E9A DNA (fig.6) reveal that on this single chromosome, as in the genome as a whole, the members of a variant class are often not significantly intermixed with members of another class.

4. DISCUSSION

We have identified a previously undescribed family of repetitive sequences in the human genome which appears to be another family of alphoid sequences. This is at least the eighth distinct alphoid family to be characterized [4,7,8,17–20], and together these families comprise nearly 3% of the human genome [5]. While all the alphoid families are composed of 170 bp repeating units, members of one family can be as much as 50% divergent in sequence from members of

another family [18]. Such divergence from other alphoid families would also appear to be the case for the family represented by pHH 550, since even at low stringencies pHH 550 does not cross-hybridize to any significant extent with cloned members of the *Bam*HI 2.0 kb [7] or the *Eco*RI 340 bp alphoid families (our unpublished observations). The pHH 550 family is at least the third alphoid family to be found on chromosome 21 [3,17], and, as is the case for other alphoid families [6–9,18,20], it exhibits some chromosome-specificity in its organization. Although the function of alphoid sequence families is still a matter of speculation, it is interesting that those families studied to date are all preferentially located at centromeres [3,17,19].

The pHH 550 family appears to have a higher degree of sequence variation between its members than that seen in other alphoid families. There are variant classes whose members tend to occupy discrete chromosomal domains rather than be intermixed with members of another class. We have recently detected a similar organization for variants of the *Eco*RI 680 bp alphoid family [22]. Thus, the organization of alphoid sequences in the human genome is more complicated than previously believed. Not only are multiple families found on the same chromosome, but each family may itself be organized into subdomains.

Acknowledgements: We appreciate the technical assistance of K. Kendra. This work was supported by a grant from the Shriners Hospitals for Crippled Children.

REFERENCES

[1] Britten, R.J. and Kohne, D.E. (1968) Science 161, 529–540.
[2] Manuelidis, L. and Wu, J.C. (1978) Nature 276, 92–94.
[3] Manuelidis, L. (1978) Chromosoma 66, 23–32.
[4] Wu, J.C. and Manuelidis, L. (1980) J. Mol. Biol. 142, 363–386.
[5] Smookler Reis, R.J. and Goldstein, S. (1980) Cell 21, 739–749.
[6] Willard, H.F. (1985) Am. J. Hum. Genet. 37, 524–532.
[7] Waye, J.S. and Willard, H.F. (1985) Nucleic Acids Res. 13, 2731–2743.
[8] Wolfe, J., Darling, S.M., Erickson, R.P., Craig, I.W., Buckle, V.J., Rigby, P.W.J., Willard, H.F. and Goodfellow, P.N. (1985) J. Mol. Biol. 182, 477–485.
[9] Jorgensen, A.L., Bostock, C.J. and Bak, A.L. (1986) J. Mol. Biol. 187, 185–196.
[10] Blin, N. and Stafford, D.W. (1976) Nucleic Acids Res. 3, 2303–2308.
[11] Moore, E.E., Jones, C., Kao, F-T. and Oates, D.C. (1977) Am. J. Hum. Genet. 29, 389–396.
[12] Doering, J.L., Jelachich, M.L.and Hanlon, K.M. (1982) FEBS Lett. 146, 47–51.
[13] Sims, M.A., Doering, J.L. and Hoyle, H.D. (1983) Nucleic Acids Res. 11, 277–290.
[14] Peterson, R.C., Doering, J.L. and Brown, D.D. (1980) Cell 20, 131–141.
[15] Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) Proc. Natl. Acad. Sci. USA 69, 2110–2114.
[16] Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961–3965.
[17] Jabs, E.W., Wolf, S.F. and Migeon, B.R. (1984) Proc. Natl. Acad. Sci. USA 81, 4884–4888.
[18] Mitchell, A.R., Gosden, J.R. and Miller, D.A. (1985) Chromosoma 92, 369–377.
[19] Gray, K.M., White, J.W., Costanzi, C., Gillespie, D., Schroeder, W.T., Calabretta, B. and Saunders, G.F. (1985) Nucleic Acids Res. 13, 521–535.
[20] Devilee, P., Slagboom, P., Cornelisse, C.J. and Pearson, P.L. (1986) Nucleic Acids Res. 14, 2059–2073.
[21] Graham, G.J., Hall, T.J. and Cummings, M.R. (1984) Am. J. Hum. Genet. 36, 25–35.
[22] Doering, J.L., Burket, A.E., Hanlon, K.M. and Schlegel, D.S. (1986) J. Cell Biol. 103, 491a.