
Intrinsic polymorphism of variable number tandem repeat loci in the human genome

Sher Ali and R. Bruce Wallace

Department of Molecular Biochemistry, Beckman Research Institute of the City of Hope,
1450 E. Duarte Rd, Duarte, CA 91010, USA

Received April 27, 1988; Revised and Accepted June 15, 1988

ABSTRACT

In the human genome, short tandem repetitive (STR) DNA sequences often show restriction fragment length polymorphisms (RFLPs) due to variation in the number of copies of the repeat unit. For a subset of these sequences known as minisatellites or variable number tandem repeat loci (VNTR), it has been proposed that a homologous "core" sequence of 10-12 nucleotides is involved in the mechanism(s) generating the polymorphism. In our present study we have prepared oligonucleotide probes complementary to one or two repeat units of several VNTR loci. Under stringent hybridization and wash conditions these probes hybridize locus specifically thus allowing the evaluation of the intrinsic polymorphism of individual loci. Our results indicate that not all of the loci having STR DNA sequences are polymorphic despite the fact that they share the "core" sequence. This suggests that more than the DNA sequence of the locus is involved in the mechanism(s) generating the polymorphism.

INTRODUCTION

The human genome contains several families of repetitive DNA sequences (1). Individual repetitive DNA sequences may be dispersed throughout the genome or may be organized in the form of long arrays of short tandem repeats (STR (1)). Some STR sequences have been referred to as minisatellites (2-5) although they have not been isolated as sequences of unique buoyant density. Many of these minisatellites seem to be highly polymorphic due to variation in the number of copies of their tandem repeat units (6). Recently, Nakamura *et al* (7) have described several loci which contain variable numbers of tandem repeats and termed these loci VNTRs. Some human VNTR loci have been isolated as cloned DNAs and characterized (3,8). Probes consisting of many tandem repeat units hybridize to multiple loci each having considerable restriction fragment length polymorphism (RFLP). This results in an individual specific hybridization pattern and forms the basis of genetic fingerprinting (2-5,8).

Two major questions surrounding the VNTR loci are: What is the

mechanism(s) by which these loci are generated? and the related question: What accounts for their high degree of polymorphism? Jeffreys' and coworkers (3) have noted that the various VNTR loci that have been characterized share a homologous "core" sequence of approximately 10 nucleotides (GGGCAGGAXG). Nakamura *et al* (7) have extended these observations to several other VNTR loci and have proposed a somewhat different "core" sequence (GGGXXGTGGGG). Jeffreys *et al* (3) have noted that the core element is homologous to the *Chi* sequence of *E. coli* and have suggested that this sequence promotes recombination and is thus responsible for the high degree of polymorphism observed. To investigate this question in more detail, we have constructed oligonucleotide hybridization probes complementary to one or two repeat units of several VNTR loci. Under appropriate conditions these probes hybridize locus specifically. Using these probes the intrinsic polymorphism of individual loci can be assessed.

MATERIALS AND METHODS

Synthetic DNA Probes, Synthesis and Labeling

The synthetic oligonucleotide probes listed in Table I were synthesized using an automated DNA synthesizer and purified by polyacrylamide gel electrophoresis and HPLC using a PRP-1 reverse phase column (Hamilton). Each oligonucleotide (10 pmole) was labeled at the 5' terminal hydroxyl group using γ -[³²P] ATP (20 pmole NEN, 7000Ci/mmole) and T4 polynucleotide kinase 5-7 units (NEN) in 10 μ l reaction volume containing 67 mM Tris-HCl, pH 8.0, 10 mM Mg(OAc)₂, and 10mM dithiothreitol. The kinase reaction was carried out at 37°C for 45min. The labeled oligonucleotide was then separated from the γ -[³²P] ATP by chromatography over DE 52 Cellulose (Whatman) as described elsewhere (9).

Isolation and Hybridization of Genomic DNA

DNA was isolated from blood cells of unrelated individual donors according to the methods described earlier (9). 3-5 μ g DNA was digested with Hinf I (BRL) according to the supplier's specification and electrophoresed in 0.7 or 1.5% agarose gel (Seakem, ME) in recirculated TAE buffer (40 mM Tris, 12 mM sodium acetate, 2 mM EDTA, pH 8.3). For size estimation, phage lambda DNA digested with the restriction enzyme Hind III was included in the gel. After the electrophoresis, the gels were dried on a gel dryer (Bio-Rad) for 1 hr at room temperature followed by another 1 hr at 60°C. Prior to the hybridization, the DNA within the dried gel was denatured by soaking the gel in 400 ml of 0.5 N NaOH, 0.15 M NaCl at room temperature for 25 min and then

the gel neutralized in an equal volume of 0.5 M Tris-HCl pH 8.0, 0.15 M NaCl for 20 min at room temperature (10). After a gentle rinse in 6X SSC, the gel was transferred to a seal-a-meal bag. The DNA within the gel matrix was then hybridized directly by incubating the processed gel for 16 to 20 hr at 37°C in 5xSSPE [1xSSPE = 180 mM NaCl, 10 mM sodium phosphate, 1 mM EDTA, pH 8.0] 0.1% SDS, 10 µg/ml sonicated and denatured *E. coli* DNA and 1 x 10⁶ cpm/ml labeled oligonucleotide probe. After the hybridization, gels were washed in 6xSSC at room temperature for 3-4 hr, and exposed to X-ray film

TABLE I
SEQUENCES AND HYBRIDIZATION CHARACTERISTICS OF THE OLIGONUCLEOTIDE PROBES

PROBE NAME	SEQUENCE 5'→3'	TOTAL LENGTH	LENGTH REPEAT UNIT	NO. REPEAT UNITS	CATEGORY ^a OF HYBRIDIZATION TO HUMAN DNA	REF.
O-33.6-22	(CCTCCAGCCCT) ₂	22	11	2	1	5
O-33.6-37	GCCCTTCCTCCGGAGCCCTCCTCCAGCCCTTCCTCCA	37	37 ^b	1	1	3
O-33.15-32	(CACCTCTCCACCTGCC) ₂	32	16	2	3	3
O-33.15-80	(CACCTCTCCACCTGCC) ₅	80	16	5	2	3
O-AY-29	GAGGARYAGAAAGGYGRYRVGTGGGCG ^c	29	37	1	1	8,13
O-YNH24	TCCTGAACAACCCACTGTACTTCCCA	27	31	1	1	7
O-33.1	GTGCCTGCTTCCCTCCCTCTCTGTGTC	27	62	1	1	3
O-34BHI	CCTGCTCCGCTCACGTGGCCACGCAC	27	?	1	1	7
O-CCR-26	(CCR) ₈ CC	26	3	8.67	2	7,17
O-CCA-26	(CCA) ₈ CC	26	3	8.67	2	7,17
O-H-Ras	CACTCCCCTTCTCTCCAGGGACGCCA	28	28	1	4	14
O-GACA-16	(GACA) ₄	16	4	4	2	9
O-GACA-24	(GACA) ₆	24	4	6	2	17

^a These oligonucleotides gave one of four different kinds of hybridization patterns under varying hybridization and wash criteria. (1) Probes produce highly polymorphic, multiloci, patterns under low stringency and a locus specific polymorphic pattern under high stringency. (2) Probes produce a highly polymorphic, multiloci, pattern under both low as well as high stringency, although a few minor bands are lost in subsequent washes. (3) Probes produce a polymorphic multiloci pattern under low stringency and allele specific but non-polymorphic pattern under high stringency. (4) This probe produces a locus specific and polymorphic pattern under low as well as high stringent condition.

^b The repeat unit for this sequence has been alternatively described as 11 (5) or 37 (3).

^c R = A and G, Y = C and T, V = not T, ? = repeat length not known.

(Kodak X-AR) with two intensifying screens (Dupont lightning plus) overnight. After obtaining the first autoradiogram, the gels were re-washed in 6xSSC at increasing temperatures including 37, 40, 45, 50, 55, 60, 65°C, and some of the gels even at 70°C for 5-10 min or longer depending upon the signal intensity and after each wash the autoradiograms were obtained. The ability of these wash conditions to dissociate a duplex containing even a single base pair mismatch has been previously documented (10-12). For the DNA hybridized in the 1.5% agarose gels, a final wash was done in 20 ml of 3M tetramethylammonium chloride, 50mM Tris-HCl, pH8.0, 2mM EDTA and 2.5% SDS at 62°C for 2 hr (13).

RESULTS AND DISCUSSION

For the present study, instead of using cloned DNA probes, synthetic DNA probes have been used. The rationale is based on the fact that tandemly repeated DNA sequences are expected to hybridize to many related sequences which are interspersed in the genome producing complex hybridization patterns (DNA fingerprints). On the other hand, a synthetic oligonucleotide probe complementary to a sequence present in at least one member of the repeat family will hybridize to those members of the family containing those sequences but not to those containing a related but not identical ones. Therefore, it is reasonable to expect that short synthetic DNA probes will hybridize to a subset of the repeat family, producing a less complex pattern than obtained with cloned probes and revealing any RFLPs present in that subset (intrinsic polymorphism). Further, if during the course of hybridization oligonucleotide-DNA duplexes containing mismatches do form, they are easily dissociated in the subsequent stringent wash. In this way, a very specific hybridization pattern can be obtained comprised of only stable duplexes. This approach has already been applied to the detection of the sickle cell (β^S) globin allele (10) as well as to the detection of a highly polymorphic locus useful for documenting bone marrow engraftment following bone marrow transplantation (14).

A summary of the hybridization results for probes to several VNTRs are presented in Table I. All the probes listed here, except the 28 base long H-Ras probe, hybridize to multiple loci under low hybridization stringency. Gels containing DNA samples from unrelated individuals were hybridized to each probe at 37°C. After each hybridization, the gels were washed in 6xSSC at various temperatures ranging from 37°-70°C and after each wash.

autoradiograms were recorded. Some of the representative hybridization results are presented in Fig. 1.

The oligonucleotide O-33.6-22, a 22 base long probe complementary to two consensus repeat units of the VNTR sequence 33.6 (5), produces a highly polymorphic pattern giving rise to a number of shared and unshared bands when the gel is exposed without any stringent wash (Fig 1A). There are four bands of 3.8, 2.8, 2.4 and 2.1 kb which appear to be common in all the individuals. These non-polymorphic fragments appear together with several polymorphic fragments ranging from 6-18 kb. After washing the gel at increasingly higher temperatures (see legend of Fig.1), a locus specific hybridization pattern emerges, where all the bands except those in the size range of 1.3 to 1.5 kb are eliminated (Fig 1B). These bands are obviously polymorphic, however since the repeat unit of this locus is only 11 nucleotides, the different alleles are not resolved well. Those bands which are removed by higher stringency washes probably reflect related sequences which are not completely complementary to the probe.

Similarly, the oligonucleotide O-33.1, a 27 base long probe complementary to a single consensus repeat unit of the VNTR sequence 33.1 (3), produces a complex polymorphic pattern under low stringency conditions (Fig 1C) comprised mostly of unshared bands in the size range of 9.5 to 1.9 kb except for one of 10.1 kb which is relatively faint and seems to be common amongst all six individuals. After the high stringency wash all but one or two bands are removed. Of the six individuals analyzed, four of them have two bands in the size range of 1.7 to 2.5 kb whereas two of them have only one band each of about 2.5 kb (Fig 1D). When DNA from families is analyzed these bands are inherited in a pattern consistent with a single Mendelian locus (not shown). Thus, under high stringency conditions this oligonucleotide probe revealed five alleles in the six individuals examined (2.5, 2.3, 2.1, 2.0. and 1.7 kb).

In contrast to the results obtained with the two probes discussed above, the oligonucleotide O-33.15, a 32 base long probe complementary to two consensus repeat units of the VNTR sequence 33.15 (3), produces a complex DNA fingerprint pattern when the gel is hybridized and washed at low stringency (Fig 1E) but a non-polymorphic locus specific pattern when washed at high stringency (Fig 1F). Under the high stringency wash condition, a 1.3 kb non-polymorphic band is seen in all individuals (Fig 1F). In addition to this 1.3 kb non-polymorphic band, there is an additional 1.45 kb band present in one individual (Fig 1F, lane 6). It is difficult to account for

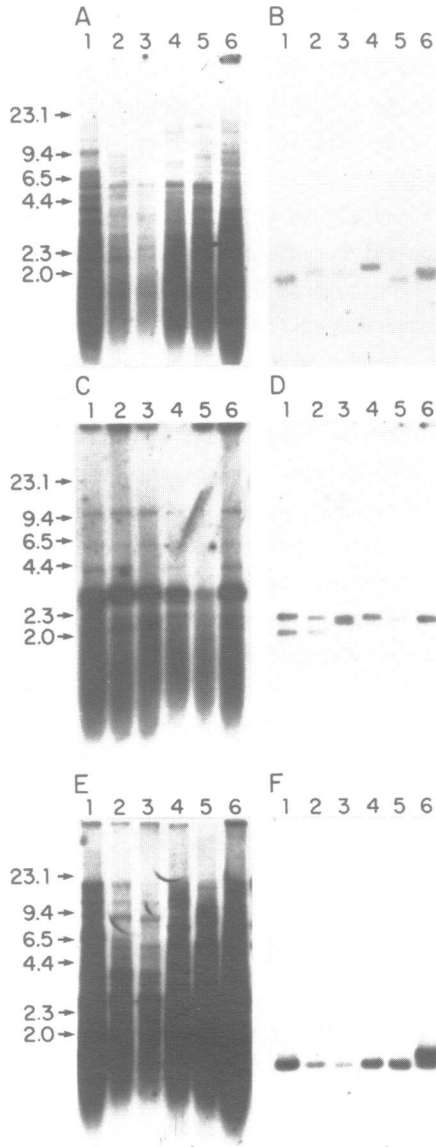


Figure 1. Hybridization patterns for probes complementary to the repeat unit(s) of three VNTR loci. The genomic DNA of six unrelated individuals was isolated, digested with *Hinf* I, subjected to electrophoresis on an agarose gel and denatured in situ. The agarose gel was dried and the dried gel hybridized with various probes. (A) Autoradiogram of the gel hybridized with oligonucleotide probe 0-33.6 (Table I) and exposed without any stringent wash. (B) Same gel after a wash at 65°C for 10 min. Note the

locus specific polymorphic bands. All other bands are lost during the sequential washes at the higher temperature (not shown). (C) Autoradiogram of the gel hybridized with 0-33.1 probe (Table I) and exposed without any stringent wash. Several polymorphic bands are seen ranging from 10.1 to 1.8kb. (D) The same gel after a final wash at 65°C for 10 min. Note the one or two bands in each lane. (E) Autoradiogram of the gel hybridized with the 0-33.15 probe (Table I) and exposed without any stringent wash. (F) The same gel after the wash at 65°C for 15 min. All the bands are eliminated except a 1.3 kb band present in all the individuals. In lane 6, there is an additional band of about 1.45 kb. This is the only sample that shows this additional band otherwise all the individuals are alike. The sizes of marker DNA fragments (bacteriophage lambda DNA digested with Hind III) are shown (in kb).

the presence of this additional band. Only the 1.3 kb band was found when the genomic DNA of several other unrelated individuals was analysed (not shown). Based on the level of hybridization of the probe to the two bands in this individual, it is reasonable to assume that this additional band is not an allele of the 1.3 kb band but rather is due to the presence of an additional locus complementary to the probe (no family members of this individual are available to confirm this suggestion). If this is the case, then it may be an example of a new VNTR locus. It is interesting to note that the locus from which the 33.15 sequence was derived is itself relatively non-polymorphic, while a DNA probe comprised of the 33.15 repeat

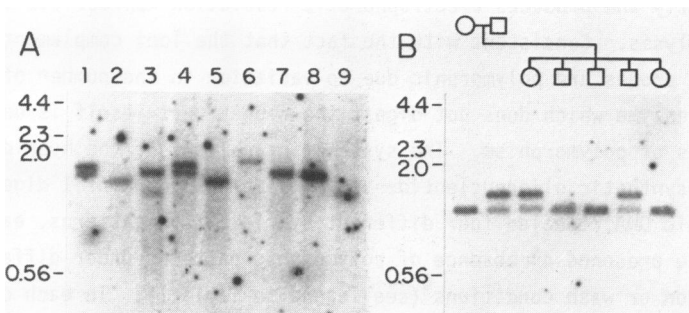


Figure 2. Intrinsic polymorphism of the VNTR locus 33.6. Genomic DNA was digested with *Hinf* I and electrophoresed in a 1.5% agarose gel in TBE buffer for better resolution than that obtained in Fig. 1B. The DNA in the gel was denatured *in situ* and hybridized with ^{32}P -labeled 0-33.6-37 at 63°C and washed in 6X SSC at RT for 2 hr followed by a wash at 60°C for 45 min. The gels were washed finally in 3M tetramethylammonium chloride at 62°C for 2 hr (13). A. Autoradiograph of a gel containing the DNA of nine unrelated individuals. B. Autoradiograph of a gel containing the DNA from a family with five siblings. The sizes of marker DNA fragments (bacteriophage lambda DNA digested with *Hind* III) are shown (in kb).

sequence produces the most informative fingerprint patterns of any of the cloned VNTR loci (14).

The cloned VNTR sequence 33.6 has been alternatively described as having an 11 nucleotide repeat unit (5) or as having a 37 nucleotide repeat unit made up of three 11 nucleotide repeats plus additional nucleotides (3). An oligonucleotide O-33.6-37, a 37 nucleotide long probe complementary to one consensus 37 nucleotide repeats, was synthesized. Genomic DNA from several unrelated individuals was digested with Hinf I and electrophoresed on a 1.5% agarose gel. The gel was hybridized with the 37 base long probe under stringent hybridization and post-hybridization wash conditions. As can be seen in Fig. 2A, the 33.6 locus is highly polymorphic, with three of the nine individuals examined being heterozygous. Expectedly, the shorter alleles give a lower hybridization signal due to the fact that they contain fewer copies of the repeat unit which binds less probe. At least 5 alleles are resolved in the 18 chromosomes studied. This probe thus reveals the intrinsic polymorphism of the 33.6 locus which was suggested by the results in Fig. 1B. Genomic DNA was also isolated from a family consisting of the mother father and five siblings. This DNA was analyzed in the same way (Fig. 2B). The results are consistent with a single Mendelian locus.

The autoradiograms presented in Fig. 1 and 2 are based on the DNA samples digested with the restriction enzyme Hinf I. This enzyme cleaves DNA frequently and produces electrophoretic resolution appropriate for the present analysis. Consistent with the fact that the loci complementary to the various probes are polymorphic due to variation in the number of repeat units, any enzyme which does not digest the repeat unit itself is useful for the analysis of polymorphism. The systematic analysis of the hybridization of several synthetic oligonucleotide probes (Table I) to Hinf I digested human genomic DNA revealed four different hybridization patterns, each being based on the presence or absence of polymorphic patterns under different hybridization or wash conditions (see legend to Table I). In each case where the probe consists of one or two repeat units, a locus specific hybridization pattern is obtained under high stringency conditions.

The fact that the locus encoding the 33.15 sequence appears to have limited or no polymorphism while closely related (cross-hybridizing) loci are highly polymorphic suggests that either there exist some constraints limiting the divergence of this locus or that this locus has arisen recently and that insufficient time has elapsed to generate polymorphism in the population. The latter possibility seems unlikely if the mutation frequency

for this locus were as high as that reported for other VNTR loci (10^{-3} /gamete (5)). It is not known what constraints could restrict the mutation of a VNTR locus. Just being present near a functional gene is not sufficient to inhibit changes in the copy number of the repeat unit, since polymorphism due to VNTR sequences exists in the H-Ras gene (15), the insulin gene (16) and the α -globin gene (17).

The 33.6 sequence has been alternatively described as a repeat of an 11 base trimer (5) or a 37 base repeat (3). A probe that is complementary to two of the 11 base long repeats (0-33.6-22) hybridizes to the same single locus as does a probe that is complementary to a single 37 base long repeat (0-33.6-37). Obviously, the former probe must be hybridizing with some degree of mismatch, since there does not exist a tandem repeat of the 11 base element in the 33.6 locus (3). The fact that under low stringency the 0-33.6-22 probe produces a fingerprint pattern suggests that the 11 base element is a common component of several different VNTR loci.

If the sequence of the VNTR repeat unit is important in the mechanism generating the polymorphism as proposed by Jeffreys *et al* (3-5), then it is possible that the sequence of the 33.15 repeat unit is not functional in stimulating recombination. One notable difference between the sequence of the 33.15 consensus and the core sequence of Jeffreys (5) is that the 33.15 sequence contains a T at the 14th base of the consensus while the core contains an A. An alternative explanation is that it is the nature of the flanking DNA of the 33.15 locus (sequence context) or chromatin conformation (eg, transcriptional activity or replication time) which may impart stability to this locus. While it is clear that we do not understand the mechanisms which generate VNTR polymorphism, the locus specific probes described in this paper represent an important new tool for investigating these mechanisms.

ACKNOWLEDGEMENTS

This work was supported by grant DCB-8515365 from the National Science Foundation (RBW). RBW is a member of the Cancer Center of the City of Hope (NIH CA33572).

REFERENCES

1. Martinez-Zapater, J.M., Estelle, M.A. and Sommerville, C.R. (1986) *Mol. Gen. Genet.* **204**, 417-423.
2. Jeffreys, A.J. and Morton, D.B. (1987) *Animal Genet.* **18**, 1-15.
3. Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) *Nature* **314**, 67-73.
4. Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) *Nature* **314**, 76-79.

5. Jeffreys, A.J., Wilson, V., Thein, S.L. Weatherall, D.J. and Ponder, B.A.J. (1986) *Amer. J. Hum. Genet.* **39**, 11-24.
6. Jeffreys, A.J., Royle, N.J., Wilson, V. and Wong, Z. (1988) *Nature* **332**, 278-281.
7. Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987) *Science* **235**, 1616-1622.
8. Wong, Z., Wilson, V., Jeffreys, A.J. and Thein, S.L. (1986) *Nucleic Acid Res.* **14**, 4605-4616.
9. Ali, S., Muller, C.R. and Epplen, J.T. (1986) *Hum. Genet.* **74**, 239-245.
10. Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L. and Wallace, R.B. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 278-282.
11. Ikuta, S., Takagi, K., Wallace, R.B. and Itakura, K. (1987) *Nucleic Acids Res* **15**, 797-811.
12. Miyada, C.G. and Wallace, R.B. *Methods Enzymol.* **154**, 94-107.
13. Wood, W.I., Gitschier, J., Lasky, L.A. and Lawn, R.M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1585-1588.
14. Yam, P., Petz, L.D., Ali, S., Stock, D. and Wallace, R.B. (1987) *Amer. J. Hum. Genet.* **41**, 867-881.
15. Capon, D.J., Chen, E.Y., Levinson, A.D., Seeburg, P.H. and Goedell, D.V. (1983) *Nature* **302**, 33-37.
16. Bell, G.I., Selby, M.J. and Rutter, W.J. (1984) *Nature* **295**, 31-35.
17. Jarman, A.P., Nicholls, R.D., Weatherall, D.J., Clegg, J.B. and Higgs, D.R. (1986) *EMBO J.* **5**, 1857-1863.
18. Ali, S. and Wallace, R.B. (unpublished results).