Int J Legal Med (1995) 108:31–35

ORIGINAL ARTICLE

M. D. Barber · R. C. Piercy · J. F. Andersen
B. H. Parkin

Structural variation of novel alleles at the Hum vWA and Hum FES/FPS short tandem repeat loci

Received: 1 February 1995 / Received in revised form: 5 April 1995

Abstract This paper reports the sequences of novel alleles identified during population databasing studies on the short tandem repeat loci HumvWA and HumFES/FPS. Two HumFES/FPS alleles follow the simple repeat pattern $(ATTT)_7$ and $(ATTT)_{15}$. Sequence variation corresponding to an A to C transversion occurred in the 5′ flanking region in two individuals possessing the designated allele 7. Two HumvWA alleles exhibited compound repeat regions comprising TCTA and TCTG repeat units. Sequence analysis confirmed the putative designation of 11 for a 127 base pair allele. However, a 131 base pair allele, putatively designated as 12, exhibited a more complex sequence. Two different types of repeat unit structures were identified which also exhibited sequence variation in the 3′ flanking region.

Key words Polymerase chain reaction · Short tandem repeat loci · HumFES/FPS · HumvWA · Sequence variation

Introduction

Tandemly repeated sequences of 2–5 base pairs in length have been shown to occur throughout the human genome with a frequency of one locus every 6–10 kilobases (Beckmann and Weber 1992). These regions have been termed short tandem repeat (STR) loci. Many are highly polymorphic, differing in the number of repeat units (Craig et al. 1988) and, in some cases, the base sequence between individuals (Adams et al. 1993; Urquhart et al. 1993; Möller and Brinkmann 1994). Their polymorphic nature and accessibility to amplification using the polymerase chain reaction (PCR) have led to their increasing use in mapping and genetic linkage studies of the human genome (Edwards et al. 1991).

M. D. Barber (✉) · R. C. Piercy · J. F. Andersen · B. H. Parkin
The Metropolitan Police Forensic Science Laboratory,
109 Lambeth Road, London SE1 7LP, UK

Recent studies have evaluated the use of multiplex PCR of STR loci combined with an automated fluorescent detection system for personal identification in the forensic sciences (Frégeau and Fourney 1993; Kimpton et al. 1993, 1994; Lygo et al. 1994). Such an approach has the advantage over current non-PCR based profiling techniques in that approximately 100-fold less DNA is required for analysis. The relatively small size of STR alleles allows highly degraded samples to be amplified (Hagelberg et al. 1990; Hochmeister et al. 1991; Jeffreys et al. 1992). The precision with which individual alleles can be resolved on polyacrylamide gel systems also obviates the requirement for continuous allele distribution models used in the analysis of variable number tandem repeat (VNTR) loci (Gill et al. 1990; Budowle et al. 1991; Evett and Gill 1991).

This laboratory recently introduced a multiplex PCR system (Kimpton et al. 1994) for the analysis of casework samples. The system is based on the co-amplification of four tetranucleotide repeat loci which are sized against a *Pst*1 digest of bacteriophage λ DNA using an ABI 373A DNA sequencer and GENESCAN 672 software. The four loci employed are HumvWA (Kimpton et al. 1992), HumFES/FPS (Polymeropoulos et al. 1991a), HumF13A1 (Polymeropoulos et al. 1991b) and HumTH01 (Edwards et al. 1991).

Sequencing studies have shown that the HumFES/FPS locus follows a simple repeat pattern, whereas the HumvWA, HumTH01 and HumF13A1 loci are compound repeats, each containing at least one non-consensus allele (Puers et al. 1993; Urquhart et al. 1994). Allelic frequencies for each locus have been calculated for Caucasian, Afro-Caribbean and Asian populations based on data collected from 1469 unrelated individuals (Greenhalgh, personal communication). During the course of this work we observed alleles for the HumvWA and HumFES/FPS loci that lay outside the allelic size range defined during precision studies of these loci. This paper reports the sequences of these alleles and data from both the repeat unit and the flanking DNA are presented.

Material and methods

Sample preparation

Bloodstains were prepared on pieces of clean cotton from donated blood samples and then allowed to dry at room temperature. The stains were stored at –70° C until ready for use. Genomic DNA was extracted from a bloodstain using the chelex procedure as described previously (Walsh et al. 1991). The recovered DNA was quantified by dot blot hybridisation to a higher primate specific probe (Walsh et al. 1992).

First round PCR amplification

A first round PCR amplification was performed with 10 ng of genomic DNA in a final concentration of 1 × GeneAmp PCR buffer (Perkin Elmer), 200 µM dNTPs (Sigma), 0.25 µM each primer (see Fig. 1) and 1.25 U AmpliTaq DNA polymerase (Perkin Elmer) in a total reaction volume of 50 µl. The reactions were performed in 0.5 ml GeneAmp thin walled reaction tubes (Perkin Elmer) and were overlaid with 50 µl of mineral oil (Sigma). Amplification was carried out in a Perkin Elmer Cetus 9600 thermal cycling block for 32 cycles of 94° C for 45 s, a 2 min cooling ramp to 54° C for 60 s, 72° C for 60 s followed by a final extension at 72° C for 10 min.

Second round PCR amplification

Gel purification of individual STR alleles was carried out by horizontal gel electrophoresis in 4.5% Metaphor agarose (Flowgen) containing 1 × TAE buffer (40 mM Tris-acetate, 1 mM EDTA pH 8.0, 0.5 µg/ml ethidium bromide). Electrophoresis was carried out at 10 V/cm for 5 h at 15° C in recirculated 0.5 × TAE running buffer. DNA bands, visualised under ultraviolet irradiation, were excised and then purified using a Mermaid gel purification kit (BIO 101) according to the manufacturer's instructions. An aliquot of the recovered PCR product containing approximately 1 ng of DNA was re-amplified over 12 cycles using the same conditions described above with the exception that primer concentration was reduced to 0.1 µM. In order to obtain sequence data from both DNA strands two separate reactions were carried out using primers carrying a 5′ biotin label on either the sense or anti-sense strands.

Solid phase sequencing reactions

An aliquot of each second round amplification containing 0.75–1.00 µg of DNA was added to 300 µg of Dynabeads M-280 Streptavidin (Dynal) in the presence of a final concentration of 2 M lithium chloride and incubated at 48° C for 15 min. Each of the following wash steps was carried out in a total volume of 500 µl with the aid of an MPC-E magnetic block (Dynal). The immobilised DNA was washed in TT buffer (250 mM Tris-HCl pH 8.0, 0.1% Tween 20) and then in double distilled water prior to being denatured in 50 µl 1.5 M NaOH for 4 min at room temperature. The eluted non-biotinylated strand was removed and the DNA/bead complex washed sequentially in TT buffer and TET buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1% Tween 20) before being resuspended in an appropriate volume of water.

Solid phase sequencing of both the sense and anti-sense strands was carried out using a Prism Sequenase dye terminator single stranded sequencing kit (Applied Biosystems) in a total reaction volume of 20 µl according to the manufacturer's instructions. The sequencing primers used were the non-biotinylated primers used in each second round PCR amplification reaction. Unincorporated labelled terminators were removed by washing the DNA/bead complex once in TT buffer, twice in 80% ethanol containing 1.35 M ammonium acetate and once in TET buffer. The samples were then resuspended in 4 µl of loading buffer (deionised formamide containing 5 mM EDTA pH 8.0), denatured at 37° C for 4 min and placed on ice.

The samples were loaded onto a 6% polyacrylamide gel containing 8.3 M urea and 1 × TBE buffer (90 mM Tris-HCl, 90 mM boric acid, 0.125 mM EDTA, pH 8.3) and analysed on a model 373A automated DNA sequencer (Applied Biosystems). Electrophoresis was carried out at 2500 V, 40 mA, 30 W, 40° C for 12 h. Sequence data were analysed using 373A version 1.2.1 analysis software (Applied Biosystems). Consensus sequences were created from the sense and anti-sense strands using SeqEd version 1.0.3 software (Applied Biosystems).

Results

DNA polymerase from *Thermus aquaticus* is known to catalyse a non-template mediated addition of a deoxyribonucleotide to the 3′ hydroxyl of PCR products. This has been termed the 'n + 1' product and has been shown to occur with relatively high efficiency (Clark 1988). Therefore, the multiplex PCR amplification conditions used in population data basing studies have been optimised to favour the 'n + 1' product in order to eliminate the formation of double peaks (Kimpton et al. 1993, 1994; Lygo et al. 1994). For consistency, allele sizes quoted throughout this paper are those observed in the initial databasing studies, and are therefore larger, by one base pair, than have been determined by sequencing methods.

Structural variation at the HumFES/FPS locus

The polymorphic unit of HumFES/FPS alleles has been described as a simple repeat following the pattern $(ATTT)_n$ (Urquhart et al. 1994). Studies using the primers in Fig. 1 have shown that the seven commonly observed HumFES/FPS alleles increase in size by 4 base pair increments over the range 211–235 base pairs (Lygo et al. 1994). The alleles have been designated as 8–14 reflecting the number of repeat units present (DNA Commission of the International Society for Forensic Haemogenetics recommendations 1992; ISFH recommendations 1994).

During population databasing studies HumFES/FPS alleles of 207 base pairs were observed in five individuals (two Caucasians, two Asians and one Afro-Caribbean) and of 239 base pairs in one individual (Afro-Caribbean). This suggests allele designations of 7 and 15 respectively. We have confirmed these putative designations for both alleles, with each type exhibiting a simple repeat pattern following the form $(ATTT)_n$.

HumvWA1 5′ CCCTAGTGGATGATAAGAATAATC 3′

HumvWA2 5′ GGACAGATGATAAATACATAGGATGGATGG 3′

HumFES/FPS1 5′ GGGATTTCCCTATGGATTGG 3′

HumFES/FPS2 5′ GCGAAAGAATGAGACTACAT 3′

Fig. 1 Primer sequences for first and second round amplifications. HumvWA sequences are from Kimpton et al. (1994) whilst HumFES/FPS sequences are from Polymeropoulos et al. (1991a)

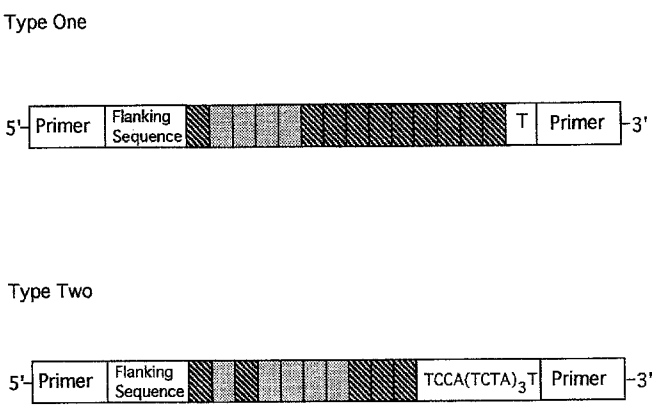

Fig. 2 Schematic representations of the type one and type two HumvWA allele 12. Only the TCTR strand is shown. TCTA repeat units are represented as ▨ whilst TCTG repeat units are represented as ▩

Comparison of the 5′ flanking sequences revealed an A to C transversion at base 34 (read 5′ to 3′) in two of the five individuals possessing the designated allele 7 (207 base pairs). Both individuals were of Asian appearance. The remaining 3′ and 5′ flanking sequence was identical to that deposited in the GenBank database (accession code X06292).

Structural variation at the HumvWA locus

The polymorphic unit of HumvWA alleles has been described as a compound repeat (Urquhart et al. 1994) following the pattern $(TCTR)_n$ (Kimpton et al. 1992), where R represents A or G. Nine commonly observed HumvWA alleles, which increase in size by 4 base pair increments over the range 135–167 base pairs, have been designated as 13–21 (Lygo et al. 1994; Urquhart et al. 1994).

In addition to these commonly occurring alleles, we have observed HumvWA alleles of 127 base pairs in five individuals (four Afro-Caribbeans and one Caucasian) and of 131 base pairs in three individuals (two Afro-Caribbeans and one Caucasian). This suggests allele designations of 11 and 12 respectively. We have confirmed the putative 11 designation of the 127 base pair allele, with the repeat region following the pattern $TCTA(TCTG)_3(TCTA)_7$ in all five cases. Both the 3′ and 5′ flanking sequences of this allele were identical to those deposited in the GenBank database (accession code M25858).

Sequence analysis of the three 131 base pair alleles has revealed two distinct types which differed in both the repeat and 3′ flanking sequences as shown in Fig. 2. Type one was observed in two Afro-Caribbean individuals and contained fourteen consecutive repeat units. A single thymine base was noted prior to the start of the 3′ primer sequence with no other 3′ flanking sequence being observed. The repeat region of type two, observed in a single Caucasian individual, comprised only ten consecutive repeat units. The repeat region differed from that observed for the majority of HumvWA alleles, which conform to the pattern $TCTA(TCTG)_{3-4}TCTA_m$ (Möller et al. 1994). The 3′ flanking region also exhibited an altered sequence structure $[TCCA(TCTA)_3T]$ to that deposited in the GenBank database (TCCA TCTA T). For both the type one and type two alleles, the 5′ flanking sequence was identical to that deposited in the GenBank database.

Discussion

Short tandem repeat alleles are named according to the number of consecutive repeat units each allele contains. This simplifies the presentation of STR evidence in court and allows direct comparison of STR typing results from different laboratories. Such a system also eliminates problems associated with a nomenclature based on amplified fragment size which would require very precise sizing of individual alleles. Therefore, we have sequenced novel alleles identified at the HumFES/FPS and HumvWA loci in order to establish the repeat unit sequence and assign allelic designations.

Sequence analysis of two novel alleles at the HumFES/FPS locus has confirmed the putative allele designations of 7 and 15 which were based on the size of the PCR amplified fragments. These observations are consistent with the results of other workers who have shown that alleles at this locus follow the simple repeat pattern $(ATTT)_n$ (Möller et al. 1994; Urquhart et al. 1994). Sequence variation was noted, however, in the 5′ flanking sequence in two individuals possessing the allele 7 where an A to C transversion was observed at base 34. This transversion has also been noted (Möller et al. 1994) in three Caucasian individuals possessing the allele 10 (219 base pairs) and five Caucasian individuals possessing the allele 11 (223 base pairs). Both of the individuals exhibiting the altered allele structure in the present study were of Asian appearance. Therefore, this mutation may be ancient, having become established before the divergence of the two ethnic groups.

Sequence analysis has confirmed the putative designation of 11 for the 127 base pair alleles observed in five individuals at the HumvWA locus. However, 131 base pair alleles, putatively designated as 12, exhibited more complex sequence structures. Two types of repeat unit and 3′ flanking sequence were identified in the three individuals possessing this allele. Type one contained fourteen consecutive repeat units and also exhibited an altered 3′ flanking sequence. A cytosine to thymine transition at position 86 could account for this, converting eight base pairs of the 3′ flanking sequence into an additional two repeat units as shown in the putative mutational scheme in Fig. 3.

The altered repeat and flanking sequences of the type two allele 12 may be explained in terms of two separate nucleotide transition events and a putative mutational scheme is shown in Fig. 4. One mutational event is a thymine to cytosine transition at position 78 in the penultimate repeat unit of the putative parent allele, reducing the number of consecutive repeats from twelve to ten. The

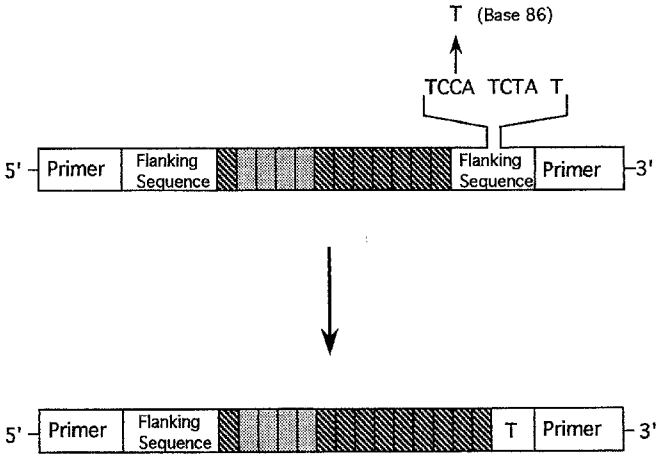

Fig. 3 Putative mutational mechanism for the generation of the HumvWA type one allele 12. Only the TCTR strand is shown. TCTA repeat units are represented as ▨ whilst TCTG repeat units are represented as ▩. Bases are numbered 5′ to 3′

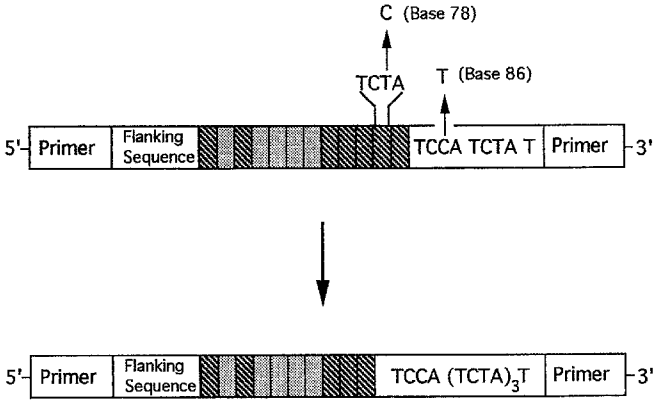

Fig. 4 Putative mutational mechanism for the generation of the HumvWA type two allele 12. Only the TCTR strand is shown. TCTA repeat units are represented as ▨ whilst TCTG repeat units are represented as ▩. Bases are numbered 5′ to 3′

other is a cytosine to thymine transition at position 86 converting eight base pairs of the 3′ flanking sequence into an additional two repeat units. The latter transition is similar to that shown in Fig. 3, generating the type one allele 12. However, the type one allele cannot be regarded as an intermediate in the formation of a type two allele as they exhibit different structures at the 5′ end of the repeat region.

Taken together these results suggest that a nomenclature based purely on the number of repeat units present in alleles at the HumvWA locus may not be appropriate. This is further supported by the findings of Möller and colleagues who have sequenced fifty HumvWA alleles from Caucasian individuals (Möller et al. 1994). They sequenced twelve HumvWA alleles designated as allele 14 and in each case they reported the repeat unit to be identical to that of the type two allele 12 characterised in the present study. The eight base pair difference in size between allele 12 and allele 14 is accounted for by a duplication of the first eight bases of the 3′ primer sequence [$(CCAT)_2$ to $(CCAT)_4$].

The work reported in this paper demonstrates that the designation of two alleles by the same allele number cannot be taken to mean that they have the same nucleotide sequence. This does not affect the reliability of a test based on molecular size, irrespective of the underlying sequences, nor the current statistical procedures in preparing evidence for court. It may be relevant to note, however, that some samples with the same allele designation could be shown to differ if further analytical procedures are used.

References

Adams M, Urquhart A, Kimpton C, Gill P (1993) The human D11S554 locus: four distinct families of repeat pattern alleles at one locus. Hum Mol Genet 2:1373–1376

Beckmann JS, Weber JL (1992) Survey of human and rat microsatellites. Genomics 12:627–631

Budowle B, Baechtel FS, Fourney RM, Adams DE, Presley LA, Deadman HA, Monson KL (1991) Fixed bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci for use in forensic comparisons. Am J Hum Genet 48:841–855

Clark JM (1988) Novel non-template nucleotide addition reactions catalysed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res 16:9677–9686

Craig J, Fowler S, Burgoyne LA, Scott AC, Harding HW (1988) Repetitive deoxyribonucleic acid (DNA) and human genome variation. A concise review relevant to forensic biology. J Forensic Sci 33:1111–1126

DNA Commission of the International Society for Forensic Haemogenetics (1992) Recommendations of the DNA Commission of the International Society for Forensic Sci Int 55:1–3

DNA recommendations – 1994 report concerning further recommendations of the DNA commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems. Int J Legal Med 107:159–160

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746–756

Evett IW, Gill P (1991) A discussion of the robustness of methods for assessing the evidential value of DNA single locus profiles in crime investigations. Electrophoresis 12:226–230

Frégeau CJ, Fourney RM (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. Biotechniques 15:100–119

Gill P, Sullivan KM, Werrett DJ (1990) The analysis of hypervariable DNA profiles: problems associated with the objective determination of the probability of a match. Hum Genet 85:75–79

Hagelberg E, Sykes B, Hedges R (1990) Ancient bone DNA amplified. Nature 342:485

Hochmeister MN, Budowle B, Borer UV, Eggman U, Comey CT, Dirnhofer R (1991) Typing of DNA extracted from compact bone tissue from human remains. J Forensic Sci 36:1649–1661

Jeffreys AJ, Allen MJ, Hagelberg E, Sonnberg A (1992) Identification of the skeletal remains of Josef Mengele by DNA analysis. Forensic Sci Int 56:65–76

Kimpton CP, Walten A, Gill P (1992) A further tetranucleotide polymorphism at the vWF gene. Hum Mol Genet 1:287

Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods Applic 3:13–22

Kimpton CP, Fisher D, Watson S, Adams M, Urquhart A, Lygo J, Gill P (1994) Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Int J Legal Med 106:302–311

Lygo JE, Johnson PE, Holdaway DJ, Woodroffe S, Whitaker JP, Clayton TM, Kimpton CP, Gill P (1994) The validation of short tandem repeat (STR) loci for use in forensic casework. Int J Legal Med 107:77–89

Möller A, Brinkmann B (1994) Locus ACTBP2 (SE33): sequencing data reveals considerable polymorphism. Int J Legal Med 106:262–267

Möller A, Meyer E, Brinkmann B (1994) Different types of structural variation in STR's: HumFES/FPS, HumVWA and HumD21S11. Int J Legal Med 106:319–323

Polymeropoulos MH, Rath DS, Xiao H, Merril CR (1991a) Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES). Nucleic Acids Res 19:4018

Polymeropoulos MH, Rath DS, Xiao H, Merril CR (1991b) Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene. Nucleic Acids Res:3753

Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW (1993) Identification of repeat sequence heterogeneity at the polymorphic repeat locus HUMTHO1[AATG]$_n$ and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 53:953–958

Urquhart A, Kimpton CP, Gill P (1993) Sequence variability of the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. Hum Genet 92:637–638

Urquhart A, Kimpton CP, Downes TT, Gill P (1994) Variation in short tandem repeat sequences- A survey of twelve microsatellite loci for use as forensic identification markers. Int J Legal Med 107:13–20

Walsh PS, Metzgar DA, Higuchi R (1991) Chelex® 100 as a medium for the simple extraction of DNA for PCR based typing from forensic material. Bio Techniques 1:91–98

Walsh PS, Varlaro J, Reynolds R (1992) A rapid chemiluminescent method for quantitation of human DNA. Nucleic Acids Res 20:5061–5065