Int J Legal Med (1998) 111 : 244–247

ORIGINAL ARTICLE

M. V. Lareu · S. Barral · A. Salas · C. Pestoni
A. Carracedo

Sequence variation of a hypervariable short tandem repeat at the D1S1656 locus

Received: 6 October 1997 / Received in revised form: 29 December 1997

Abstract A short tandem repeat at the D1S1656 locus was sequenced in 45 selected alleles and 13 different alleles were found which were designated according to the total number of repeats. This STR is a compound hypervariable STR consisting of blocks of (TAGR) repeats with a basic sequence structure $(TAGA)_4(TGA)_{0\text{-}1}(TAGA)_{6\text{-}16}$-$(TAGG)_{0\text{-}1}(TG)_5$. The presence of a TGA, probably due to an A deletion in the fifth TAGA repeat leads to intermediate a.3 alleles. Population data showed that this is a highly polymorphic STR with a heterozygosity of more than 0.89. This fact together with its simple structure and small size (129–168 bp) makes this STR one of the most interesting DNA polymorphisms for forensic and genetic purposes.

Key words D1S1656 · Short tandem repeats · Human genome · Sequence variation

Introduction

Analysis of short tandem repeat (STR) sequences by the polymerase chain reaction (PCR) (Litt and Luty 1989; Tautz 1989; Weber and May 1989; Edwards et al.1991) is nowadays the method of choice for forensic identification (Edwards et al. 1992; Kimpton et al. 1992; Urquhart et al. 1994; Pestoni et al. 1995) and genetic linkage analysis (Beckmann and Weber 1992; Weissenbach et al. 1992) Apart from the increased sensitivity inherent in any PCR technique, with STRs there is also the advantage of detecting discrete alleles, thus eliminating the need for the continuous allele distribution models currently employed with VNTR systems (Gill et al. 1990). Also, because of their small sizes STRs are more likely to be successful on old or badly degraded material (Hagelberg at al. 1991).

M. V. Lareu · S. Barral · A. Salas · C. Pestoni · A. Carracedo (✉)
Institute of Legal Medicine,
Faculty of Medicine, c/San Francisco s/n,
E-15705 Santiago de Compostela, Spain
e-mail: apimlang@usc.es
Tel. (34-81) 582327; Fax (34-81) 580336

Dinucleotide STRs are the most common STRs in the human genome and are the genetic markers most used for linkage analysis although they are not used in forensic science. The reason for this is that analysis of these STRs is affected by strand slippage during amplification, producing artifactual stutter bands (Hauge and Litt 1993). Nevertheless, trinucleotides, tetranucleotides and pentanucleotide repeats appear to be less prone to slippage and are more suitable for forensic purposes (Edwards et al. 1991). For all these reasons it is not surprising that many of these STRs (especially highly polymorphic tetranucleotide repeats) are included in many national and international trials in order to be validated for forensic use.

The choice of STRs is crucial. STRs range from the extremely complex to the most simple (Urquhart et al. 1994; Brinkmann 1996). Complex STRs have the advantage of hypervariability and simple STRs have the advantages of easy standarization and low mutation rates (Brinkmann 1996). Ideally, STRs for forensic purposes should combine both the characterististics of hypervariability and low mutation rate. In addition they should have other characteristics such as robustness, easy multiplexing, low stutter characteristics and small size since the latter is crucial for the successful typing of degraded samples (Alvarez-Garcia et al. 1996).

Using a GATA probe, a short tandem repeat was found in the locus D1S1656 (CHLC-GATA 44E05,P16122). Preliminary population data indicated the potential usefulness of this system for forensic purposes, but the nucleotide sequences of all the common alleles have not been determined. We report here the sequence of 13 different alleles, isolated from the Galician population (NW Spain). In addition a nomenclature for this system is proposed and population data from Galicia is reported.

Materials and methods

Genomic DNA was isolated from human blood as described (Valverde et al. 1993). Primers: forward primer 5′ GTG TTG CTC AAG GGT CAA CT, reverse primer *5′ GAG AAA TAG AAT CAC TAG GGA ACC (*Fluorescein labeled at the 5′ end).

Table 1 Sequence composition of the D1S1656 STR alleles

Allelic group	n° alleles sequenced	Sequence
10 (129 bp)	2	$(TAGA)_{10}(TG)_5$
11 (133 bp)	3	$(TAGA)_{11}(TG)_5$
12 (137 bp)	2	$(TAGA)_{12}(TG)_5$
	3	$(TAGA)_{11}(TAGG)_1(TG)_5$
13 (141 bp)	4	$(TAGA)_{12}(TAGG)_1(TG)_5$
14 (145 bp)	5	$(TAGA)_{13}(TAGG)_1(TG)_5$
15 (149 bp)	7	$(TAGA)_{14}(TAGG)_1(TG)_5$
15.3 (152 bp)	4	$(TAGA)_4(TGA)_1(TAGA)_{10}(TAGG)_1(TG)_5$
16 (153 bp)	3	$(TAGA)_{15}(TAGG)_1(TG)_5$
17 (157 bp)	3	$(TAGA)_{16}(TAGG)_1(TG)_5$
17.3 (160 bp)	3	$(TAGA)_4(TGA)_1(TAGA)_{12}(TAGG)_1(TG)_5$
18.3 (164 bp)	3	$(TAGA)_4(TGA)_1(TAGA)_{13}(TAGG)_1(TG)_5$
19.3 (168 bp)	3	$(TAGA)_4(TGA)_1(TAGA)_{14}(TAGG)_1(TG)_5$

The PCR reaction mixture was performed using 5 ng of genomic DNA in a 50 µl reaction volume with 10 mM TRIS-HCl (pH 8.3), 50 mM KCl, 0.01% gelatin, 1.5 mM $MgCl_2$, 200 µM each dNTP, 0.25 µM each primer and 1.25 U AmpliTaq DNA polymerase (Cetus, Emerville, Calif.). PCR conditions were 30 cycles denaturation 94° C for 45 s, annealing 60° C for 60 s, and extension 72° C for 60 s in a Perkin Elmer Thermocycler. The size of the PCR products were first determined in a 6% PAGE gel with 6 M urea in an automated sequencer (A.L.F. Pharmacia, Uppsala, Sweden) and then purified from a PAGE gel after silver staining. DNA sequences were obtained using a PCR Fentomol Sequencing Kit (Promega, Madison, Wis.). The cycle sequencing had the same cycle conditions as the first PCR. The resultant PCR products were denatured and applied to a 6% PAGE DNA sequencing gel with a separation distance of 24 cm.

Results and discussion

Nomenclature

Allelic designation was done according to the recommendations of the DNA Commission of the International Society for Forensic Haemogenetics (Bär et al. 1997). The allele designation is based on the number of repeats in the repeat unit and therefore it has been defined as (TAGR)n.

A total of 13 different allelic groups were found with the repeats ranging from $(TAGA)_{10}$ to $(TAGR)_{19}$ and 4 intermediate a.3 alleles due to the presence of a TGA after the first 4 TAGA repeats.

Sequencing variation

The STR in the D1S1656 locus was sequenced in 45 arbitrarily chosen alleles, including at least 2 individuals from each allelic group and 13 different alleles were found. The sequence composition of the D1S1656 alleles is displayed in Table 1 and Fig. 1. No variation was found in the constant regions.

D1S1656 is a compound hypervariable STR consisting of blocks of (TAGR) repeats with a basic sequence structure $(TAGA)_4(TGA)_{0\text{-}1}(TAGA)_{6\text{-}16}(TAGG)_{0\text{-}1}(TG)_5$. The presence of a TGA, probably due to an A deletion in the fifth TAGA repeat leads to intermediate a.3 alleles. The small

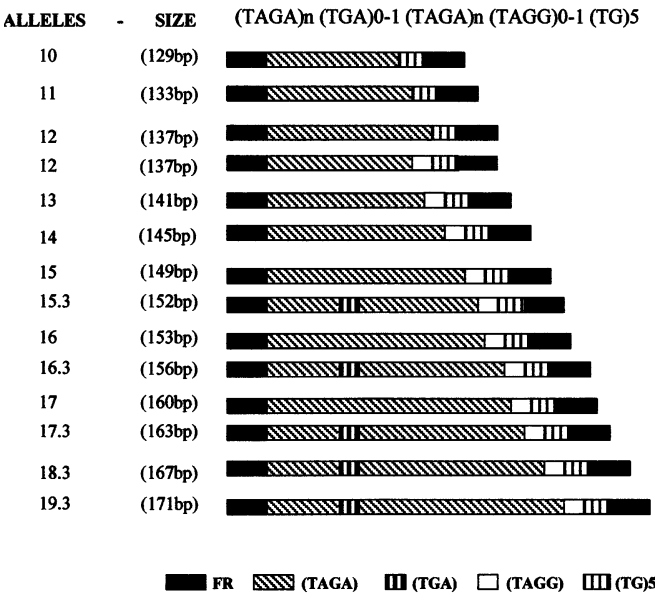

Fig. 1 Sequence structure of the different alleles found in our study

alleles 10 and 11 have variations in the (TAGA) unit only. Allele 12 shows a (TAGG) repeat in some individuals. The largest alleles always show a (TAGG) repeat at the end of the tandem array.

Population data

In a total of 125 healthy unrelated individuals from Galicia (NW Spain) 13 different allelic groups (TAGR) were found. Allele and genotype frequencies are given in Table 2. Frequencies range from 0.004 (alleles 10 and 19.3) to 0.164 (allele 15). The system was found to be in Hardy-Weinberg equilibrium and the exact test (Guo and Thompson 1992) gave a *P* value of 0.29. The discrimination power (Fisher 1951) was 0.98 and the heterozygosity (Nei and Roychoudhury 1974) 0.89.

Table 2 Allele frequencies of D1S1656 in the Galician population (n: 125)

Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
10	0.004	13	0.076	15.3	0.080	17	0.088	19.3	0.004
11	0.064	14	0.092	16	0.088	17.3	0.128		
12	0.160	15	0.164	16.3	0.016	18.3	0.036		

Hardy-Weinberg equilibrium: Exact test: p = 0.29
Heterozygosity: 0.89 CE: 0.77 PD: 0.98

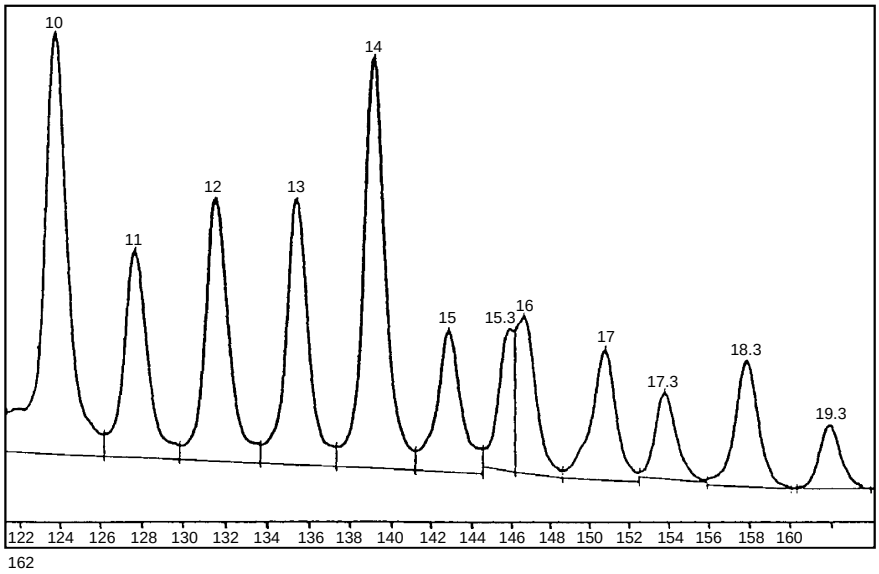

Fig. 2 Representation of D1S1656 allelic ladder composed of 12 sequenced alleles. The allele designation is based on the number of (AGAR) repeats

Additional information

An allelic ladder was constructed with the allele types 10, 11, 12, 13, 14, 15, 15.3, 16, 17, 17.3, 18.3, 19.3 (Fig. 2). This ladder is freely available from the authors. The accurate identification of the neighbouring alleles 15.3 and 16 can be problematic. However under the conditions described both alleles can be clearly distinguished. In a series of independent experiments $\delta 1 - \delta 2$ values (observed bp values for alleles 15.3 and 16) were calculated. The average was 0.754 and the standard deviation was 0.143.

The STR at the D1S1656 locus is a highly polymorphic STR with a relatively simple structure which is uncommon since highly polymorphic STRs are usually complex in structure.

The characteristics of this system, including easy amplification, high heterozygosity, easy multiplexing with other systems, small size and sequence simplicity make this STR one of the more interesting DNA polymorphisms for forensic purposes.

Acknowledgements This work was supported by a grant from the Xunta de Galicia (XUGA 20816B96). The authors would like to thank S.Rand for his constructive suggestions to this work.

References

Alvarez-Garcia A, Muñoz I, Pestoni C, Lareu MV, Rodriguez-Calvo M, Carracedo A (1996) Effect of environmental factors on PCR-DNA analysis from dental pulp. Int J Legal Med 109: 125–129

Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B (1997) DNA recommendations-Further report of the DNA Commission of ISFH regarding the use of short tandem repeat systems (1997) Int J Legal Med 110:175–176

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

Brinkmann B (1996) The STR approach. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics 6. Springer, Berlin Heidelberg New York, pp 24–32

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

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253

Fisher RA (1951) Standard calculations for evaluating a blood group system. Heredity 5:95–102

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

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

Hagelberg E, Gray IC, Jeffreys AC (1991) Identification of the skeletal remains of a murder victim. Nature 352:427–429

Hauge XY, Litt M (1993) A study of the origin of shadows bands seen when typing dinucleotide repeat polymorphisms by the PCR. Hum Mol Genet 2:411–415

Kimpton CP, Walton A, Gill P (1992) A further tetranucleotide repeat polymorphism in the vWF gene. Hum Mol Genet 1:28

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. Genetics 76:379–390

Pestoni C, Lareu MV, Rodriguez M, Muñoz I, Barros F, Carracedo A (1995) The use of the STRs HUMTH01, HUMVWA31/A, HUMF13A1, HUMFES/FPS, HUMLPL in forensic application. Int J Legal Med 107:283–290

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17: 6463-6471

Urquhart A, Kimpton CP, Downes TJ, Gill P (1994) Variation in short tandem repeat sequences. Int J Legal Med 107:13–20

Valverde E, Cabrero C, Cao R, Rodriguez-Calvo M, Díez A, Barros F, Alemany J, Carracedo A (1993) Population genetics of three VNTR polymorphisms in two different Spanish populations. Int J Legal Med 105:251–256

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Weissenbach J, Gyapay G, Dib A, Vignal J, Morisette P, Millasseau G, Vaysseix G, Lathrop M (1992) A second generation linkage map of the human genome. Nature 359:794–801