Int J Legal Med (1996) 109:134–138

ORIGINAL ARTICLE

M. V. Lareu · C. Pestoni · M. Schürenkamp · S. Rand
B. Brinkmann · A. Carracedo

A highly variable STR at the D12S391 locus

Received: 22 April 1996 / Received in revised form: 29 June 1996

Abstract A total of 103 fragments in the STR D12S391 locus were sequenced. 24 different alleles were found which can be grouped into 12 allelic classes based on the total number of repeats. The structure of this compund STR consists of blocks of (AGAT) and (AGAC) repeats with a consensus structure $(AGAT)_{8-17}$ $(AGAC)_{6-10}$ $(AGAT)_{0-1}$. Whereas shorter alleles only have (AGAT) repeats, > 225 bp alleles are more complex, having two motifs (AGAT) and (AGAC). Population data showed that this to be a highly polymorphic STR with a heterozygosity of 0.9. This fact together with its simple structure make this STR very suitable for forensic and genetic purposes.

Key words Short tandem repeats · D12S391 · Sequencing · Population data · Forensic usefulness

Introduction

Analysis of short tandem repeat (STR) sequences by the polymerase chain reaction (PCR) (Weber and May 1989; Tautz 1989; Litt and Luty 1989; Edwards et al.1991) is currently the method of choice for forensic identification (Edwards et al. 1992; Kimpton et al. 1992; Brinkmann 1992).

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 being used in forensic science. The reason is that analysis of these STRs has been affected by enzyme slippage during amplification, producing artifactual stutter bands (Hauge and Litt 1993). Nevertheless, tri, tetra and pentanucleotide repeats appear to be less prone to slippage and are more suitable for forensic purposes (Edwards et al.1991).

The choice of STRs is critical. STRs range from the extremely complex STRs to the most simple (Urquhart et al. 1994; Brinkmann 1996). Complex STRs have the advantages of hypervariability. Simple STRs are easier to standardize and have low mutation rates (Brinkmann 1996).

Using a GATA probe, a short tandem repeat was recently found in the locus D12S391 (CHLC-GATA 11 H08). A preliminary population study and sequencing data indicate the potential usefulness of this system for forensic purposes (Lareu et al. 1996).

The aim of this study was to determine the variation of sequence structure and to evaluate the potential forensic utility of this STR system. The latter includes a population genetic study in two caucasian populations and the construction of appropriate sequenced allelic ladders.

Material and methods

DNA was extracted from blood samples of 166 healthy unrelated individuals from Galicia (NW Spain) and 188 from the Münster area (Germany).

The DNA was quantified using a Perkin-Elmer 552 UV/Vis Spectophotometer or, alternatively the slot-blot technique with the human specific probe D17Z1 (Gibco BRL).

Primers sequence

Forward primer: 5'AACAGGATCAATGGATGCAT3'

Reverse primer: *5'TGGCTTTTAGACCTGGACTG3'
(* 5'fluorescent labeling)

Amplification conditions

PCR amplification of the STR in D12S391 was performed using 1–5 ng of genomic DNA in a 25 µl reaction volume. Reaction included 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, CA, USA). 30 cycles of PCR (94°C–45 s, 60°C-1 min, 72°C-1 min) were performed.

M. V. Lareu · C. Pestoni · A. Carracedo (✉)
Institute of Legal Medicine, Faculty of Medicine,
E-15705, Santiago de Compostela, Spain
Fax: +34 (81) 580 336

M. Schürenkamp · S. Rand · B. Brinkmann
Institute of Legal Medicine, Westfälische Wilhelms-Universität,
Von-Esmarch-Strasse 86, D-48149 Münster, Germany

Taq-Cycle sequencing

Isolation of DNA fragments from heterozygous individuals was carried out after polyacrylamide gel electorphoresis and silver staining as described (Möller and Brinkmann 1994). Gel pieces containing DNA bands were cut from the gel and transferred to microfuge tubes. Elution was carried out using the „crush and soak" method (Maniatis et al 1989).

A total of 103 fragments were sequenced using two different automated DNA sequencers (A.L.F., Pharmacia, Uppsala, Sweden and ABI 373A, Applied Biosystems, Foster City, Calif.).

A.L.F. Protocol

Eluted DNA fragments were reamplified and purified with MicroSpin S-200 HR Columns (Pharmacia) before the Cycle Sequence. The sequencing reaction was carried out using a ftmol Sequencing Kit (Promega, Madison USA). 100–400 ng of template DNA and 0.5 µM sequencing primer (D12S391 *5′TGGCTTTTAGACCTGGACTG 3′). 30 cycles of PCR (94° C–45 s, 60° C-1 min, 72° C–1 min) were performed using a MJR Thermocycler.

When the cycle sequencing was finished, 4 µl of loading buffer (5 mg/ml dextran blue/formamide) was added to each sample. The samples were heat denaturated at 94° C for 2 min before being loaded onto a 6% polyacrylamide sequencing gel on an Automated DNA Sequencer (A.L.F., Pharmacia) at 1600 V, 38 mA and 45 W for 5 h. Sequenced data was analysed automatically on the A.L.F. sequencer using the A.L.F. manager software (Pharmacia, Uppsala, Sweden).

ABD protocol

Eluted DNA fragments were reamplified, PCR products were desalted and concentrated using Centricon 100 (Amicon, Beverly. USA). Sequencing reactions were carried out using the Taq Dye-Deoxy-Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA), 2 µl of the centricon eluate (200–500 ng template DNA) and 0.5 µM sequencing primer (both primers were used as sequence primers to confirm the sequence data). 25 cycles of PCR (96° C–15 s, 60° C–15 s, 60° C–4 min) were performed using a Perkin Elmer 9600 Thermocycler. DNA was purified by phenol/chloroform extraction and ethanol/NaAc precipitation. The DNA pellet was dried and 4 µl loading buffer was added (formamide/50 mM EDTA (5:1))

After heat denaturation at 94° C for 2 min, the samples were separated in a 6% PAA sequencing gel on the ABD 373A Sequencer at 2500V, 40 mA and 30 W. Sequence data was analysed automatically on the ABI 373A Sequencer using Data collection and SeqEd software (Applied Biosystems, Foster City, Calif.).

Population study

The samples from Galicia were typed using the A.L.F sequencer and the samples from Münster using an A.B.I sequencer. Common sequenced allelic ladders were used and some samples were typed with both systems for validation purposes.

A.L.F. protocol

5 µl of loading buffer (5 mg dextran blue in 1 ml of 95% formamide) were added to 1 µl of each PCR product and combined with internal lane standars (Parmacia, Uppsala Sweden). The samples were heat denatured at 94° C for 2 min before being loaded onto a standard 6% polyacrylamide denaturing sequencing gel (7.0 M urea, 5.7% acrylamide, 0.3% bisacrylamide, 100 mM tris-borate pH 8.3 and 1 mM Na_2EDTA). The gels were run for 4 h at constant power (45 W) 1600V and 42 mA on the Automatic Laser Fluorescent DNA sequencer (A.L.F., Pharmacia, Sweden). Fragment sizes were determined automatically using the „Fragment Manager" software, and typed by comparison with a sequenced allelic ladder.

A.B.D. protocol

The alleles were separated on a 6% denaturing gel (8.3 M urea, 6% acrylamide/bis-solution and 1 × tris/borate buffer.) using the fluorescence detection system (Applied Biosystems, Foster City, Calif.). The gel were run at 800 V, 45 mA and 30 W for 7 h. The fragment sizes were automatically analyzed using the internal standard Genescan 500 labelled with ROX (6-carboxyrhodamin X) and the Genescan software 672 (Applied Biosystems, Foster City, Calif.).

Allelic designation and allelic ladder

Allelic designation was made according to the recommendations of the DNA Commision of the International Society for Forensic Haemogenetics (ISFH) (1992, 1994). The allele designation is based on the number of repeats in the repeating unit and it has been therefore defined for $(AGAR)_n$.

A total of 11 sequenced alleles were selected to be used as allelic ladder under denaturing conditions (Fig. 1). The allelic ladder can be easily reamplified using the same protocol and it is available from the authors.

Mobility studies

Polyacrylamide gels and different T and C values and non-denaturing conditions were used as described in Pestoni et al. (1995).

Statistical analysis

Hardy-Weinberg equilibrium was tested for with conventional Pearson's chi-square methods (χ^2), and with the two exact tests

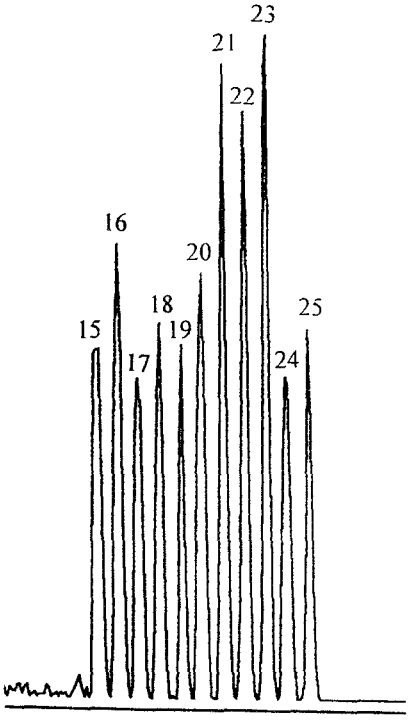

Fig. 1 Representation of D12S391 allelic ladder composed of eleven sequenced alleles. The allele designation is based on the number of (AGAR) repeats

proposed by Guo and Thompson (1992) based on conventional Monte Carlo methods and the Markov chain approach, respectively. For the Monte Carlo method the number of batches was 100 and the size of each batch 170. For the Markov chain method the number of batches was 20 and the size of each batch 5000. For the Markov chain method the dememorization period was 1000 steps.

Comparison of population data was carried out using a 2-way RxC contingency table test comparing allele distributions for population sample homogeneity and using chi-square as statistical parameters.

Other statistical parameters of genetic and medico-legal interest that were used: power of discrimination (PD) was calculated following Fisher's method (Fisher 1951), heterozygosity value (h) was calculated as described by Nei and Roychoudhury (1974) and chance of exclusion (CE) was calculated as described by Ohno et al. (1982).

Results

Sequence structure and variation of D12S391 alleles

A total of 103 different fragments were sequenced and analysed and 24 different alleles were found ranging from 209 bp to 253 bp. Alleles were selected using native gels. The sequence composition of the D12S391 alleles is displayed in Figure 2 and Table 1. D12S391 is a compound STR consisting of (AGAT) and (AGAC) repeats varying in number between alleles. This system has a basic sequence structure of $(AGAT)_{8-17}$ $(AGAC)_{6-10}$ $(AGAT)_{0-1}$.

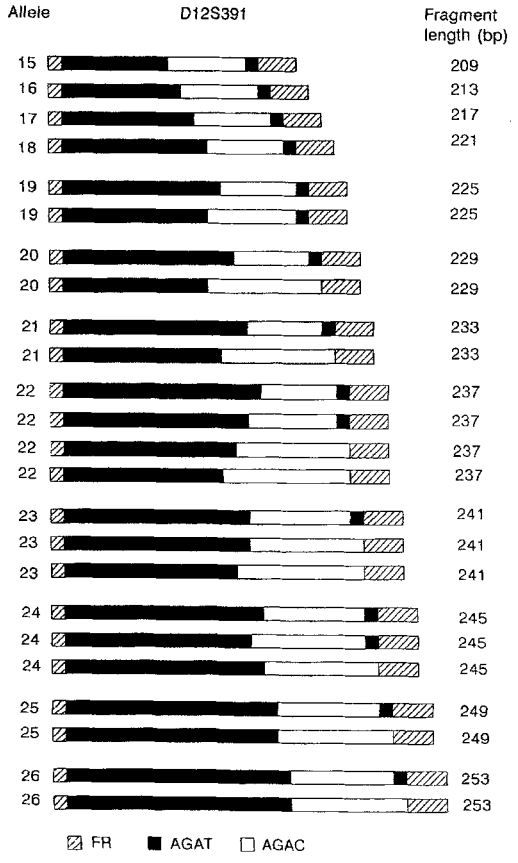

Fig. 2 Sequence structure of the different alleles found in our study

Table 1 Sequence composition of the D12S391 STR alleles

Allelic group	No. sequenced alleles	Sequence
15 (209 bp)	6	$(AGAT)_8$ $(AGAC)_6$ (AGAT)
16 (213 bp)	6	$(AGAT)_9$ $(AGAC)_6$ (AGAT)
17 (217 bp)	6	$(AGAT)_{10}$ $(AGAC)_6$ (AGAT)
18 (221 bp)	10	$(AGAT)_{11}$ $(AGAC)_6$ (AGAT)
19 (225 bp)	10	$(AGAT)_{12}$ $(AGAC)_6$ (AGAT)
19^2 (225 bp)	3	$(AGAT)_{11}$ $(AGAC)_7$ (AGAT)
20 (229 bp)	4	$(AGAT)_{11}$ $(AGAC)_9$ (AGAT)
20^2 (229 bp)	9	$(AGAT)_{11}$ $(AGAC)_9$
21 (233 bp)	3	$(AGAT)_{14}$ $(AGAC)_6$ (AGAT)
21^2 (233 bp)	7	$(AGAT)_{12}$ $(AGAC)_9$
22 (237 bp)	3	$(AGAT)_{15}$ $(AGAC)_6$ (AGAT)
22^2 (237 bp)	2	$(AGAT)_{14}$ $(AGAC)_7$ (AGAT)
22^3 (237 bp)	2	$(AGAT)_{13}$ $(AGAC)_9$
22^4 (237 bp)	6	$(AGAT)_{12}$ $(AGAC)_{10}$
23 (241 bp)	2	$(AGAT)_{14}$ $(AGAC)_8$ (AGAT)
23^2 (241 bp)	5	$(AGAT)_{14}$ $(AGAC)_9$
23^3 (241 bp)	1	$(AGAT)_{13}$ $(AGAC)_{10}$
24 (245 bp)	2	$(AGAT)_{15}$ $(AGAC)_8$ (AGAT)
24^2 (245 bp)	2	$(AGAT)_{14}$ $(AGAC)_9$ (AGAT)
24^3 (245 bp)	4	$(AGAT)_{15}$ $(AGAC)_9$
25 (249 bp)	3	$(AGAT)_{16}$ $(AGAC)_8$ (AGAT)
25^2 (249 bp)	5	$(AGAT)_{16}$ $(AGAC)_9$
26 (253 bp)	1	$(AGAT)_{17}$ $(AGAC)_8$ (AGAT)
26^2 (253 bp)	1	$(AGAT)_{17}$ $(AGAC)_9$

Small alleles from allele 15 to 18 have variations in the (AGAT) repeating unit only, whereas the rest of the sequence is constant. The complexity of this system increases in larger alleles with variations in the number of both AGAT and AGAC motifs.

Additional variation in the largest alleles (from allele 19 to allele 27) occurs due to the presence or absence of a AGAT unit at the end of the tandem array.

Population data

The allele and genotype frequencies from Galicia and from Münster are given in Tables 2 and 3.

Hardy-Weinberg equilibrium was tested using conventional Pearson's chi-square methods (χ^2), and the exact test proposed by Guo and Thomson (1992). The two populations were in Hardy Weinberg equilibrium for both tests. Other statistical parameters of genetic and forensic interest are shown in Table 3.

Comparison of populations

No other population studies for this STR have been published but no significant differences were observed be-

Table 2 D12S391 genotype frequencies of the populations studied

Genotypes	Galicia		Germany	
	Observed	Expected	Observed	Expected
15/15	1	0.25		
15/17	3	1.06	1	1.23
15/18	2	2.86	4	2.05
15/19	3	1.96	1	1.40
15/20	1	1.29		
15/21	1	1.29	3	1.23
15/22			1	1.20
15/23	1	1.06	1	1.02
16/17	1	0.57	1	1.01
16/18	2	1.54	1	1.68
16/19	1	1.05	2	1.15
16/20	1	0.70	1	1.36
16/21	1	0.70	1	1.01
16/22			1	0.98
16/23			1	0.84
16/24	1	0.36	1	0.31
17/17	1	1.10	1	2.35
17/18	4	5.94	11	7.82
17/19	1	4.07	4	5.36
17/20	3	2.68	5	6.37
17/21	2	2.68	7	4.69
17/22	5	3.66	7	4.58
17/23	4	2.20	1	3.91
17/24	1	1.38	2	1.45
17/25			1	0.89
17/26	1	0.16		
18/18	7	8.03	7	6.52
18/19	13	10.99	4	8.94
18/20	7	7.26	16	10.61
18/21	6	7.26	4	7.82
18/22	11	9.89	7	7.63
18/23	7	5.94	8	6.52
18/24	6	3.74		
18/25	1	1.10	1	1.48
19/19	2	3.77	7	3.06
19/20	5	4.97	5	7.28
19/21	4	4.97	5	5.36
19/22	9	6.78	3	5.23
19/23	6	4.07	3	4.47
19/24	2	2.56	4	1.66
19/25	2	0.75	3	1.02
20/20	2	1.64	3	4.32
20/21	5	3.28	6	6.37
20/22	4	4.47	8	6.22
20/23	2	2.68	5	5.31
20/24	1	1.69	3	1.97
20/25			2	1.21
21/21	2	1.64	2	2.35
21/22	7	4.47	5	4.58
21/23	3	2.68	6	3.91
21/24			1	1.45
22/22	1	3.05	3	2.24
22/23	2	3.66	3	3.81
22/24	5	2.30		
23/23			2	1.63
23/24	1	1.38	2	1.21
23/25	1	0.41	1	0.74
25/26	1	0.03		
Others		11.98		9.19
Total	166	166.00	188	188.00

Table 3 Allele frequencies of D12S391 system in both populations studied and other statistical parameters of genetic and forensic interest

Allele	Galicia	Germany
15	0.039	0.029
16	0.021	0.024
17	0.080	0.112
18	0.220	0.186
19	0.152	0.127
20	0.099	0.152
21	0.099	0.112
22	0.135	0.109
23	0.082	0.093
24	0.052	0.035
25	0.015	0.021
26	0.006	
	Galicia	Germany
Expected heterozygosity:	0.87539	0.87982
Observed heterozygosity:	0.90361	0.86702
CE:	0.74542	0.75204
DP:	0.97149	0.97286
Exact Test p:	0.54132	0.5776

tween the two populations studied here ($\chi^2 = 62.19$; df = 57; $P > 0.1$).

Meioses

No mutations were obtained in a total of 40 paternal and 35 maternal meioses.

Discussion

Although this STR system was found with a GATA probe, for simplicity and nomenclature purposes AGAT and AGAC were defined as the repetitive motifs within this STR.

The allele designation is based on the number of repeats in the repeating unit and it has been therefore defined for $(AGAR)_n$. 12 different allelic groups were found with a total number of repeats ranging from $(AGAR)_{15}$ to $(AGAR)_{26}$.

Sequence data of 103 different fragments of D12S391 showed that this STR has not only a length polymorphism

but also some structural variations as well. 24 different alleles were found. Therefore the D12S391 system can be classified as a STR with intermediate microvariation (Brinkmann 1996) or compound STR (Urquhart et al. 1994).

As for other STRs in this group (i.e. HUMVWFA31/A) small alleles have a considerable conservativeness of their structure with variations only in the number of AGAT units. From the allele 19 variations in the number of AGAT and AGAC units occur, probably due to evolutionary events with original variations in the number of AGAT units, then occuring T→C transition.

What is an exception in this system is the high heterozygosity with a relatively simple structure. No insertions or delections of single units were found and the structure resembles that of the most simple STR. The heterozygosity of 0.9 is similar to that of complex STRs with high microvariation.

There are no significant differences in the two caucasian population studied and both are in Hardy-Weinberg equilibrium for this system. In both populations allele 18 is more common and the frequencies range from 0.0076 (allele 26) to 0.2252 (allele 18) in the Galician population and from 0.021 (allele25) to 0.186 (allele18) in the German population.

Although most of the structural variation of this STR can be detected using non-denaturing conditions which increases the discrimination power of this system and therefore its forensic usefulness, we recommend the use of denaturing conditions for standardization reasons. This STR has a normal electrophoretic behaviour but the use of non-denaturing conditions would necessitate a rigorous standardization of electrophoretic conditions.

This STR can be typed using manual electrophoretic methods or automated sequencers. We have used two different sequencers and we always obtained reproductible results both in sequencing and in fragment analysis. Preliminary experiments have shown that this STR is suitable for multiplex combinations (data not shown).

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

Acknowledgements This work was supported by grants from Xunta de Galicia (XUGA 20801B95) and Fondo de Investigación Sanitaria (FISS 94/1053).

References

Brinkmann B (1992) The use of STRs in stain analysis. In: Proceedings from the Third International Symposium on Human Identification. Promega Corporation, Madison, USA, pp 357–373

Brinkmann B (1996) The STR approach. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics 6. Springer, Berlin Heidelberg (in press)

DNA recommendations-1992 report concerning recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms. Int J Legal Med 105:63–64

DNA recommendations-1994 report concerning further recommendations of the DNA Commission of the International Society for Forensic Haemogenetics regarding PCR-based polymorphisms in STR 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

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

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

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

Lareu MV, Barros F, Salas A, Pestoni C, Muñoz I, Rodríguez-Calvo, MS, Carracedo A (1996). D12S391: a highly useful STR for forensic purposes. in: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics 6. Springer, Berlin Heidelberg (in press)

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

Maniatis T, Fritsch EF, Sambrook, J (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

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

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

Ohno Y, Sebetan IM, Akaishi S (1982) A simple method for calculating the probability of excluding paternity with any number of codominant alleles. Forensic Sci Int 19:93–98

Pestoni C, Lareu MV, Rodríguez MS, 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

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