Int J Legal Med (2001) 115 : 40–44

ORIGINAL ARTICLE

G. Mertens · N. Mommers · L. Boutrand · M. Gielis
A. Vandenberghe

Flemish population data and sequence structure of the hypervariable tetranucleotide repeat locus D12S1090

Received: 5 October 2000 / Accepted: 2 April 2001

Abstract The allele frequency and sequence structure of the STR locus D12S1090 were investigated in 598 Flemish individuals. The locus shows a complex organisation with repetitions of GATA interrupted by TA and other tetra- and pentanucleotide blocks. No deviation from Hardy-Weinberg equilibrium was observed. The extensive polymorphism makes it a powerful tool for identity as well as paternity testing and even permits differentiation of closely related populations, such as Flemish and Germans. D12S1090 seems to be one of the most informative STRs, however, as seen in other highly variable STRs, the observed mutation frequency of 5.1×10^{-3}, is relatively high.

Keywords Short tandem repeats · D12S1020 · Flemish · Population genetics · Mutation rate

Introduction

Short tandem repeat (STR) polymorphisms amplified by PCR are at present the most widely used genetic markers for identity and paternity testing purposes. They are highly sensitive and allow typing of stains that are severely degraded as the amplified fragments are usually shorter than 300 bp [1]. On the other hand, most STRs, having only 3–6 common alleles, are distinctly less polymorphic than VNTRs [2]. Therefore, the STR approach requires typing of a larger number of systems to obtain comparable results, detracting from some its advantages, e.g. the saving of time and material. One possibility to limit the number of STR loci is typing highly polymorphic STRs such as ACTBP2 [3]. We thus investigated the tetranucleotide (GATA) repeat locus D12S1090 (GDB accession number GDB:376560, also known as GATA5A09) in a population sample from Flanders, the Dutch speaking part of Belgium. Although this STR is considered to be highly polymorphic [4], compared with other STRs, it has only scarcely been studied.

G. Mertens (✉) · N. Mommers · M. Gielis
Antwerp Blood Transfusion Centre, Wilrijkstraat 8,
2650 Edegem, Belgium

L. Boutrand · A. Vandenberghe
Laboratory of Human Molecular Genetics, Faculty of Pharmacy,
University Claude Bernard Lyon I, 69008 Lyon Cedex 08, France
e-mail: gerd.mertens@bl.rodekruis.be,
Tel.: +32-3-8290000, Fax: +32-3-8290161

Material and methods

Genomic DNA was extracted from blood of 598 unrelated Flemish Caucasians by the salting out method [5]. These individuals were the mother and alleged father in 299 paternity cases and 433 paternity trios were used to search for mutations.

For PCR the Multiplex I kit (Lifecodes, Stamford, Conn.) was used according to the manufacturer's instructions, thus co-amplifying D12S1090 with D3S1744 and D18S849. Primer sequences for D12S1090 were neither provided by the producer nor available from the GDB, but were determined as follows. PCR products obtained by amplification of the Multiplex I kit were cloned in pGEM-T Easy (Promega, Madison, Wis.). Of the clones, 20 were picked at random and sequenced using T7 and T6 primers specific for the cloning vector. This allowed the identification of the primer-pair sequences for the three loci. Amplification with these primers and comparison of allele lengths led to the identification of the D12S1090 locus with its corresponding primers:

– Forward: AAG CCC AAA GAT GTA AGG CT
– Reverse: ACC AAC CTA GGA AAC ACA GT

This primer pair was further used to characterise the D12S1090 alleles. Alleles from homozygous individuals were sequenced directly on genomic DNA while alleles from heterozygotes were cloned. Sufficient clones chosen at random, were sequenced to obtain both allelic sequences. The ABI Prism dye terminator cycle-sequencing ready reaction kit and an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, Calif.) were used for sequencing.

Amplification was performed using 5 ng of template DNA in a 25 µl volume. The reaction mixture included 0.625 U Taq polymerase (Goldstar, Eurogentec, Belgium), 2.5 µl primer mix, 200 µM of each nucleotide, 2.5 µl 10 × PCR buffer. PCR was carried out in a GeneAmp PCR System 9600 (Applied Biosystems, Foster City, Calif.) thermal cycler using the following conditions: 30 cycles of denaturation at 95 °C for 1 min, annealing at 57 °C for 1 min, extension at 72 °C for 1 min. After electrophoresis on a denaturing 4% (w/v, 39:1 acrylamide:bisacrylamide, 7 M urea, 0.4 mm thick) gel, alleles were revealed by silver staining [6].

Alleles were named according to the recommendations of the DNA Commission of the International Society for Forensic Haemogenetics [7] and were based on the number of repeats. Alleles were identified with reference to the ladder supplied by the manufacturer. Using the same strategy as for allele characterisation, the ladder supplied by Lifecodes was cloned in pGEM-T Easy (Promega, Madison, Wis.) and the alleles were sequenced.

Allelic frequencies were estimated by direct gene counting. Hardy-Weinberg equilibrium (HWE) was tested by means of the exact test of Guo and Thompson [8] using Arlequin Version 2000 software [9] to perform the calculations. Population differentiation was examined by the exact test of Raymond and Rousset [10] as implemented in the Arlequin package.

Some parameters of forensic interest were determined using Powerstat software (Promega, Madison, Wis.). The polymorphism information content (PIC) was calculated as described by Botstein et al. [11], the power of discrimination (POD) was computed following Jones [12], the power of exclusion (PEX) and the mean paternity index (MPI) were enumerated according to Brenner and Morris [13].

Table 1 Structure of the repeat regions in 66 sequenced alleles of D12S1090 (small sequence differences 5′ of the repeat are *underscored*; *n* number of sequenced alleles)

Allele	bp	*n*	Sequence						
9	209	2	TATA GATAGATA	6 × (GATA)					
10	213	1	TATA GATAGATA	7 × (GATA)					
11	217	1	TATA GATAGATA	8 × (GATA)					
12	221	3	TATA GATAGATA	9 × (GATA)					
13	225	3	TATA GATAGATA	10 × (GATA)					
13[a]	225	1	TATA GATAGATA	10 × (GATA)					
14	229	1	TATA GATAGATA	11 × (GATA)					
15	233	1	TATA GATAGATA	12 × (GATA)					
16	238	3	TATA GATAGATA	6 × (GATA)		GATTA	6 × (GATA)		
17	241	1	TATA GATACATA	14 × (GATA)					
17	242	1	TATA GATAGATA	7 × (GATA)		GATTA	6 × (GATA)		
18	246	2	GATA GATAGATA	9 × (GATA)		GATTA	5 × (GATA)		
18	246	3	TATA GATAGATA	8 × (GATA)		GATTA	6 × (GATA)		
18	246	1	TATA GATAGATA	9 × (GATA)	GATT	GATTA	4 × (GATA)		
18	248	3	TATA GATAGATA	9 × (GATA) TA 2 × (GATA)		GATTA	3 × (GATA)		
19	250	1	TATA GATAGATA	10 × (GATA)		GATTA	5 × (GATA)		
19	250	2	TATA GATAGATA	9 × (GATA)		GATTA	6 × (GATA)		
19	250	1	TATA GATAGATA	10 × (GATA)		GATTA	4 × (GATA)		
20	254	3	TATA GATAGATA	10 × (GATA)		GATTA	6 × (GATA)		
20	254	1	GATA GATAGATA	11 × (GATA)	GATT	GATTA	4 × (GATA)		
20	254	1	TATA GATACATA	9 × (GATA)		GATTA	7 × (GATA)		
20	254	2	TATA GATACATA	8 × (GATA)		GATTA	8 × (GATA)		
21	258	1	TATA GATAGATA	12 × (GATA)		GATTA	5 × (GATA)		
21	258	1	TATA GATACATA	10 × (GATA)		GATTA	7 × (GATA)		
21	258	1	TATA GATAGATA	11 × (GATA)		GATTA	6 × (GATA)		
21[a]	258	1	TATA GATAGATA	11 × (GATA)		GATTA	6 × (GATA)		
22	262	1	TATA GATAGATA	11 × (GATA)		GATTA	7 × (GATA)		
22	262	3	TATA GATACATA	11 × (GATA)		GATTA	7 × (GATA)		
22	262	1	TATA GATACATA	10 × (GATA)		GATTA	8 × (GATA)		
23	266	2	TATA GATACATA	12 × (GATA)		GATTA	7 × (GATA)		
23	266	1	TATA GATAGATA	10 × (GATA)	GATT	GATTA	8 × (GATA)		
24	270	1	TATA GATAGATA	10 × (GATA)	GATT	GATTA	9 × (GATA)		
25	273	1	TATA GATAGATA	10 × (GATA)		GATTA	4 × (GATA)	TA 2 × (GATA)	GATTA 3 × (GATA)
25	273	1	TATA GATACATA	× (GATA)		GATTA	5 × (GATA)	TA 2 × (GATA)	GATTA 3 × (GATA)
25	274	1	TATA GATAGATA	11 × (GATA)	GATT	GATTA	9 × (GATA)		
25	274	1	TATA GATACATA	11 × (GATA)		GATTA	10 × (GATA)		
26	277	3	TATA GATAGATA	10 × (GATA)		GATTA	5 × (GATA)	TA 2 × (GATA)	GATTA 3 × (GATA)
26	278	2	TATA GATACATA	11 × (GATA)		GATTA	11 × (GATA)		
27	281	1	TATA GATAGATA	11 × (GATA)		GATTA	5 × (GATA)	TA 2 × (GATA)	GATTA 3 × (GATA)
27[a]	281	1	TATA GATAGATA	11 × (GATA)		GATTA	5 × (GATA)	TA 2 × (GATA)	GATTA 3 × (GATA)
28	285	2	TATA GATAGATA	13 × (GATA)		GATTA	4 × (GATA)	TA 2 × (GATA)	GATTA 3 × (GATA)
29	289	1	TATA GATAGATA	14 × (GATA)		GATTA	4 × (GATA)	TA 2 × (GATA)	GATTA 3 × (GATA)
30	293	1	TATA GATAGATA	14 × (GATA)		GATTA	5 × (GATA)	TA 2 × (GATA)	GATTA 3 × (GATA)

[a] Sequenced alleles from the Lifecodes ladder

```
aagcccaaag atgtaaggct GGGAGATAGC AAACCAGGTC TTGATGTTGC AGATTTGTTG
TATATATATT TATATAGATA GATAGATAGA TAGATAGATA GATAGATAGA TAGATGTTAT
AGATATATAT ATTATGTATT TGATATAAAC TAATGCATAT CAAACATTGT TTTTTTGGTA
GAGATGAGAT CTCactgtgt ttcctaggtt ggt
```

Fig. 1 Sequence of allele 10 of locus D12S1090. The total length is 231 bp. The six GATA repeats are preceded by a TATA tetranucleotide (Primer sequences are written in *lowercase*, the variable region is presented in *italics* and is detailed for other alleles in Table 1)

Results and discussion

Sequence structure of D12S1090 alleles

In 598 Flemish individuals, 22 different types of alleles were found using denaturing PAGE followed by silver staining. The sequence structure was determined from 66 alleles, including examples of all 22 types found in this Flemish population sample. The locus shows a clearly complex organisation, the complexity increasing with the allelic number (Table 1 and Fig. 1). Only alleles 18 and 20 contained the corresponding number of GATA repeats, while in all other alleles a TATA tetranucleotide block instead of GATA occurs at the 5′-end of the sequence. We observed five major types of sequence organisation. The first, as present in alleles 9–15 and 17, contains uninterrupted repetitions of the GATA core. In the second type, seen in alleles 16–23 and in 25 and 26, the regular repetition of GATA is interrupted by GATTA. In alleles 18, 20, 23, and 25, we found examples of breaking the GATA repetitions by GATT, followed by GATTA. A fourth type of non-consensus organisation (allele 18) has a TA insertion. Finally, from allele 25 onwards, the GATA repetitions are interrupted by TA and GATTA. The sequence analysis revealed that the majority of the alleles are composed of compound repeats having the same size but different sequence structure. Examples were observed for alleles 17 through 26, showing two (allele 23) to even four (allele 20) structural forms with the same number of nucleotides but a different sequence. While Perlee et al. [14] failed to reveal sequence differences between same size alleles of D12S1090 from unrelated individuals, the kind of homoplasmy we observed has been demonstrated in other STR loci such as D14S299 [15] and D22S683 [16]. Therefore, D12S1090 not only has a length polymorphism, but also some structural variation in most of its alleles.

The allelic structure is in agreement with the previous description by Perlee et al. [14], except for the fact that we find it systematically and repeatedly three nucleotides shorter. The Lifecodes ladder was sequenced, showing concordance with the allelic designations used by Lifecodes and in our population study. In their ladder sequence however, we found that Lifecodes modified the reversed primer sequence in CAACCAGATCGAAACACACT. Furthermore, we observed a different sequence for ladder allele 27 (Table 1) than published by Perlee et al. [14]. The sequence for allele 27 proposed by Perlee et al. is probably erratic, while our result fits with the general structure of the shorter and longer alleles.

Population data

For HWE calculations, allelic frequency counting and population comparison, same size alleles with different sequences were pooled, as well as allelic forms differing 1 bp in length from the consensus repeat. The following alleles are concerned: 17, 18, 19, 20, 21, 22, 23, 25 and 26 (see Table 1).

A total of 147 genotypes were found in the 598 subjects tested. Table 2 gives the frequencies of the alleles of D12S1020 in these Flemish Caucasians. The frequencies show a bimodal distribution pattern. Only four alleles were more frequent than 10%: 20 (11.2%), 22 (11.0%), 23 (10.4%) and 24 (10.3%).

No significant deviation from HWE was observed (exact test $p = 0.594$). The statistical parameters of forensic interest can be found in Table 2. The degree of polymorphism of D12S1090 is higher than that of most STRs used in forensic practice, ranging between ACTBP2 [17] and D12S391 [18], making it a very powerful tool for identity as well as parentage testing. Since the chromosomal distance between D12S1090 and D12S391 (GDB:686474) is only 32 cM, these loci are in linkage disequilibrium, precluding their combined use in paternity casework.

In addition, D12S1090 proved to be a sensitive marker for population differentiation. While comparing the allelic frequencies of several populations (Table 3), significant differences were shown between the Flemish and Moroc-

Table 2 D12S1020 allele frequencies and forensic parameters in a Flemish population sample (598 individuals)

Allele	Frequency	Allele	Frequency
9	0.002	20	0.112
10	0.008	21	0.091
11	0.038	22	0.110
12	0.053	23	0.104
13	0.024	24	0.103
14	0.018	25	0.066
15	0.008	26	0.080
16	0.008	27	0.058
17	0.005	28	0.012
18	0.015	29	0.008
19	0.074	30	0.002
H obs	0.918		
H exp	0.921		
POD	0.986		
PIC	0.91		
CE	0.811		
MPI	5.42		

Table 3 Comparison of allele frequencies for D12S1090 between the Flemish and other populations [18, 19].

D12S1090	German	US Caucasian	Moroccan	Colombian Mestizo	US Hispanic	Colombian Black	US Black	US Oriental
Flemish	$p = 0.000$	$p = 0.723$	$p = 0.001$	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$
n	149	110	49	486	231	248	103	208

p-value of the exact test of population differentiation (steps in the Markov chain 10,000, dememorisation steps 1,000).
n: number of individuals

cans (own data, collected on Moroccans residing in Belgium), Colombian Mestizos and Blacks [19], US Blacks [20], US Hispanics and US Orientals (data provided by Lifecodes). Even when the Flemish data were compared with the neighbouring Germans [4], a clear distinction was possible! The only population which D12S1020 could not separate from the Flemish, were US Caucasians [20]. When all nine populations were compared pairwise, all pairs proved significantly (exact test for population differentiation $p < 0.05$) different, except for Flemish – US Caucasians ($p = 0.723$), Colombian Black – Moroccan ($p = 0.502$) and US Black – Moroccan ($p = 0.485$). It would be of interest to obtain data on more Caucasian populations to evaluate whether or not they can be differentiated on the basis of the D12S1090 polymorphism. Thus, this STR also appears to be of potential interest for anthropological research.

Mutations

We observed isolated parent/child mismatches in 3 out of 584 (433 maternal, 151 paternal) meioses, yielding an overall mutation rate of 5.1×10^{-3}. Although small in number, the age distribution of the mutations seemed in accordance with the overall distribution of parental age at conception (Table 4).

Of the mutations two were of maternal origin, one contraction of the allele 13 to a 12 repeat unit (aged 24 years) and one contraction of the allele 27 to a mutant 26 allele (aged 27 years). The paternal mutation (aged 28 years) was an expansion with one GATA block of the allele 28 leading to an allele 29. The structure of these mutations was confirmed by sequencing. Compared with other STRs, the mutation rate was close to the 6.8×10^{-3} reported for ACTBP2 [21], which also has a slightly higher polymorphism and mean number of uninterrupted repeats than D12S10190. It supports the observation [21] that the mutation rate correlates with the geometric mean of the repeat lengths.

Table 4 Distribution of age of mothers and fathers at the time of conception

Age at conception (years)	Number of mothers (%)	Number of fathers (%)
15–19	18 (4.0)	2 (1.3)
20–24	102 (23.6)	24 (15.9)
25–29	141 (32.6)	39 (25.8)
30–34	118 (27.2)	37 (24.5)
35–39	47 (10.9)	25 (16.6)
40–44	7 (1.7)	9 (6.0)
45–49	–	9 (6.0)
50–54	–	3 (2.0)
55–59	–	2 (1.3)
60–64	–	1 (0.7)

References

1. Brinkmann B (1992) The use of STRs in stain analysis. In: Proceedings from the third international symposium on human identification. Promega, Madison, USA, pp 357–373
2. 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
3. Polymeropoulos MJ, Rath DS, Xiao H, Merril CR (1992) Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2). Nucleic Acids Res 20:1432
4. Stein C, Lange T, Ferecik S, Grosse-Wilde H, Henssge C (1998) German population data of three tetrameric short tandem repeat loci – D3S1744, D12S1090 and D18S849. Forensic Sci Int 91: 103–107
5. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215
6. Bassam BG, Annoles GC, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80–83
7. Bär W, Brinkmann B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B (1997) DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. Int J Legal Med 110:175–176
8. Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48: 361–372
9. Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2000: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland (http://anthro.unige.ch/arlequin)
10. Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280–1283
11. Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331
12. Jones DA (1972) Blood samples: probability of discrimination. J Forensic Sci Soc 12:355–359
13. Brenner C, Morris J (1990) Paternity index calculations in single locus hypervariable DNA probes: validation and other studies. In: Proceedings for the International Symposium on Human Identification 1989. Promega Corporation, Madison, WI, pp 21–53
14. Perlee L, Neuweiler J, Balazs I (1996) Analysis of sequence variations in the alleles from three STR loci. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics 6. Springer, Berlin Heidelberg New York, pp 52–54

15. Yoshimoto T, Tamaki K, Katsumata S, Huang XL, Uchichi R, Tanaka M, Uchida H, Yamamoto T, Chen S, Armour J, Katsumata Y (1999) Sequence analysis at a microsatellite locus D14S299 (wg1c5) and population genetic comparisons. Int J Legal Med 113:15–18
16. Barral S, Lareu MV, Salas M, Carracedo A (2000) Sequence variation of two hypervariable short tandem repeats at the D22S683 and D6S477 loci. Int J Legal Med 113:146–149
17. Klintschar M, Neuhuber F (1998) A study on the short tandem repeat system ACTBP2 (SE33) in an Austrian population sample. Int J Legal Med 111:329–330
18. Lareu MV, Pestoni C, Schürenkamp M, Rand S, Brinkmann B, Carracedo A (1996) A highly variable STR at the D12S391 locus. Int J Legal Med 109:134–138
19. Yunis JJ, Garcia O, Baena A, Arboleda G, Uriarte I, Yunis E (2000) Population frequency for the short tandem repeat loci D18S849, D3S1744 and D12S10190 in Caucasian-Mestizo and African descent populations of Colombia. J Forensic Sci 45: 429–431
20. Neuweiler J, Perlee L, Venturini J, Balazs I (1996) Properties of an STR multiplex marker system suitable for paternity and forensic determinations. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics 6. Springer, Berlin Heidelberg New York, pp 148–150
21. Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62: 1408–1415