Int J Legal Med (2000) 113 : 76–80

ORIGINAL ARTICLE

Eva-Maria Dauber · Barbara Glock
Dieter W. M. Schwartz · Wolfgang R. Mayr

Further sequence and length variation at the STR loci HumFES/FPS, HumVWA, HumFGA and D12S391

Received: 14 October 1998 / Received in revised form: 18 February 1999

Abstract This paper reports population data and statistics for the HumFES/FPS, HumVWA, HumFGA and D12S391 loci in Austria. The sequences of some rare and new variant alleles which have been identified in the course of the present population study and other investigations are described. Sequence variation occurred in a HumFES/FPS allele revealing an $(ATTT)_9$ structure and an A to C transversion in the 5′ flanking region. At the HumVWA locus the structural type of the common allele 14 has been found in one allele 13 and in three examples of allele 15. Additionally the TCTA $(TCTG)_3(TCTA)_n$ structure has been observed in three examples of allele 13 and one allele 14, which is very uncommon. Another allele 14 showed a C to T transition in the third of nine TCTA repeats. The sequences of three length variations at the HumFGA locus, namely the alleles 16, 19.2 and 21.2 are reported. At the D12S391 locus a novel 19.1 allele was found in this study. An extended nomenclature is proposed for the HumVWA locus to denominate sequence variants in a precise but simple way.

Key words Short tandem repeats (STR) · Sequencing · Polymerase chain reaction · Polymorphism · Alleles

Introduction

Short tandem repeats (STRs) have become routine markers for human identity and paternity testing purposes in the forensic field. Accurate typing of STRs requires a precise knowledge of the structural variation of alleles. In the course of population studies to investigate allele distributions in the Austrian population at the STR loci HumFES/FPS [1], HumVWA [2], HumFGA [3] and D12S391 [4], novel and rare variant alleles have been identified. In additional, but selected Austrian blood samples, which could therefore not be included to the population sample, more variant alleles have been found. This paper reports population data and sequences of novel and rare alleles at the four STR loci in comparison with commonly occurring alleles.

E.-M. Dauber (✉) · B. Glock · D. W. M. Schwartz · W.R. Mayr
Clinical Department for Blood Group Serology,
University of Vienna, Währinger Guertel 18-20,
A-1090 Wien, Austria
e-mail: eva.maria.dauber@univie.ac.at
Tel. +431-40400-5337; Fax +431-40400-5321

Materials and methods

Genomic DNA was extracted from peripheral blood samples taken from healthy, unrelated Caucasoid individuals of the Vienna and Lower Austria region by the salting out method [5].

Amplification was carried out in a Hybaid Omnigene Thermocycler using 8 ng of genomic DNA, 0.4 μM each primer (HumFES/FPS: 1 μM), 200 μM dNTPs (Pharmacia Biotech), 1 × PCR buffer (50 mM KCl, 10 mM TrisCl, pH 8.8 at 25 °C, 0.1% Triton-X-100 and 1.5 mM $MgCl_2$) except for HumFES/FPS, where 0.8 × PCR buffer was used. After a first denaturation step (98 °C, 5 min), 2 U DNA polymerase (Dynazyme II, Finn Zymes Oy) was added to a final reaction volume of 50 μl. The oligonucleotide primer sequences and amplification conditions were as published elsewhere [1, 2, 4, 6–8].

Native polyacrylamide gel electrophoresis was carried out for HumFES/FPS (T = 6%; 112 mM Tris-acetate), HumVWA (T = 7%, 120 mM Tris-acetate) and HumFGA (T = 6%, 120 mM Tris-acetate) in Tris-acetate as rehydration buffer and 0.2 mM Tris-tricine as electrode buffer as described elsewhere [9].

Denaturing gel electrophoresis for D12S391 and solid phase single-strand sequencing were conducted on an A.L.F. sequencer as previously described [10, 11].

To calculate χ^2 analysis for examination of Hardy-Weinberg expectations and indicator values for the discriminatory potential of the four polymorphic loci, the HWE-Analysis software Version 3.0 (Christoph Puers, Institute for Forensic Medicine, Münster, Germany) was used.

Results and discussion

HumFES/FPS

Alleles at the HumFES/FPS locus showed a simple repeat structure following the pattern $(ATTT)_n$ ranging in size from 211–239 bp. Some of the 219 bp and 223 bp alleles

showed an A to C transversion at position 34 in the 5′ flanking region (ATTT strand) leading to a different electrophoretic mobility in native polyacrylamide gels [7, 12]. The same base substitution has been observed for rare 207, 227 and 231 bp alleles in various populations [13–15].

Table 1 Allele frequencies

HumFES/FPS *n* = 429		HumVWA *n* = 431	
Allele	Frequency	Allele	Frequency
8^{Ade} (8)	0.015	11	0.001
9^{Ade} (9)	0.001	13	0.001
9^{Cyt} (9a)	0.001	14	0.088
10^{Ade} (10)	0.031	rare 14	0.001
10^{Cyt} (10a)	0.245	new 14	0.001
11^{Ade} (11)	0.424	15	0.116
11^{Cyt} (11a)	0.022	rare 15	0.001
12^{Ade} (12)	0.221	16	0.190
13^{Ade} (13)	0.037	17	0.258
14^{Ade} (14)	0.001	18	0.216
		19	0.109
		20	0.020
HumFGA *n* = 518		D12S391 *n* = 317	
Allele	Frequency	Allele	Frequency
16	0.001	15	0.036
17	0.001	16	0.028
18	0.009	17	0.112
19	0.062	17.3	0.009
20	0.146	18	0.191
20.2	0.001	18.3	0.016
21	0.196	19	0.098
21.2	0.002	19.1	0.002
22	0.190	19.3	0.003
22.2	0.006	20	0.119
23	0.132	21	0.126
23.2	0.007	22	0.128
24	0.132	23	0.082
24.2	0.001	24	0.032
25	0.085	25	0.009
26	0.026	26	0.009
27	0.004		

In this study 10 different alleles were found in a population of 429 unrelated Caucasoid individuals including a 215 bp allele provisionally designated 9a due to its particular electrophoretic migration. Sequence analysis exhibited 9 reiterations of the core repeat unit and the same A to C transversion at position 34 of the 5′ flanking region as was already reported for the alleles 10a and 11a [7], but not yet for a 215 bp allele. These variant alleles should be named 9^{Cyt}, 10^{Cyt} and 11^{Cyt} respectively, whereas the consensus alleles should be named 9^{Ade}, $10^{Ade}$and 11^{Ade} [16].

As the allele 10^{Cyt} (allele 10a) is quite common in Caucasoid populations [14, this study], it seems more likely that the new variant allele 9^{Cyt} (allele 9a) has evolved by the deletion of a 4 bp repeat unit from an allele 10^{Cyt} (10a) than by a single base substitution in a consensus allele 9.

Base sequences of other alleles sequenced in this study were in absolute agreement with published data [7]. Allele frequencies and parameters of forensic interest are shown in Tables 1 and 2.

HumVWA

The HumVWA locus is classified as a compound STR locus consisting of two different repeat motifs TCTA and TCTG [12] and most VWA alleles (13, 15–22) follow the consensus repeat structure TCTA $(TCTG)_4$ $(TCTA)_n$ [7]. As shown in Fig. 1 three types of sequence variations have been described [17].

1. The first type of variation has been observed for all alleles 11 sequenced so far, some alleles 15 and 16 and isolated alleles 14; the rare TCTA $(TCTG)_5$ $(TCTA)_n$ structure has been considered as the same type of variation [7, 13, 17–20]

2. The second type of variation has been found restricted to black populations [13, 17]

3. The third type of variation was observed in the majority of alleles 14 (138 bp) and exceptional alleles 12 (130 bp) and 15 (142 bp) [7, 13, 18, 21].

In this study, 12 different alleles ranging in size from 126 bp to 162 bp (alleles 11, 13–20) have been observed in a population of 431 unrelated Caucasoid individuals (allele frequencies and forensic parameters see Tables 1 and 2). Alleles have been named by comparison with the

Table 2 Population statistics and parameters of forensic interest

Parameter	HumFES/FPS	HumVWA	HumFGA	D12S391
Observed heterozygosity	0.695	0.821	0.855	0.868
Expected heterozygosity	0.709	0.818	0.852	0.867
χ^2	20.71	31.96	43.82	40.03
df	28	28	36	28
p	0.834	0.663	0.178	0.295
Mean paternity exclusion chance	0.468	0.636	0.701	0.728
Mean exclusion probability	0.442	0.633	0.700	0.729
Polymorphism information content	0.611	0.792	0.834	0.851
Probability of match	0.129	0.062	0.041	0.038
Discrimination power	0.871	0.938	0.959	0.962

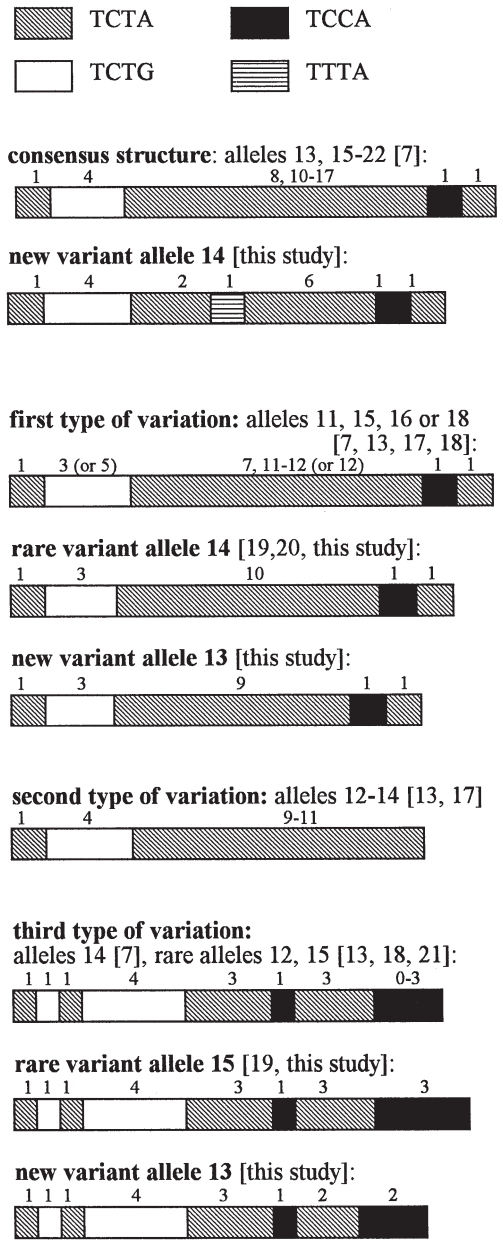

Fig. 1 HumVWA: structural variation and sequencing results repeat and 3 flanking region (except the last base)

sequenced allelic ladder except for three alleles, which showed an electrophoretic mobility which did not match with the ladder alleles: sequencing revealed a new and a rare variant allele 14 (138 bp) as well as a rare variant allele 15 (142 bp) (see Fig. 1).

The base sequence of the new variant allele 14 corresponds best to the consensus repeat pattern apart from a C to T transition in the third of the nine consecutive TCTA repeats, which is very unusual for this locus. The most likely explanation for the generation of this allele is a single base substitution in a TCTA repeat of a consensus allele 14, but it must be noted that the sequence of this putative precursor allele has not yet been reported.

Another rare variant allele was an allele 14 observed in the population sample due to its distinct electrophoretic behaviour. In contrast to most of the 138 bp alleles it exhibited a structure which corresponds to the first type of variation. An identically structured 138 bp allele has only been reported twice so far: in one case the variant allele originated from an individual of Asian origin [19], in the other case the ethnic origin of the individual carrying the variant was not stated [20]. Therefore the rare variant allele 14 sequenced in this study must be considered as new variant allele in this population. Out of four alleles 13 (134 bp) sequenced in this study (one allele derived from the population sample, two more examples of allele 13 have been identified in additional Austrian blood samples) three can be related to the first type of variation. These 134 bp alleles are in contrast to previously published data [7] and must therefore be considered as novel alleles. The single allele 11 (126 bp) sequenced in this study was identical to all 126 bp alleles sequenced so far [13, 17, 22, 23] and corresponds to the first type of variation. The sequence structure of all the alleles 11 and 13–16 sequenced in this and other studies suggests that the first type of variation occurs more frequently in the smaller alleles.

The sequence structure of the rare variant allele 15 (142 bp) allele fits best to the third type of variation, which is typical for the common allele 14. In additional, selected Austrian blood samples two identically structured alleles 15 and a structurally related new variant allele 13 (134 bp) were found due to their distinct electrophoretic migration and sequenced. The three examples of rare variant alleles 15 (142 bp) and the new variant allele 13 (134 bp) exhibiting the third type of variation sequenced in this study, differed from the common allele 14 (138 bp) by the loss of a TCTA repeat and an additional TCCA repeat respectively. As the common allele 14, which exhibits the third type of variation, is much more frequent in this population than the neighbouring structurally related variant alleles 13 and 15, it can therefore be suggested that they might have been generated by deletions/insertions of complete TCTA or TCCA repeats from the common type of allele 14. The existence of these allelic variants is very uncommon and unexpected in this population [24]. The only variant allele 15 reported so far derived from a Turkish Caucasian individual [21]. Therefore not only the variant allele 13 but also the variant allele 15 described in this paper have to be considered as new variants in this population.

Furthermore the third type of structural variation can no longer be considered as being restricted to the common allele 14 [24]. But more alleles will have to be sequenced to elucidate the evolutional mechanism for the generation of the first and third type of variation alleles, which has not yet been considered.

In the light of the repeat structure of the alleles investigated in this and previous sequencing studies [13, 17], alleles at the HumVWA locus have turned out to be more polymorphic than the first analyses suggested [7, 12]. It has already been stated that a nomenclature based purely on the number on consecutive TCTA and TCTG repeats is

not appropriate for this locus [13]. Another nomenclature, which includes the TCCA and TCTA repeats of the 3′ flanking region has therefore been proposed by other authors [17]. As a nomenclature for HumVWA alleles [7] is already established in the forensic field, the allelic designations of the rare and novel alleles have been assigned according to the fragment length [16, 25].

Hum FGA

Alleles at the HumFGA locus range in size from 168 to 294 bp and follow the repeat pattern $(YYBY)_n$ where Y represents C or T and B represents C, G or T. Some alleles possess an incomplete repeat and differ in size by 2 bp from the regular alleles. Larger alleles (30.2, 34.2, 46.2) differing markedly from the regular 4 bp increment pattern were found to be restricted to Afro-Carribean populations [6].

In this study 17 different alleles ranging in size from 172 bp to 216 bp have been observed in a population of 518 individuals (allele frequencies and parameters of forensic interest see Table 1 and 2) including the alleles 16 (172 bp) and 21.2 (194 bp), which are reported in the manual of a commercially available multiplex kit (Perkin Elmer – Profiler Plus) but base sequences have not yet been published. This is also the case for an isolated 19.2 allele (186 bp), which was found in our investigations in a single black individual of African origin (Table 3).

Allele 16 completes the series of alleles possessing complete tetrameric repeats (15, 17–29) and the alleles 19.2 and 21.2 complete the series of interalleles showing 2 bp deletions (18.2, 20.2, 22.2, 23.2, 24.2, 30.2, 34.2, 46.2) which were previously described [6]. As all interalleles are of low frequency we suggest one of the more frequent alleles 22.2 or 23.2 as a parent allele for the generation of the other alleles showing an incomplete repeat. The smaller and larger alleles might have been generated by the insertion/deletion of a complete repeat unit from the putative parent allele.

D12S391

D12S391 has been described as a compound STR locus consisting of two different repeat motifs (AGAT) and (AGAC). Most of the alleles follow the regular repeat structure $(AGAT)_{8–17}$ $(AGAC)_{6–10}$ [4]. Some alleles (220 bp, 224 bp) show an incomplete repeat (GAT) at the second position of the repeat array [8].

In this study an additional allele at this locus has been identified in a Caucasoid individual, which was 1 bp longer than the consensus allele 19. Sequence analysis exhibited a single thymine base after the third of 12 AGAT repeats. This allele was designated 19.1 [16, 25].

The underlying mechanism for the evolution of this allele could be the loss of three bases (AGA) of a complete AGAT repeat or the insertion of a single base. As the additional thymine base is located within a block of 12 AGAT repeats, the origin from a complete repeat and the loss of three bases seems to be the more likely mechanism. The coexistence of allelic variants 1 bp shorter (18.3, 19.3) and longer (19.1) than consensus alleles and additional sequence variations confirm the recommendation for the use of denaturing electrophoresis at this highly variable locus [4,8].

Apart from the base substitution in the 5′ flanking region of the novel HumFES/FPS alleles and the altered 3′ flanking region of the type three variation alleles, the 5′ and 3′ flanking regions of all sequenced alleles were identical with the GenBank (accession numbers: HumFES/FPS: X06292, HumVWA: M25858, HumFGA: M64982) and the EMBL sequence (accession number D12S391: G08921) and previously published data [4, 6, 7].

Nomenclature

Due to the increasing number of sequence variants at the HumVWA locus it has become more and more difficult to designate HumVWA variant alleles in a precise but simple way. Allelic designations according to the already established nomenclature refer to the number of repeats which is proportional to the fragment length. This is convenient for routine applications and interlaboratory exchange of data. Such a nomenclature, however, cannot meet the needs of an unambiguous designation of sequence variants, which is interesting for the study of microheterogeneities. An extended nomenclature should only complement the existing allelic designations and not replace them. Therefore we propose that an extended nomenclature should only be applied when dealing with base sequences, as it has already been introduced for the large number of different sequenced HLA alleles [26].

At the sequence level HumVWA alleles could be named in addition to the already existing allele designation by further figures written after a decimal point indicating the sequence structure. To avoid confusion with the designation of alleles possessing incomplete repeats a zero should be introduced between the allele name and the figures indicating the sequence structure. A consensus allele 15, for example, would therefore be denominated 15.01, where 15.0 refers to the number of repeats in the

Table 3 HumFGA: sequencing results

Allelic designation	Repeat region	Allele size in bp	Number of sequenced alleles
16	$(TTTC)_3$ TTTT TTCT $(CTTT)_8$ CTCC $(TTCC)_2$	172	2
19.2	$(TTTC)_3$ TTTT TT . . $(CTTT)_{12}$ CTCC $(TTCC)_2$	186	1
21.2	$(TTTC)_3$ TTTT TT . . $(CTTT)_{14}$ CTCC $(TTCC)_2$	194	2

originally defined repeat region and the last figure (1 in this case) to the consensus structure. Variant alleles with the same number of repeats but different base sequences would be named 15.02, 15.03 etc.

In an analogous way, it is possible to name alleles in systems with incomplete repeats: e.g. in D12S391, the allele with 19 repeats showing the $(AGAT)_{12}$ $(AGAC)_6$ (AGAT) structure would be named 19.01 and with the $(AGAT)_{11}(AGAC)_7$ (AGAT) structure 19.02; the first allele described with 19 repeats plus an incomplete repeat of 3 bases 19.31 and the second sequence structure with 19 repeats plus an incomplete repeat of 3 bases 19.32. The principle of this extended nomenclature could be easily applied to other STR loci if they show higher levels of microvariation.

References

1. Polymeropoulos MH, Rath DS, Xiao H, Merril RM (1991) Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES). Nucleic Acids Res 19:4018
2. Kimpton CP, Walton A, Gill P (1992) A further tetranucleotide repeat polymorphism in the vWF gene. Hum Mol Genet 1:28
3. Mills KA, Even D, Murray JC (1992) Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA). Hum Mol Genet 1:779
4. 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
5. Miller SA, Dykes DD, Polesky HF (1988.) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research 16(3):1215
6. Barber MD, McKeown BJ, Parkin BH (1996) Structural variation in the alleles of a short tandem repeat system at the human alpha fibrinogen locus. Int J Legal Med 108:180–185
7. Möller A, Meyer E, Brinkmann B (1994) Different types of structural variations in STRs: HumFES/FPS, HumVWA and HumD21S11. Int J Legal Med 106:319–323
8. Glock B, Dauber EM, Schwartz DWM, Mayr WR (1997) Additional variability at the D12S391 locus in an Austrian population sample: sequencing data and allele distribution. Forensic Sci Int 90:197–203
9. Schwartz DWM, Jungl EM, Krenek OR, Mayr WR (1994) Simple and rapid typing for VNTR polymorphisms using high resolution electrophoresis of PCR products on rehydratable polyacrylamide gels. In: Bär W, Fiori A, Rossi U (eds) Advances in forensic haemogenetics 5. Springer, Berlin Heidelberg New York, pp 170–172
10. Glock B, Schwartz DWM, Jungl EM, Mayr WR (1996) Allelic ladder characterization of the short tandem repeat polymorphism in intron 6 of the lipoprotein lipase gene and its application in an Austrian Caucasian population study. J Forensic Sci 41:579–581
11. Glock B, Schwartz DWM, Schwartz-Jungl EM, Thorwartl GH, Dauber EM, Mayr WR (1997) A new allele at the short tandem repeat locus HumF13A01. Int J Legal Med 110:284–285
12. Urquhart A, Oldroyd NJ, Kimpton CP, 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
13. Barber MD, Piercy RC, Andersen JF, Parkin BH (1995) Structural variation of novel alleles at the HumVWA and HumFES/FPS short tandem repeat loci. Int J Legal Med 108:31–35
14. Alper B, Meyer E, Schürenkamp M, Brinkmann B (1995) HumFES/FPS and HumF13B: Turkish and German population data. Int J Legal Med 108:93–95
15. Piccinini A, Möller K, Wiegand P (1996) HumFES/FPS and HumF13B: population genetic data from North Italy. Int J Legal Med 108:283–284
16. 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 the ISFH regarding the use of short tandem repeat systems. Int J Legal Med 110: 175–176
17. Brinkmann B, Sajantila A, Goedde HW, Matsumoto H, Nishi K, Wiegand P (1996) Population genetic comparisons among eight populations using allele frequency and sequence data from three microsatellite loci. Eur J Hum Genet 4:175–182
18. Schwartz DWM, Dauber EM, Glock B, Mayr WR (1996) Typing of the HumVWA microsatellite polymorphism: allele frequencies and sequencing data. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics 6. Springer, Berlin Heidelberg New York, pp 58–59
19. Horst B, Eigel A, Sanguansermsri T, Rolf B (1997) Analysis of the short tandem repeat systems HumVWA and HumF13B in a population sample from northern Thailand. Int J Legal Med 110:235–237
20. Walsh PS, Fildes NJ, Reynolds R (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Res 24(14):2807–2812
21. Karger B, Meyer E, DuChesne A (1997) STR analysis on perforating FMJ bullets and a new variant VWA allele. Int J Legal Med 110:101–103
22. Buscemi L, Cucurachi N, Mencarelli R, Tagliabracci, Wiegand P, Ferrara SD (1995) PCR analysis of the short tandem repeat (STR) system HUMVWA31 (Allele and genotype frequencies in an Italian population sample). Int J Legal Med 107:171–173
23. Klintschar M, Kubat M (1995) A study of the short tandem repeat systems HumVWA and HumTH01 in an Austrian population sample. Int J Legal Med 107:329–330
24. 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
25. 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
26. Bodmer JG, Marsh SGE, Albert ED, Bodmer WF, Bontrop RE, Charron D, Dupont B, Erlich HA, Fauchet R, Mach B, Mayr WR, Parham P, Sasazuki T, Schreuder GMTh, Strominger JL, Svejgaars A, Terasaki PI (1997) Nomenclature for factors of the HLA system 1996. Vox Sang (73):105–130