 © Springer-Verlag 1996

ORIGINAL ARTICLE

F. Rousselet · H. Pfitzinger · P. Mangin

French Caucasian population data obtained from fluorescently detected HumvWFA31/A and HumF13A01 short tandem repeat loci

Received: 9 October 1995 / Received in revised form: 15 January 1996

Abstract Allele and phenotype frequencies for two tetranucleotide STR (short tandem repeat or microsatellite) systems, HUMvWFA31/A and HUMF13A01, were obtained from a sample of approximately 240 unrelated individuals randomly selected from the French Caucasian population. PCR (polymerase chain reaction) products were analysed on 6% polyacrylamide denaturing gels and visualized using fluorescently labelled primers on the automated 373A ABI DNA sequencer (Applied Biosystems Inc.). French Caucasian allele frequencies were compared to other published Caucasian data. Conditions were optimised for the quadruplex PCR amplification of these two STR loci together with the HUMFESFPS and HUMTH01 loci and the quadruplex PCR was also performed on various forensic DNA samples.

Key words Population data · STR · Multiplex PCR · Fluorescence

Introduction

STR loci, consisting of short repeated DNA motives (2–6 bp), provide small PCR products (100–500 bp) corresponding to discrete alleles, which are easily identified and differentiated by the number of their repeat units. STR loci show a greater sensitivity than variable number tandem repeat (VNTR) systems (Kimpton et al. 1993) as they are more tolerant of degraded, small or mixed DNA samples. Because of their small allele size ranges, the microsatellite loci can also be amplified simultaneously in a single tube as a multiplex PCR reaction (Lygo et al. 1994) thus improving the efficiency and the speed of the analysis and consequently reducing costs.

In order to apply these STR systems to forensic identifications or parental testing (Whitaker et al. 1995) in routine analyses, it was essential to determine the allele frequencies in the population of interest.

In a previous study using a manual method we obtained French Caucasian population data for HUMFESFPS (FES) and HUMTH01 (TH01) STR loci (Pfitzinger et al. 1995). The use of an automated ABI 373A sequencer allowed us to obtain rapidly, through fluorescent duplex amplification, the HUMvWFA31/A (vWF) and the HUMF13A01 (F13A) allele distributions. Finally, the fluorescent detection of the quadruplex amplified alleles (Robertson et al. 1991) at these four STR loci was developed and optimized for identification purposes.

Materials and methods

STR primer sequences

The HUMTH01 (Polymeropoulos et al. 1991a), HUMFESFPS (Polymeropoulos et al. 1991b), HUMvWFA31/A (Kimpton et al. 1992) and HUMF13A01 (Polymeropoulos et al. 1991c) primer sequences were as described in the literature (Kimpton et al. 1993).

Multiplex PCR mixtures

Coamplification of the data base samples for vWF and F13A loci: the DNA extracted from approximately 240 blood samples of unrelated randomly selected individuals was amplified in a 25 μl reaction mixture containing 20 ng of DNA, 1 × buffer (Gibco-BRL), 1.5 mM $MgCl_2$, 0.05% Tween, 200 μg/ml BSA (Gibco-BRL), 200 μM each dNTP, 0.625 U Taq DNA-polymerase (Gibco-BRL), 0.06 μM each vWF primer and 0.15 μM each F13A primer. For quadruplex amplification, the same conditions were used except for primer concentrations which were 0.2 μM each vWF and TH01 primers, 0.25 μM each F13A primer and 0.5 μM each FES primer.

PCR cycles

Duplex and quadruplex amplifications were carried out in a Perkin Elmer Gene Amp PCR system 9600, for 28 cycles, as follows: 45 s at 94°C, 30 s at 54°C and 30 s at 72°C.

Polyacrylamide gel electrophoresis

Of the PCR product 2 μl were loaded on a 6% denaturing polyacrylamide gel as described in the ABI GENESCAN 672 software

F. Rousselet · H. Pfitzinger (✉) · P. Mangin
Institut de médecine légale, 11 rue Humann,
F-67085 Strasbourg Cedex, France

user's manual and the fragment sizes determined automatically using the same software.

Detection

In order to distinguish the overlapping alleles at the four STR loci analysed in the quadruplex PCR mixture, one of the two primers for each locus was labelled with the JOE or the FAM dyes (ABI) as described (Kimpton et al. 1993). The final allele designation was based on the conversion of base pairs into repeat units present in the amplified allelic fragment as proposed (DNA recommendations 1994).

Allelic ladders

The designation was achieved against four allelic ladders corresponding to each STR system and run simultaneously on the same gel. TH01, FES, vWF and F13A allelic ladders consisting of sequenced alleles were provided by the Forensic Science Service laboratory (Aldermaston, UK) as part of an EDNAP (European DNA Profiling group) inter-laboratory exercise (Kimpton et al. 1995).

Statistical evaluations

Evaluation of the Hardy-Weinberg (H-W) equilibrium was based on comparisons of observed and expected phenotypes, using the χ^2 test (Schwartz 1963). For these calculations, phenotype classes with less than five expected events were pooled and the degrees of freedom (*df*) were calculated according to the formula (c-1-a) where c is the number of compared phenotype classes and a is the number of alleles appearing in the compared classes. Furthermore, the allele frequencies occurring in different Caucasian population samples were compared according to the alleles observed for each population sample and using a computer programme from G. Carmody (Carleton University, Canada) calculating G-statistics values (i.e. likelihood-ratio test). The programme calculates the probability by generating 1000 random tables (Roff and Bentzen 1989), each having the same marginal totals as the observed tables. It also calculates the χ^2 and the G-statistics for the observed table and for all of the 1000 simulated tables, and gives the probability of obtaining random tables that have larger χ^2 and G-statistics values than the observed table (Lewontin and Felsenstein 1965). This test avoids combining data with less than five observations and tolerates tables having cells with no observations. If the probability is less than 0.05, the data are not homogeneous enough to be grouped together. On the contrary, probabilities greater than 0.05 indicate that the data can legitimately be amalgamated. When the comparison using the χ^2 and G-statistic tests showed heterogeneity, we performed the test again, but for combinations of two populations. An additional test concerns the comparison of each allele class separately for all the population samples of interest, in order to localize the observations leading to heterogeneity.

Results and discussion

HUMvWFA31/A

The nine different vWF alleles (13–21) observed were unimodally distributed in the French population sample ($n = 242$) as already observed for other populations (Table 1). The amplified allele sizes ranged from 135 to 167 base pairs (*bp*). The smallest allele observed migrated as allele 13, containing 13 repeat units, whereas allele 11 observed in the Italian population sample from Parma (Buscemi et al. 1995), was not detected in our French sample (Table 1). The largest allele observed migrated as allele 21. The three most frequent alleles were alleles 16, 17 and 18 as already detected in the other population samples (Table 1). Observed heterozygosity (0.81) and allelic diversity (h = 0.81) were identical and the discrimination power of the vWF system in the French population was 0.94. According to the χ^2 test performed with the observed (Table 2) and expected phenotypes, no significant deviation from Hardy-Weinberg equilibrium was found for our French Caucasian sample ($\chi^2 = 9.12$; $df = 10$; $0.50 < P < 0.90$). On another hand, the homogeneity G-statistics test performed led to homogeneity for the 10 allelic distributions analysed and shown in Table 1 ($\chi^2 = 102.17$; $P = 0.06$; G = 95.58; $P = 0.07$).

Table 1 HUMvWFA31/A allelic distributions for ten Caucasian population samples. The allele size is indicated in base pairs (bp). The allele identity (ID) corresponds to the number of repeat units present in the amplified fragment and deduced from its size. *n* represents the total number of alleles observed in each population sample. (a) Möller et al. 1994, (b) Sajantila et al. 1994, (c) Drozd et al. 1994, (d) Hochmeister et al. 1994, (e) Dupuy et al. 1993, (f) Lorente et al. 1994, (g) Buscemi et al. 1995 (Italy 1 = Ancona; Italy 2 = Parma), (h) Furedi et al. 1995.

Allele size (bp)	Allele ID	Allele frequencies									
		France $n = 484$	Germany $n = 642$ (a)	Finland $n = 350$ (b)	UK $n = 400$ (c)	Switzerland $n = 200$ (d)	Norway $n = 600$ (e)	Spain $n = 240$ (f)	Italy 1 $n = 228$ (g)	Italy 2 $n = 194$ (g)	Hungary $n = 488$ (h)
127	11	0	0	0	0	0	0	0	0	0,005	0
131	12	0	0	0	0	0	0	0	0	0	0
135	13	0,002	0,0047	0,003	0,0025	0	0	0	0	0	0
139	14	0,103	0,11	0,12	0,13	0,1	0,088	0,163	0,096	0,077	0,115
143	15	0,097	0,098	0,054	0,1025	0,095	0,087	0,129	0,074	0,123	0,125
147	16	0,231	0,21	0,191	0,1775	0,225	0,22	0,208	0,197	0,175	0,195
151	17	0,267	0,27	0,3	0,2675	0,29	0,27	0,242	0,28	0,314	0,277
155	18	0,182	0,21	0,166	0,2	0,215	0,218	0,163	0,223	0,221	0,186
159	19	0,101	0,77	0,14	0,1	0,06	0,088	0,096	0,096	0,072	0,086
163	20	0,015	0,17	0,023	0,02	0,01	0,027	0	0,03	0,01	0,014
167	21	0,002	0,016	0,003	0	0,005	0,002	0	0	0	0,002

Table 2 HUMvWA31/A observed phenotypes for the French Caucasian population sample. *n* is the total number of individuals in the sample and n corresponds to the number of observations for each phenotype. Non-observed phenotypes are not listed.

Phenotypes ID	n ($n = 242$)
13–17	1
14–14	3
14–15	4
14–16	11
14–17	12
14–18	11
14–19	6
15–15	2
15–16	18
15–17	10
15–18	8
15–19	3
16–16	13
16–17	30
16–18	15
16–19	10
16–20	2
17–17	18
17–18	23
17–19	14
17–20	3
18–18	8
18–19	12
18–20	2
18–21	1
19–19	2

HUMF13A01

The HUMF13A01 locus possesses a larger number of identified alleles compared to the vWF locus and to the TH01 and FES loci studied previously (Pfitzinger et al. 1995). Fifteen distinct F13A alleles ranging from 181 to 235 bp were identified in this French population sample ($n = 234$) (Table 3). The smallest allele observed migrated as allele 3.2 and the largest allele observed migrated as allele 17. Alleles 2.2, 9 and 10 did not appear in our population sample. The five smallest alleles (3.2–7) were also the five most frequent alleles as observed for the American, Swiss, German and Norwegian population samples (Table 3). The largest F13A alleles are rare and present at less than 2%. Allele 2.2, never reported before, was observed in the Norwegian Caucasian population sample (Table 3). Alleles 9 and 10, absent from four Caucasian population samples (French, American, German and Norwegian), were reported for Afro-Carribbean (Kimpton et al. 1993) and black-American (Gill and Evett 1995) individuals whereas allele 9 was observed once in the Swiss Caucasian sample. The observed heterozygosity (0.76) is very similar to allelic diversity ($h = 0.77$) and the discrimination power of the F13A system corresponded to 0.90. Observed (Table 4) and expected F13A French phenotypes did not deviate significantly from Hardy-Weinberg equilibrium ($\chi^2 = 8.85$; df = 7; $0.20 < P < 0.30$). The G-statistics test shows that the French, American, Swiss, German and Norwegian samples are not homogeneous ($\chi^2 = 77.61$; $P = 0.03$; G = 80.03; $P = 0.02$). In fact, according to the values of the allele frequencies in the different population samples reported in Table 3, this heterogeneity may be due either to the under-representation of allele 3.2 or to the over-representation of allele 16 in the Swiss Caucasian population, which is also the smallest sample studied for the F13A locus. According to the G-test for each pair of population samples, it appeared that the Swiss allelic distribution is significantly different from the French, American and Norwegian ones but not from the German one (Table 5). Furthermore, the same G-test applied to each F13A allele class separately in the five

Table 3 HUMF13A01 allelic distributions for five Caucasian population samples. The allele size is indicated in base pairs (bp). The allele identity (ID) corresponds to the number of repeat units present in the amplified fragment and deduced from its size. *n* represents the total number of alleles observed in each population sample. (a) Hammond et al. 1994, (b) Hochmeister et al. 1994, (c) Puers et al. 1993, (d) Dupuy et al. 1993.

Allele size (bp)	Allele ID	Allele Frequencies				
		France $n = 468$	USA $n = 348$ (a)	Switzerland $n = 198$ (b)	Germany $n = 326$ (c)	Norway $n = 600$ (d)
177	2.2	0	0	0	0	0,002
181	3.2	0,098	0,083	0,02	0,055	0,077
183	4	0,051	0,02	0,045	0,018	0,042
187	5	0,229	0,192	0,202	0,187	0,184
191	6	0,271	0,345	0,323	0,325	0,297
195	7	0,306	0,325	0,354	0,359	0,354
199	8	0,009	0,006	0	0,012	0,01
203	9	0	0	0,005	0	0
207	10	0	0	0	0	0
211	11	0,002	0,003	0,015	0,003	0
215	12	0,002	0	0	0,003	0,003
219	13	0,005	0	0,005	0,009	0,003
223	14	0,009	0,009	0	0,006	0,008
227	15	0,01	0,017	0,01	0,012	0,017
231	16	0,006	0	0,02	0,009	0,003
235	17	0,002	0	0	0	0,002

Table 4 HUMF13A01 observed phenotypes for the French Caucasian population sample. *n* is the total number of individuals in the sample and n corresponds to the number of observations for each phenotype. Non-observed phenotypes are not listed.

Phenotypes ID	n (*n* = 234)
3.2–3.2	4
3.2–4	2
3.2–5	11
3.2–6	11
3.2–7	12
3.2–16	1
3.2–17	1
4–4	1
4–5	3
4–6	2
4–7	12
4–12	1
4–15	1
4–16	1
5–5	11
5–6	30
5–7	36
5–8	1
5–14	2
5–15	2
6–6	20
6–7	40
6–11	1
6–14	2
6–15	1
7–7	20
7–13	2
7–15	1
8–8	1
8–16	1

population samples studied led to homogeneity for all allele classes except for allele class 3.2 (G = 17.5210; p = 0.0020). This result shows that the difference between the samples can be due to the sensitive variation in the occurrence of allele 3.2 in the different samples (Table 3).

Quadruplex PCR at HUMTH01, HUMFESFPS, HUMvWFA31/A and HUMF13A01 loci

The effects of different DNA concentrations or primer concentrations and of cycling parameters were investigated and led to the final conditions described in material and methods for the amplification of a minimum of 20 ng of DNA. The efficiency of the quadruplex amplification was verified on DNA extracted from blood samples and on various forensic DNA samples. The PCR products obtained from the DNA of a single individual purified from liver, spleen, kidney, lymph nodes, psoa muscles and heart and analysed on the same gel gave identical alleles, thus showing the reliability of the system (data not shown). This quadruplex STR system has been very helpful in numerous routine forensic analyses where the low quantity or bad quality of the DNA gave negative results using the RFLP technique. However, it appeared that with some contaminated and small DNA samples, singleplex PCR can be a solution to the multiplex inhibition, and to type the STR systems that did not amplify in the quadruplex reaction mixture. Furthermore, the use of the Genescan-2500 Rox internal ladder (ABI), containing large DNA fragments, prevents two series of samples from being run successively on the same gel. We therefore decided to use a regular DNA ladder with 25 bp gaps and consisting of smaller DNA fragments up to 500 bp long (Hammond et al. 1994).

Conclusion

Our previous study of the HUMTH01 and HUMFESFPS STR loci using manual silver staining detection (Pfitzinger et al. 1995) was extended in the present work to automated multiplex fluorescence-based detection of two additional STR loci: HUMvWFA31/A and HUMF13A01. Allele and phenotype frequencies for vWF and F13A loci were easily and rapidly determined by duplex amplification of about 240 DNA samples and were in

Table 5 G-statistics values for each possible combination of two population samples out of the five population samples studied for the HUMF13A01 locus. S̲ indicates that the difference between the two allelic distributions is statistically significant and ns indicates that the difference between the two allelic distributions is statistically not significant.

F13A Caucasian population data	Norway	Germany	Switzerland	USA
France	g = 11.2975 *P* = 0.7610 ns	g = 18.3428 *P* = 0.1550 ns	g = 32.3467 *P* = 0.0010 S̲	g = 19.6501 *P* = 0.1320 ns
USA	g = 15.4684 *P* = 0.4350 ns	g = 14.1633 *P* = 0.3680 ns	g = 32.6704 *P* = 0.0030 S̲	
Switzerland	g = 35.1072 *P* = 0.0000 S̲	g = 19.4847 *P* = 0.1440 ns		
Germany	g = 12.6931 *P* = 0.6450 ns			

agreement with Hardy-Weinberg equilibrium expectations. The amplification of STR loci, either singleplex or quadruplex, has been used for routine analysis, in our laboratory, for almost one year and was very helpful in many cases where RFLP analyses gave negative results. In our laboratory, the current development of additional multiplex PCR systems for other loci, such as D6S502, D18S51, D21S11, HUMFIBRA, or amelogenin for sex diagnosis (Urquhart et al. 1995), will lead to a higher discrimination equivalent to that obtained with RFLP techniques and single locus probes.

Acknowledgements We wish to thank Estelle Carniel for helpful technical assistance, Michel Ross for statistical evaluations and Peter Gill from the Forensic Science Service for his great help. This work was supported by a grant from the French ANRT (Association Nationale de la Recherche Technique).

References

Buscemi L, Cucurachi N, Mencarelli R, Tagliabracci A, 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

DNA recommendations-1994 report concerning further recommendations of the DNA commission of the International Society for Forensic Haemogenetics regarding PCR-based polymorphisms in STR (Short Tandem Repeat) systems. Int J Legal Med 107: 159–160

Drozd MA, Archard L, Lincoln PJ, Morling N, Nellemann LJ, Phillips C, Soteriou B, Syndercombe Court D (1994) An investigation of the HUMVWA31A locus in British Caucasians. Forensic Sci Int 69: 161–170

Dupuy BM, Berg ES, Olaisen B (1993) Four STRs in 300 Norwegians. In: Bär W, Fiori A, Rossi U (eds) Advances in forensic haemogenetics 5. Springer, Berlin Heidelberg New York, pp 539–541

Furedi S, Woller J, Padar Z (1995) Hungarian population data for the STR systems TH01 and VWA. Int J Legal Med 108: 48–49

Gill P, Evett I (1995) Population genetics of short tandem repeat (STR) loci. Genetica 96: 69–87

Hammond HA, Lin J, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of thirteen short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55: 175–189

Hochmeister MN, Jung JM, Budowle B, Borer UV, Dirnhofer R (1994) Swiss population data on three tetrameric short tandem repeat loci-VWA, HUMTH01, and F13A1-derived using multiplex PCR and laser fluorescence detection. Int J Legal Med 107: 34–36

Kimpton CP, Walton A, Gill P (1992) A further tetranucleotide repeat polymorphism in 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 Appl 3: 13–22

Kimpton C, Gill P, D'Aloja E, Anderson JF, Bär W, Holgersson S, Jacobson S, Johnsson V, Kloosterman AD, Lareu MV, Nellemann L, Pfitzinger H, Phillips, Rand S, Schmitter H, Schneider PM, Sternersen M, Wide MC (1995) Report on the second EDNAP collaborative exercise. Forensic Sci Int 71: 137–152

Lewontin RC, Felsenstein J (1965) The robustness of homogeneity tests in 2XN tables. Biometrics 21: 19–33

Lorente JA, Lorente M, Budowle B, Wilson MR, Villanueva E (1994) Analysis of short tandem repeat (STR) HUMVWA in the Spanish population. Forensic Sci Int 65: 169–175

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, Wiegand P, Grüschow C, Seuchter SA, Baur MP, Brinkmann B (1994) Population data and forensic efficiency values for the STR systems HUMVWA, HUMMBP, and HUMFABP. Int J Legal Med 106: 183–189

Pfitzinger H, Ludes B, Kintz P, Tracqui A, Mangin P (1995) French Caucasian population data for the HUMTH01 and HUMFESFPS short tandem repeat (STR) systems. J Forensic Sci 40: 270–274

Polymeropoulos MH, Xiao H, Rath DS, Merril CR (1991a) Tetranucleotide repeat polymorphism at the Human Tyrosine Hydroxylase Gene (TH). Nucleic Acids Res 19: 3753

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

Polymeropoulos MH, Rath SD, Xiao H, Merril CR (1991c) Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene (F13A1). Nucleic Acids Res 19: 4036

Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW (1993) Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01 (AATG)n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 53: 953–958

Robertson J, Ziegle J, Kronick M, Madden D, Budowle B (1991) Genetic typing using automated electrophoresis and fluorescence detection. In: Burke T, Dolf G, Jeffreys AJ, Wolff R (eds) DNA fingerprinting: approaches and applications. Birkhauser, Basel, pp 391–398

Roff DA, Bentzen P (1989) The statistical analysis of mitochondrial DNA polymorphisms and the problem of small samples. Mol Biol Evol 6: 539–545

Sajantila A, Pacek P, Lukka M, Syvänen AC, Nokelainen P, Sistonen P, Peltonen L, Budowle B (1994) A microsatellite polymorphism in the von Willebrand factor gene: comparison of allele frequencies in different population samples and evaluation for forensic medicine. Forensic Sci Int 68: 91–102

Schwartz D (1963) Comparaison d'un pourcentage observé à un pourcentage théorique. Principe des tests (in French). In: Méthodes statistiques à l'usage des médecins et des biologistes. Collection statistique en biologie et en médecine, Flammarion Medecine Sciences 3ème édition, pp 37–44

Urquhart A, Oldroyd NJ, Kimpton CP, Gill P (1995) Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. Biotechniques 18(1): 116–121

Whitaker JP, Clayton TM, Urquhart AJ, Millican ES, Downes TJ, Kimpton CP, Gill P (1995) Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. Biotechniques 18: 670–677