Int J Legal Med (1995) 108 : 145–149

© Springer-Verlag 1995

ORIGINAL ARTICLE

Pablo Martin · Antonio Alonso · Bruce Budowle
Cristina Albarrán · Oscar García · Manuel Sancho

Spanish population data on 7 tetrameric short tandem repeat loci

Received: 19 May 1995 / Received in revised form: 3 August 1995

Abstract Allele and genotype frequencies for 7 tetrameric short tandem repeat loci were determined in a Spanish population sample (N = 186–244) using PCR and subsequent analysis of the PCR products by denaturing polyacrylamide gel electrophoresis followed by silver staining. The loci were HUMFES/FPS, HUMVWA, HUMTHO1, HUMF13B, HUMCSF1PO, HUMF13A1 and HUMTPOX and all loci met Hardy-Weinberg expectations. In addition, little evidence was found for association of alleles among the 7 loci. Thus the allele frequency data can be used in identity testing to estimate the frequency of a multiple PCR-based DNA profile in the Spanish population.

Key words PCR · STR · HUMFES/FPS · HUMVWA · HUMTHO1 · HUMF13B · HUMCSF1PO · HUMF13A1 · HUMTPOX · Population genetics · Spain · Hardy-Weinberg equilibrium · Linkage equilibrium

P. Martin (✉) · A. Alonso · C. Albarrán · M. Sancho
Instituto Nacional de Toxicología, Sección de Biología, Luis Cabrera, 9, E-28002-Madrid, Spain

B. Budowle
Forensic Science Research and Training Center, FBI Academy, Quantico, Virginia 22135, USA

O. García
Area de laboratorio UTAP, Mª Díaz de Haro, 3, E-48013 Bilbao, Spain

Introduction

Short tandem repeat (STR) loci are highly polymorphic markers composed of repetitive sequence elements of 2–7 base pairs in length [1, 2] which are amenable to analysis by the polymerase chain reaction [3]. Because of their small size (< 400 bp), STRs can be efficiently amplified even when the DNA template is severely degraded and generally do not have the problem of unequal amplification among alleles [4–6]. In addition, STR amplification products differing in size by as little as one base can be clearly resolved by denaturing polyacrylamide gel electrophoresis to give precise and discrete allelic data. These advantages make the PCR-based STR typing one of the methods of choice for forensic identification of body fluids and human remains and for paternity testing [7].

This study presents allele/genotype frequency data in a Spanish population sample for 7 STR loci: HUMFES/FPS

Table 1 Characteristics of the STR loci analysed

STR	Gene chromosome location	Primer sequence references	Product length (bp)	Alleles in the ladder
HUMFES/FPS	c-fes/fps proto-oncogene (15q25-qter)	[7, 8]	222–250	7, 8, 9, 10, 11, 12, 13, 14
HUMVWA	Von Willebrand factor (12-p12-pter)	[9, 13]	139–167	13, 14, 15, 16, 17, 18, 19, 20
HUMTHO1	Tyrosine hydroxylase (11p15.5)	[1]	179–203	5, 6, 7, 8, 9, 10, 11
HUMF13B	Factor XIII b (1q31–q32.1)	[10]	169–185	6, 7, 8, 9, 10
HUMCSF1PO	c-fms proto-oncogene for CSF-1 receptor (5q33.5–q34)	[7, 14]	295–327	7, 8, 9, 10, 11, 12, 13, 14, 15
HUMF13A1	Coagulation factor XIII a (6p24–p25)	[7, 11]	281–331	4, 5, 6, 7, 8, 9 11, 12, 13, 14, 15, 16
HUMTPOX	Thyroid peroxidase (2p13)	[12, 14]	232–248	8, 9, 10, 11, 12

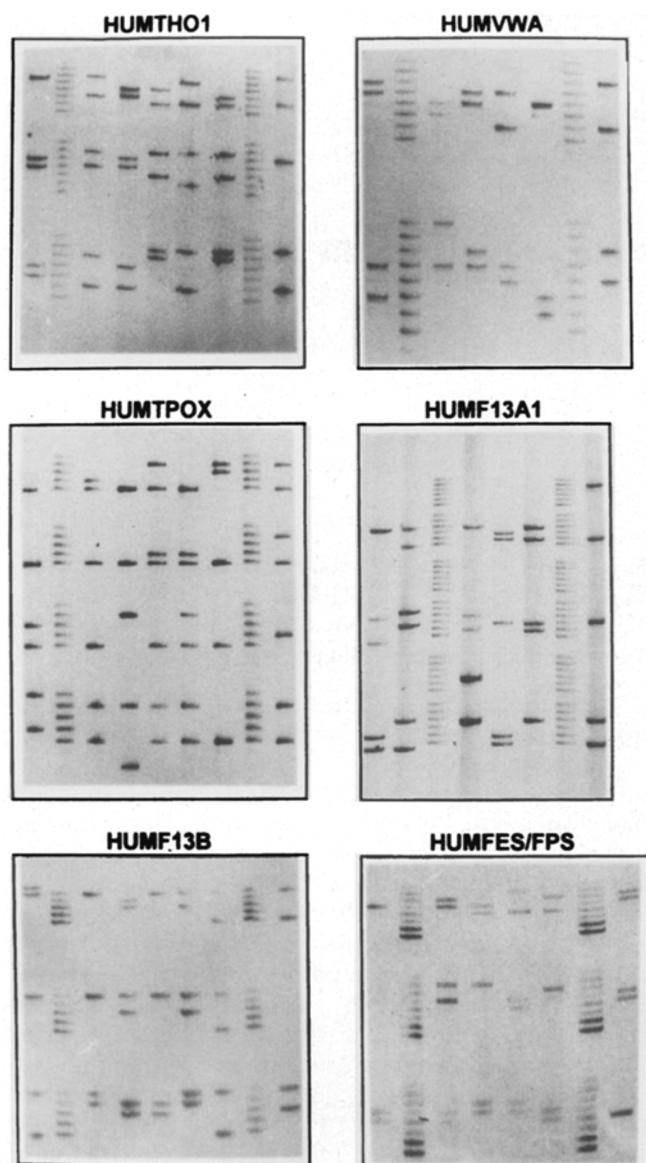

Fig. 1 Representative STR profiles of 6 different loci obtained by single-locus PCR reactions and sequential loading of the PCR products on denaturing polyacrylamide gel and subsequent detection by silver stain. Samples were loaded four (HUMTPOX), three (HUMTHO1, HUMFES/FPS, HUMF13B and HUM13A1) or two (HUMVWA) separated times at 15–25 min intervals

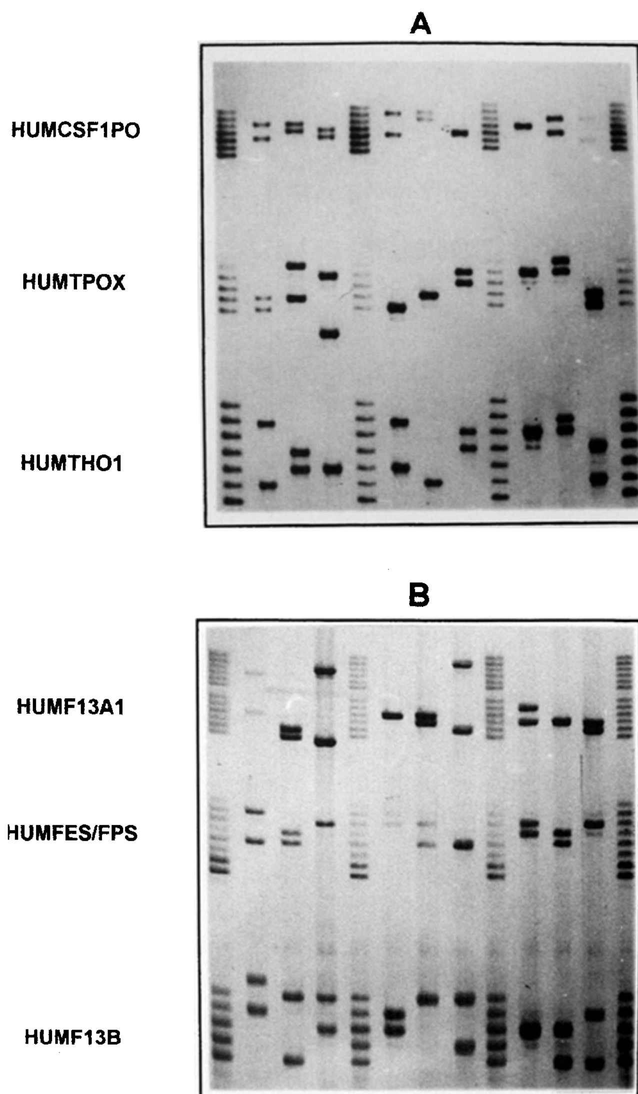

Fig. 2 Analysis of different STR loci by single-locus PCR reactions and simultaneous loading of the PCR products on denaturing polyacrylamide gel and subsequent detection by silver stain. **A** Simultaneous analysis of PCR products amplified separately at the HUMCSF1PO, HUMTPOX and HUMTHO1 loci. **B** Simultaneous analysis of PCR products amplified separately at HUMF13B, HUMFES/FPS and HUMF13A1 loci

[8], HUMVWA [9], HUMTHO1 [1], HUMF13B [10], HUMCSF1PO [7], HUMF13A1 [11] and HUMTPOX [12].

Materials and methods

Population sample

Blood samples were taken from routine forensic cases investigated in this Institute. The population sample examined in this study encompass a total of 244 unrelated individuals living predominantly in the following geographical areas of Spain: communities of Madrid, Castilla-La-Mancha, and Castilla-Leon (central Spain) and community of Valencia (Eastern Spain).

Sample preparation

The DNA was extracted by the standard phenol/chloroform extraction procedure. The quality and the quantity of DNA in each sample were estimated by ethidium bromide agarose gel electrophoresis and fluorimetry, respectively.

PCR typing

The amplification of HUMFES/FPS, HUMVWA, HUMTHO1, HUMF13B, HUMCSF1PO, HUMF13A1 and HUMTPOX (see Table 1 for chromosome location and primer sequence references) were performed separately according to the manufacturer's recommendations using the GenePrint STR System (Promega Corporation, Madison, Wis.). The PCR was carried out in 25 μl reaction volumes containing 5–20 ng template DNA and 0.5 units of AmpliTaq DNA polymerase (Perkin-Elmer) in a 480 Perkin-Elmer

thermal cycler. PCR products (0.5–5 μl) were mixed with 2.5 μl of loading dye (10 m*M* NaOH, 95% formamide, 0.05% bromophenol blue and 0.05% xylenecyanol FF), denatured at 95°C for 3 min and loaded onto denaturing polyacrylamide gels cross-linked with piperazine diacrylamide (6–4% T, 3% C, 24 cm long and 0.75 mm thick) containing 8M urea and 0.5X Tris-Borate-EDTA. Electrophoresis was carried out for 2–3 h on a Hoefer SE 620 vertical slab gel electrophoresis unit (Hoefer Scientific Instruments, Calif.). The conditions for electrophoresis were set at a constant voltage of 500–800 V and carried out at 50°C. Temperature was controlled by means of a LKB 2219 MultiTemp II thermostatic circulator (Pharmacia, Sweden), connected to the heat exchanger of the Hoefer electrophoresis unit. After electrophoresis gels were stained with silver as described previously [15]. Alleles were classified by comparison of the sample fragments with those of the allelic ladders supplied by Promega GenePrint System (see Table 1). Allele designations were made according to recommendations of the DNA Commission of the International Society for Forensic Haemogenetics [16].

Statistical analysis

The frequency of each allele for each STR was calculated from the numbers of each genotype in the sample (i.e. the gene count method). Unbiased estimates of expected heterozygosity were computed as described by Edwards et al. [2]. The expected numbers of distinct homozygous and heterozygous genotypes and standard errors (SE) were calculated according to the method described by Chakraborty et al. [17, 18]. Possible divergence from Hardy-Weinberg expectations (HWE) was determined by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies [17, 19, 20], the likelihood ratio test [2, 18, 21] and the exact test [22]. An interclass correlation criterion [23] for two-locus association was used for detecting disequilibrium between the STR loci. Independence across all loci was determined by examining whether or not an observed variance of the number of heterozygous loci in the population sample is outside its confidence interval under the assumption of independence [24–26].

Results and discussion

Genotype data for 7 tetrameric STR loci were determined by single-locus PCR reactions. The STR profiles displayed in Figs. 1 and 2 were selected from the population study as a fair representation of the clarity with which allele designations were made. Figure 1 illustrates different STR profiles obtained by sequencial loading of several amplifications corresponding to the same loci, while Fig. 2 illustrates different STR profiles obtained by simultaneous loading of several amplifications corresponding to different loci. Both typing approaches minimize labor, materials and analysis time, thereby increasing throughput. All the observed alleles differed in size by one repeat unit (i.e. 4 base pairs) expect for the HUMTHO1 allele 9.3

Table 2 Allele frequencies for STR loci in a Spanish population sample

Allele	FES/FPS (n = 382)	VWA (n = 444)	THO1 (n = 488)	F13B (n = 392)	CSF1PO (n = 374)	F13A1 (n = 398)	TPOX (n = 436)
3,2						0.085	
4						0.043	
5						0.206	
6			0.217	0.115		0.256	0.002
7			0.176	0.010		0.372	
8	0.008		0.166	0.250	0.003	0.013	0.546
9	0.003		0.178	0.189	0.011		0.096
9.3			0.256				
10	0.319		0.006	0.431	0.267		0.060
11	0.382			0.005	0.275		0.268
12	0.246				0.374		0.028
13	0.039				0.067	0.013	
14	0.003	0.106			0.003	0.005	
15		0.144				0.008	
16		0.227					
17		0.286					
18		0.149					
19		0.074					
20		0.014					

n = Number of alleles

Table 3 HWE test for independence on STR loci in a Spanish population sample

	FES/FPS	VWA	THO1	F13B	CSF1PO	F13A1	TPOX
Obs. Homozygosity	30.9%	16.2%	22.5%	28.1%	28.9%	24.1%	36.7%
Exp. Homozygosity[a]	30.8%	19.2%	20.2%	29.6%	29.0%	25.4%	38.2%
Homozygosity Test[b]	0.989	0.266	0.354	0.648	0.967	0.677	0.644
Likelihood Ratio Test[b]	0.757	0.443	0.288	0.695	0.371	0.406	0.713
Exact Test[b]	0.634	0.392	0.296	0.763	0.550	0.282	0.373

[a] Expected homozygosity is an unbiased estimate
[b] These values are probability values

Table 4 Obseverd and expected heterozygous and homozygous classes for STR loci in a Spanish population sample

	FES/FPS	VWA	THO1	F13B	CSF1PO	F13A1	TPOX
Heterozygotes Observed	10	19	12	10	10	20	11
Heterozygotes Expected ± SE[a]	9.6 ± 1.5	18.5 ± 1.2	12.2 ± 1.1	9.9 ± 1.3	10.1 ± 1.5	18.9 ± 2.1	9.9 ± 1.1
Homozygotes Observed	4	6	5	4	3	6	5
Homozygotes Expected ± SE[a]	3.3 ± 0.5	5.6 ± 0.6	5.0 ± 0.1	3.9 ± 0.3	3.6 ± 0.5	4.1 ± 0.7	3.6 ± 0.7

[a] SE = standard error

Table 5 Two locus inter-class correlation test for STR loci in a Spanish population sample

Loci	N	Probability
FES/VWA	190	0.507
FEST/THO1	189	0.838
FES/F13B	189	0.608
FES/CSF1PO	187	0.560
FES/F13A1	191	0.541
FES/TPOX	190	0.591
VWA/THO1	214	0.205
VWA/F13B	194	0.224
VWA/CSF1PO	187	0.970
VWA/F13A1	197	0.557
VWA/TPOX	208	0.115
THO1/F13B	194	0.067
THO1/CSF1PO	185	0.104
THO1/F13A1	197	0.127
THO1/TPOX	214	0.925
F13B/CSF1PO	186	0.015
F13B/F13A1	196	0.959
F13B/TPOX	195	0.112
CSF1PO/F13A1	187	0.369
CSF1PO/TPOX	187	0.809
F13A1/TPOX	198	0.551

N = Number of individuals with data on both loci

that has a single base deletion of a thymidine residue in the 5th of 10 TCAT repeats [13, 27] and for the HUMF13A1 allele 3.2 that has a 2 base-pair deletion outsite the repeat region [13]. In all cases, expect for HUMVWA, the STR profiles were free from artifactual bands. PCR amplification of the HUMVWA locus sometimes generated faint bands of 4 bases less than the authentic alleles (i.e. stutter bands). These artifactual bands are caused by repeat slippage of the Taq DNA polymerase [28]. This phenomenon does not interfere with the interpretation of HUMVWA profiles obtained from samples of known origin but should be taken into account for forensic analysis of mixed body fluids.

Table 2 shows the observed allele frequencies for the 7 STR loci in the Spanish population sample. The results of the different test procedures for testing the correspondence of the genotype frequencies with Hardy-Weinberg expectations are shown in Tables 3 and 4. All loci meet HWE based on the homozygosity test (i.e. the total frequencies of heterozygotes and homozygotes), the likelihood ratio test, and the exact test. To obtain the likelihood ratio test and the exact test data, the data were shuffled 2000 times. In addition, the test based on the number of distinct genotypes observed in the sample population shows that the observed numbers of distinct heterozygote and homozygote genotypes generally are in accordance with their respective HWE predictions (Table 4).

An interclass correlation test analysis demonstrated that there is little evidence for correlation between the alleles at any of the pairs of loci (Table 5). There was 1 example of departure from expectations (HUMF13B/HUMCSF1PO) out of a total of 21 interclass correlation tests, which is approximately 5% of the comparisons. Thus, the degree of departure was no more than expected. Also, with a Bonferroni correction [29] (used for correcting when multiple tests are performed on a population sample), there was no evidence for departures from expectations for pair-wise comparisons of STR loci. To confirm that there is little evidence for departures from expectations of independence among the STR loci when using the product rule to derive a multiple locus frequency estimate, an additional test for association that addressed all 7 STR loci at one time was considered. The test examined whether or not the observed variance (S_k^2) of the number of heterozygous loci in a population sample is outside its confidence interval under the assumption of independence using the procedure described by Brown et al. [24]. While this test criterion results in some loss of information in summarizing multiple locus genotype data, it is more powerful than the classical goodness-of-fit χ^2 test. There was no evidence of association of the 7 loci ($S_k^2 = 1.53$, 95% confidence interval of variance is 1.05–1.57).

In conclusion, a Spanish population database has been established for HUMFES/FPS, HUMVWA, HUMTHO1, HUMF13B, HUMCSF1PO, HUMF13A1 and HUMTPOX and it has been shown that the allele frequency data for the 7 STR loci can be used to estimate the frequency of a multiple PCR-based DNA profile in the Spanish population.

References

1. 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
2. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric repeat loci in four human population groups. Genomics 12:241–253

3. Saiki RK, Scharf S, Faloona T, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of β-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia. Science 230: 1350–1354
4. Hagelberg E, Gray IC, Jeffreys AJ (1991) Identification of the skeletal remains of a murder victim by DNA analysis. Nature 352: 427
5. Gill P, Kimpton CP, Sullivan KM (1992) A rapid polymerase chain reaction method for identifying fixed specimens. Electrophoresis 13: 173–175
6. Jeffreys AJ, Allen MJ, Hagelberg E, Sonnberg A (1992) Identification of the skeletal remains of Joseph Mengele by DNA analysis. Forensic Sci Int 56: 65–76
7. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55: 175–189
8. Polymeropoulos MH, Rath DS, Xiao H, Merril CR (1991) Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES). Nucleic Acids Res 19: 4018
9. Kimpton CP, Walton A, Gill P (1992) A further tetranucleotide repeat polymorphism in the vWF gene. Hum Mol Genet 1: 287
10. Nishimura DY, Murray JC (1992) A tetranucleotide repeat for the F13B locus. Nucleic Acids Res 20: 1167
11. Polymeropoulos MH, Rath DS, Xiao H, Merril CR (1991) Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene (F13A1). Nucleic Acids Res 19: 4306
12. Anker R, Steinbrueck T, Donnis-Keller H (1992) Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. Hum Mol Genet 1: 137
13. Urquhart A, Kimpton CP, Downes TJ, Gill P (1994) Variation in short tandem repeat sequences: a survey of twelve microsatellite loci for use as forensic identification markers. Int J Leg Med 107: 13–20
14. Huang NE, Schumm J, Budowle B (1995) Chinese population data on three tetrameric short tandem repeat loci – HUMTHO1, TPOX, and CSF1PO – derived using multiplex PCR and manual typing. Forensic Sci Int 71: 131–136
15. Budowle B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC (1991) Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am J Hum Genet 48: 137–144
16. DNA recommendations (1994) Report concerning further recommendations of the DNA commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems. Int J Legal Med 107: 159–160
17. Chakraborty R, Smouse PE, Neel JV (1988) Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. Am J Hum Genet 43: 709–725
18. Chakraborty R, Fornge M, Guegue R, Boerwinkle E (1991) Population genetics of hypervariable loci: analysis of PCR-based VNTR polymorphism within a population. In: Burke T, Dolf G, Jeffreys AJ, Wolff R (eds) DNA fingerprinting: approaches and applications. Birkhauser, Berlin, pp 127–143
19. Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. Genetics 76: 379–390
20. Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590
21. Weir BS (1992) Independence of VNTR alleles defined by fixed bins. Genetics 130: 873–887
22. Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48: 361–372.
23. Karlin S, Cameron EC, Williams PT (1981) Sibling and parent-offspring correlation estimation with variable family size. Proc Natl Acad Sci USA 78: 2664–2668
24. Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of Hordeum Spontaneum. Genetics 96: 523–536
25. Chakraborty R (1984) Detection of nonrandom associations of alleles from the distribution of the number of heterozygous loci in a sample. Genetics 108: 719–731
26. Chakraborty R, Zhong Y, Jin L, Budowle B (1994) Non-detectability of restriction fragments and independence of DNA-fragment sizes within and between loci in RFLP typing of DNA. Am J Hum Genet 55: 391–401
27. Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW (1993) Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTHO1 $[AATG]_n$ and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 53: 953–958
28. Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20: 211–215
29. Weir BS (1990) Multiple tests. In: Genetic data analysis. Sinauer Associate, Sunderland, Mass, pp 109–110