Int J Legal Med (1997) 110 : 184–187

ORIGINAL ARTICLE

S. Füredi · M. Angyal · Z. Kozma · J. Sétáló
J. Woller · Z. Pádár

Semi-automatic DNA profiling in a Hungarian Romany population using the STR loci HumVWFA31, HumTH01, HumTPOX, and HumCSF1PO

Received: 30 December 1996 / Received in revised form: 11 March 1997

Abstract A population study of Hungarian Romanies was carried out for the STR loci HumVWFA31, HumTH01, HumTPOX, and HumCSF1PO. After multiplex PCR amplification semi-automatic DNA profiling was performed using an ALF DNA sequencer. At the loci investigated there was little and no evidence for departures from Hardy-Weinberg expectations and linkage equilibrium, respectively. The allele sizing accuracy of the ALF DNA sequencer was increased to a high level (99.97% on average) by applying external and internal markers. Allele frequency distributions of the STR loci, with one exception, were significantly different between the Romany and other Hungarian population databases. On the other hand, however, only small differences in frequencies of individual phenotypes were found.

Key words PCR · Multiplex STR profiling · Automated DNA sequencer · Romany (Gypsy) population · Hungary

Introduction

The high variability of the short tandem repeats (STRs) and their relative simplicity when used as polymorphic markers, makes them useful tools for population genetics and genetic identification purposes [1, 2].

Prior to introduction of STR characterization to forensic casework in Hungary, population surveys have been started in the relevant populations [3–5]. In Hungary the Romany population is the largest of the ethnic minorities representing about 6% of the Hungarian inhabitants. The Romany population, an Orientalo-Indid group of Caucasians, most likely originated from the north-western part of India, and started to settle in great numbers in Hungary in the 15th century [6]. The Romanies formed genetically closed population groups throughout the country.

While Romany population data for numerous conventional blood type loci exist in Hungary [7, 8], there have not been such population data generated for forensically interesting DNA polymorphisms. The use of an ALF DNA sequencer and fluorescent quadruplex amplification allowed us to rapidly obtain the four-locus (HumVWFA31, HumTH01, HumTPOX, and HumCSF1PO [2, 9, 10]) STR profiles from a Romany population of Baranya county in Hungary. The sizing accuracy of the ALF DNA sequencer was also evaluated since relatively little data have been available for this instrument [11] as compared to other automated DNA sequencers. Furthermore, the Romany population database established here was compared with other Hungarian population databases.

S. Füredi (✉) · J. Woller · Z. Pádár
Department of Biology, Institute for Forensic Sciences,
P.O. Box 314/4, H-1903 Budapest, Hungary

M. Angyal · Z. Kozma · J. Sétáló
Department of Forensic Medicine,
Medical University School of Pécs, P.O. Box 99,
H-7643 Pécs, Hungary

Materials and methods

Population sample

The sample consisted of 135 unrelated Hungarian Romany ("Roma") individuals residing in Baranya county (south-western Hungary).

Multiplex amplification conditions and detection system

The coamplification of HumVWFA31, HumTH01, HumTPOX and HumCSF1PO loci was performed using reagents provided in the GenePrint Fluorescent Quadruplex STR System CSF1PO-TPOX-TH01-vWA (Promega, Madison, Wis.) according to the manufacturer's instructions. The PCR products were analysed on an ALF DNA sequencer (Pharmacia) using standard gel format (19 cm well-to-read distance) as previously described [12]. The fragment size of the amplified products was determined with the Promega allelic ladders loaded in lanes 3, 15, 27, and 38. Two fluorescent internal markers were used in each lane to compensate for any mobility shift between the lanes. A 100 bp internal marker was purchased from Pharmacia, and a 271 bp fragment (allele 9) of the HumHPRTB STR locus was obtained after amplification accord-

ing to Szibor et al. [13], where the reverse primer was fluorescently labeled. The measurement error of computer-generated allele sizing was determined for the Promega allelic ladders and individual samples. The accuracy was calculated for each allele using the formula% accuracy = (1 – |estimated size – expected size|/expected size) × 100 [11].

Statistical methods

Possible divergence from Hardy-Weinberg expectations (HWE) was determined by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies [14] and the exact test [15]. The interclass correlation [16] and s_k^2 criteria [17] were used to test for allelic independence between/among STR loci. Population homogeneity was examined using a computerized G-statistic test of 2 × C contingency tables. Interpopulation differences were evaluated by calculating the mean phenotype difference values (MVs) [18] and by a scatter plot comparison [19]. In the MV investigation 100 individuals were randomly selected out of each population sample. In the scatter plot analysis we pooled all individual profiles from the databases to be compared. The likelihood ratio (LR) estimates of the four-locus profiles were computed multiplying the allele frequencies from each population. The respective LRs of every target profile were then plotted on a logarithmic scale (i.e. a scatter plot). Each axis of the plot assigned from which population database a LR log was estimated.

Results and discussion

Allele frequencies for the four STR loci in the Romany population sample are shown in Table 1. All alleles differed in size by one repeat unit (i.e. 4 base pairs) for all loci, except for the HumTH01 allele 9.3. All the loci did not deviate from HWE based on the exact test (Table 1). The excess of homozygous HumVWFA31 types reached the significance level of 0.03. However, no particular homozygous type was predominantly responsible for this excess (Table 2). An explanation for the surplus is that a homozygote type could be a heterozygote on the level of sequence analysis. Fragments of the same length can have different sequence structures, especially in HumVWFA31, which consists of different repeat motifs [1, 20]. Additionally, inbreeding is a relatively likely assumption for observing excess homozygosity in such genetically closed population groups as Hungarian Romanies.

An interclass correlation analysis and an s_k^2 criterion demonstrated that there was no evidence for allelic association between/among the STR loci ($P \geq 0.064$; s_k^2 = 0.967, 95% C.I. of variance is 0.628–0.986).

A measure of the precision of the computerized allele sizing was examined by analysing the Promega allelic ladders on five different gels. The accuracy values were

Table 1 Allele frequencies and HWE test for four STR loci in 135 unrelated Hungarian Romany individuals

Allele	VWA	TH01	TPOX	CSF1PO
6		0.237		
7		0.085		0.004
8		0.233	0.433	0.019
9		0.181	0.067	0.011
9.3		0.259		
10		0.004	0.144	0.289
11			0.352	0.252
12			0.004	0.381
13				0.041
14	0.085			0.004
15	0.070			
16	0.333			
17	0.252			
18	0.233			
19	0.019			
20	0.007			
Exact test[a]	0.296	0.856	0.083	0.070

[a] These values are probability values

Table 2 Observed and expected homozygous VWA types in 135 Hungarian Romany individuals

Homozygous type	Observed	Expected
14–14	2	0.98
15–15	2	0.67
16–16	15	15.00
17–17	13	8.56
18–18	11	7.35
19–19	0	0.05
20–20	0	0.01
Homozygosity	0.319	0.239

Homozygosity test: $P = 0.030$

Table 3 Standard deviation (SD) of estimated size and maximum size range for the STR alleles obtained by semi-automatic DNA typing of 135 Hungarian Romany individuals[a]

Locus	Allele	SD (bp)	Range (bp)	N
VWA	14	0.07	0.3	21
	15	0.05	0.2	17
	16	0.06	0.3	75
	17	0.07	0.3	55
	18	0.07	0.3	52
	19	0.04	0.1	5
TH01	6	0.06	0.2	55
	7	0.06	0.2	23
	8	0.05	0.2	53
	9	0.06	0.2	46
	9.3	0.06	0.3	61
TPOX	8	0.08	0.4	96
	9	0.08	0.2	16
	10	0.05	0.2	39
	11	0.06	0.2	77
CSF1PO	8	0.08	0.2	5
	9	0.10	0.2	3
	10	0.10	0.4	68
	11	0.09	0.4	65
	12	0.11	0.6	79
	13	0.10	0.3	11

[a] Data include at least 3 observations
N = number of PCR fragments

Table 4 G statistic test (*P* values) for homogeneity on VWA, TH01, TPOX, and CSF1PO allele distributions

Populations	VWA	TH01	TPOX	CSF1PO
Romany/CBu	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.024
Romany/CBa	0.002	0.002	$< 10^{-3}$	0.759
CBu/CBa	0.184	0.678	0.458	0.668

CBu = Hungarian Caucasians residing in the Budapest area [5]
CBa = Hungarian Caucasians residing in Baranya county [21, 22]

Table 5 Phenotype frequency differences (mean values) of the STR loci between populations. Populations were compared from two directions. The abbreviation CBu as given in Table 4 A) Single STR systems. B) 2–4 systems combined

A

Populations	VWA	TH01	TPOX	CSF1PO
Romany → CBu	1.5	1.4	1.5	1.4
CBu → Romany	1.5	1.4	1.8	1.2

B

Population	2 loci (VWA, TH01)	3 loci (TPOX)	4 loci (CSF1PO)
Romany → CBu	2.1	3.2	4.6
CBu → Romany	2.1	2.8	3.3

very similar to those data obtained from the analysis of 928 allelic fragments from 135 Romany individuals (Table 3). This extremely high accuracy (99.97% on average for individual samples) allows a clear distinction between the adjacent alleles differing in size by even 1 bp.

The Hungarian Romany allele frequency values for the four STR loci, with the exception of HumCSF1PO, were significantly different from two other Hungarian population databases (Table 4). These two other databases can be considered homogeneous on all the four loci and represent the Hungarian Caucasians residing in the central [5] and the south-western parts [21, 22] of Hungary. The differences in the allele frequency distributions between the Romany and other Hungarian population databases corresponded with previous observations on conventional blood type markers [7, 8]. The combined forensic efficiency values observed in the Romany population sample [$pM(4\ loci) = 2.3 \times 10^{-4}$; $MEC(4\ loci) = 0.93$] were similar to those found in the two other Hungarian databases mentioned above.

Interpopulation comparisons with respect to individual phenotypes were carried out between the Baranya Romany and Central Hungarian population databases. The comparison regarding MVs revealed quite small interpopulation differences, which did not increase considerably when the STR loci were combined (Table 5). Thus the investigated STR systems have low discrimination power for the two populations analysed here. Our results are in good accordance with those previous findings, which suggest that generally only minor phenotype frequency differences exist within major population groups [18, 19]. In the scatter plot comparison (Fig. 1) 5.3% of the four-locus profiles had LR estimates differing by more than one order of magnitude where estimates were more common than 1/1 million (6 in log). This fact indicates that in certain cases there may be an unfair bias in applying of the Baranya Romany database to the four-locus STR profiles of the Central Hungarian population and vice versa.

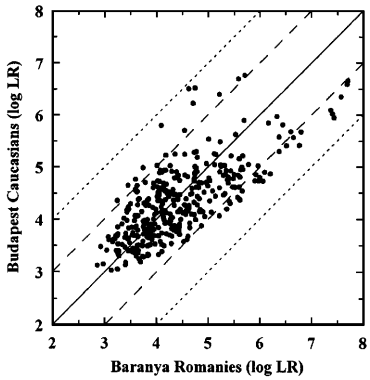

Fig. 1 A scatter plot comparison of the Baranya Romany and Budapest Caucasian population databases for 358 four-locus STR profiles. The solid diagonal line indicates the theoretical line where both databases would produce the same estimates. The dashed and dotted lines assign the one and two order magnitude differences in the profile estimates, respectively

In conclusion, a Hungarian Romany population database has been established for four STR loci using a semi-automatic DNA typing method. Because of its high precision and individualization power, the present STR profiling system is a valuable tool for forensic DNA work in Hungary.

Acknowledgements Computer programs were kindly provided by B. Budowle (FSRTC, FBI Academy, Quantico, USA) and G. Carmody (Carleton University, Ottawa, Canada). The authors wish to thank Magdolna Fucskó for her expert technical assistance.

References

1. 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
2. Sprecher CJ, Puers C, Lins AM, Schumm JW (1996) General approach to analysis of polymorphic short tandem repeat loci. Biotechniques 20:266–276
3. Furedi S, Woller J, Padar Z (1995) Hungarian population data for the STR systems TH01 and VWA. Int J Legal Med 108:48–49
4. Csete K, Schürenkamp M, Varga T (1996) The STR systems HumVWA and HumACTB2 in a Hungarian population. Int J Legal Med 108:316–317
5. Füredi S, Budowle B, Woller J, Pádár Z (1996) Hungarian population data on six STR loci – HUMVWFA31, HUMTH01, HUMCSF1PO, HUMFES/FPS, HUMTPOX, and HUMHPRTB – derived using multiplex PCR amplification and manual typing. Int J Legal Med 109:100–101
6. Tauszik T (1991) The Gypsy groups. In: Czeizel A, Benkmann H-G, Goedde HW (eds) Genetics of the Hungarian population. Springer, Berlin Heidelberg New York, pp 64–72

7. Tauszik T, Hummel K (1991) Blood groups and the results of certain blood group tests in Hungary. In: Czeizel A, Benkmann H-G, Goedde HW (eds) Genetics of the Hungarian population. Springer, Berlin Heidelberg New York, pp 116–123
8. Benkmann H-G, Goedde HW (1991) Genetic markers, serum protein and erythrocyte enzyme polymorphisms. In: Czeizel A, Benkmann H-G, Goedde HW (eds) Genetics of the Hungarian population. Springer, Berlin Heidelberg New York, pp 124–149
9. Hochmeister MN, Budowle B, Schumm JW, Sprecher CJ, Borer UV, Dirnhofer R (1995) Swiss population data and forensic efficiency values on 3 tetrameric short tandem repeat loci-HUMTH01, TPOX, and CSF1PO – derived using a STR multiplex system. Int J Legal Med 107:246–249
10. Nagai A, Yamada S, Watanabe Y, Bunai Y, Ohya I (1996) Analysis of the STR loci HUMF13A01, HUMFXIIIB, HUMLIPOL, HUMTH01, HUMTPOX and HUMVWFA31 in a Japanese population. Int J Legal Med 109:34–36
11. Decorte R, Cassiman J-J (1996) Evaluation of the ALF DNA sequencer for high-speed sizing of short tandem repeat alleles. Electrophoresis 17:1542–1549
12. Moscetti A, Boschi I, Dobosz M, Destro-Bisol G, Pescarmona M, D'Aloja E, Pascali VL (1995) Fluorescence-based classification of microsatellites using a single-wavelength semiautomatic sequencer: genotype assignment and identity tests by analysis of comigrating peak profiles. Electrophoresis 16: 1875–1880
13. Szibor R, Krause D, Kropf S (1995) Usage of X-linked markers for forensic purposes. In: Jacob B, Bonte W (eds) Advances in forensic sciences 6. Verlag Dr Köstler, Berlin, pp 141–145
14. Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. Genetics 76:379–390
15. Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372
16. 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
17. Chakraborty R (1984) Detection of nonrandom association of alleles from the distribution of the number of heterozygous loci in a sample. Genetics 108:719–731
18. Meyer E, Wiegand P, Brinkmann B (1995) Phenotype differences of STRs in 7 human populations. Int J Legal Med 107: 314–322
19. Budowle B, Monson KL, Giusti AM, Brown BL (1994) The assessment of frequency estimates of Hae III-generated VNTR profiles in various reference databases. J Forensic Sci 39:319–352
20. Barber MD, Piercy RC, Andersen JF, Parkin BH (1995) Structural variation of novel alleles at the Hum vWA and Hum FES/FPS short tandem repeat loci. Int J Legal Med 108:31–35
21. Kozma Z, Bajnóczky I, Nishi K (1997) Allele frequency distribution of the STR systems HumVWA, HumFES/FPS and HumF13A01 in two racial groups: Hungarian and Japanese population data. In: Proceedings of the 3rd International Symposium Advances in Legal Medicine, Osaka, Japan (in press)
22. Kozma Z, Usiyama I, Sétáló J, Bajnóczky I, Nishi K (1997) HumTH01, HumTPOX and HumCSF1PO STR loci: allele frequency distributions in a Hungarian and a Japanese population group investigated by a triplex PCR and manual typing. Acta Crim Jap (in press)