Int J Legal Med (2000) 113 : 272–275

ORIGINAL ARTICLE

B. Egyed · S. Füredi · M. Angyal · L. Boutrand
A. Vandenberghe · J. Woller · Z. Pádár

Analysis of eight STR loci in two Hungarian populations

Received: 18 May 1999 / Received in revised form: 26 August 1999

Abstract A multiplex reaction for the eight STR loci D3S1358, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820 was used to generate allele frequency databases for two Hungarian population samples, Caucasians from the Budapest area and Romanies from Baranya county. During the analysis two intermediate-sized alleles and a sequence variant allele were observed at the D7S820 locus. All three types of allelic variants were found to have modifications in the same block of a $(T)_9$ stretch located within the 3′ flanking region of each allele, which may indicate a possible higher mutation rate of this $(T)_9$ block. For the loci D3S1358 and D7S820 the Romany population database showed departures from Hardy-Weinberg equilibrium. The forensic efficiency values for the Romany population were slightly different from those found in the Hungarian Caucasian population. Comparing the allele frequency values by G-statistic, calculating the F_{ST} indices and with the pair-wise comparisons of inter-population variance, the two Hungarian populations could be distinguished using data from the eight STR loci.

Key words Multiplex STR profiling · Capillary electrophoresis · Variant alleles · Romany population · Population genetics

Introduction

In the last decade STR profiling has gained a central role in forensic identification and paternity testing. Using the STR markers in genetic investigations of the representative groups of human population, it also becomes possible to rapidly obtain information about the scale of population genetic effects caused by several factors such as inbreeding and substructuring [1, 2, 3]. Among the Hungarian inhabitants, the Romanies represent one of the most relevant ethnic groups of the population. Several studies on STRs have been carried out on the Hungarian Romanies [4, 5, 6]. The aim of this study was to generate and evaluate genotype databases of eight additional STRs for a Romany as well as a reference Caucasian population sample by analysing the loci D3S1358, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, and D7S820.

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

M. Angyal
Department of Forensic Medicine,
Medical University School of Pécs, P.O. Box 99,
7643 Pécs, Hungary

L. Boutrand · A. Vandenberghe
Laboratory of Human Molecular Genetics, Faculty of Pharmacy,
University Claude Bernard Lyon 1, 69008 Lyon, France

Materials and methods

Blood samples were collected from 223 unrelated Hungarian Caucasian individuals (116 males and 107 females) living in the Budapest area (Central Hungary) and 206 unrelated Romany individuals (79 males and 127 females) residing in Baranya county (southwestern Hungary). The former sample was used as a reference group for the mixed character of the Hungarian population, because the samples were collected in a blood bank and were of mixed background.

DNA samples (0.5–2 ng) were coamplified using reagents provided in the AmpF*l*STR Profiler Plus PCR amplification kit. The PCR products were analysed by fluorescence-based automated detection and capillary electrophoresis system on an ABI PRISM 310 Genetic Analyzer. All procedures were carried out in accordance with the manufacturer's instructions (PE Applied Biosystems, Foster City, Calif.). Genotyping was performed using the GeneScan Analysis v2.1 and Genotyper v2.0 software. Allelic designation was done according to Bär et al. [7].

About 100 ng of genomic DNA from samples with intermediate-sized alleles on locus D7S820 were amplified for sequencing analysis using published primers (GenBank G08616). The purification of PCR products was carried out with the Qiaquick PCR purification kit (Qiagen, Hilden, Germany). PCR products of 9var/10 and 8var/8 alleles were cloned in pGEM-T (Promega, Madison, Wisc.) and sequenced on an ABI PRISM 310 Genetic Analyzer using the ABI PRISM d-Rhodamine Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer). Sequencing was done with forward as well as reverse primers.

Possible divergence of population data from Hardy-Weinberg expectations (HWE) was determined by the exact test [8]. The allele frequency profile comparisons were performed by G-statistic

test using a software for R × C contingency tables. Population substructure was measured by calculating the unbiased single-locus coancestry coefficient F_{ST} (θ) [9] and its *Φ*-statistic analogue Φ_{ST} in the analysis of molecular variance (AMOVA) [10, 11] using the software ARLEQUIN v1.1.

Results and discussion

A sub-sample of both populations had already been analysed for the vWA locus and published previously [4, 12], but no essential differences were found between the allele frequency distributions of the sub-samples and those with the increased number of samples.

No significant differences in electrophoretic mobility were observed between the same STR alleles using the ABI Prism 310 instrument. Due to the high resolution power of this capillary electrophoresis system, two intermediate-sized alleles could be detected at the locus D7S820. The genotypes with variant alleles were 9var/10 and 8var/8. PCR products of each genotype were cloned and in order to exclude cloning artefacts, at least two clones of each allele were completely sequenced. The sample carrying the 9var allele was from the Central Hungarian population database, and the 8var allele was detected in the blood sample of an Albanian male from Yugoslavia. The variant alleles were designated as 9.+1 and 8.-1 according to their sequence structure (Fig. 1.). Allele 9.+1 was found to have a T insertion in the $(T)_9$ stretch located within the 3′ flanking region, while the allele 8.-1 lacked a T in this stretch. In the 9var/10 sample the allele 10 showed a sequence variation because a T→A transversion could be observed at the end of the $(T)_9$ block. This variation was found in all six sequenced clones of allele 10. The variant sequences may suggest that the $(T)_9$ block is the polymorphic part of the locus D7S820 due to its relatively high mutation rate. This hypothesis should be confirmed by examining further sequences from the 3′ flanking region of parent-child genotype pairs. Some poly-A tracts (poly-T on the opposite strand) belonging to Alu repetitive elements are known to be polymorphic [13] and Alu sequences are often associated and probably actively involved in the genesis of short repeat sequences [14, 15]. However, the sequence 3′ of the $(T)_9$ block does not show any homology with an Alu repetitive element. We also checked with the computer program BLAST [16] if this poly-T track could be part of a retrotransposed pseudogene, but no significant homology with genes known at present was found (search effected via the www. interface at the NCBI, http://www.ncbi.nlm.nih.gov, 4. 20. 1999). However, 100% homology was found with the human PAC clone DJ0649P17 (GenBank AC004848). The sequence of the entire D7S820 locus is contained in this clone which has been located at 7q11.23-q21 and thus permits a more refined localisation of D7S820 so far localised only to 7q [17].

The Hungarian Romany allele frequency values (Tables 1, 2, 3) for all loci were significantly different from the Hungarian Caucasian population data ($P \leq 10^{-3}$). There was evidence of departures from HWE for the loci D3S1358 and D7S820 in the Romany population sample (Table 4). These departures are likely to be sampling effects but could also be due to inbreeding effects in such a genetically closed subpopulation. Comparing the allele frequency values of the Central Hungarian population database with other Caucasians by G-statistics, there were small discrepancies at the loci FGA, D8S1179, D21S11, D18S51 between the STR data presented here and an Italian [18] as well as a south-western Hungarian [19] database. Other population databases showed no differences [20, 21, 22]. The combined forensic efficiency values observed in the Romany population sample for the eight loci (PM = 1.2×10^{-9}, PE = 0.99931) were slightly different from those found in the Hungarian Caucasian database (PM = 2.5×10^{-9}, PE = 0.99975) (PM = matching probability, PE = power of exclusion). Despite the higher observed heterozygosity values of the Caucasians, the summarised PM value proved lower in the Romany population. Calculating Wright's F_{ST} indices for the two populations, a relatively high level of F_{ST} values (Table 4) were found for the eight loci compared to the previous observations in other Caucasian populations [1, 3]. At four loci this genetic correlation was reinforced by Φ-statistics, where genetic variation was considered at the molecular level. The results of pair-wise population comparisons were very sim-

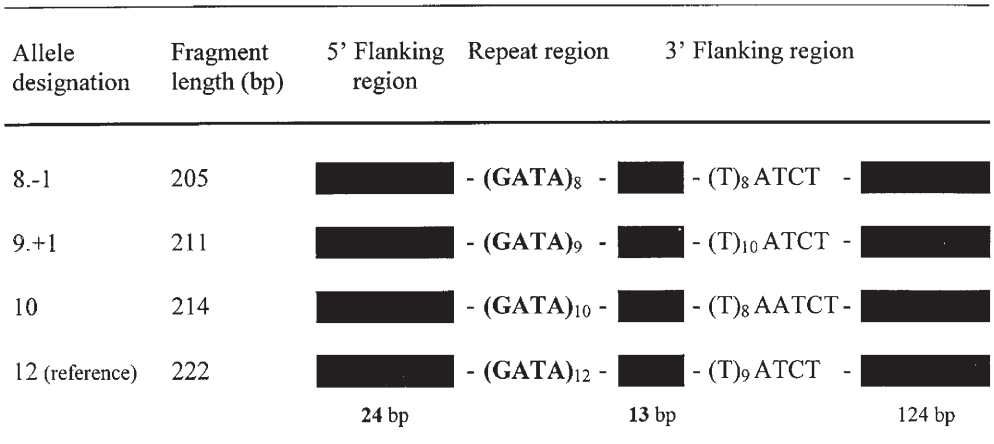

Fig. 1 Sequence structure and fragment length of four D7S820 alleles. 8.-1 and 9.+1 are intermediate-sized alleles; 10 was found to be a sequence variant allele as compared to the published reference sequence (GenBank G08616) containing 12 (GATA) repeats. For sequencing the original STS primers were used that differ from the primer pair contained in the kit and thus lead to shorter PCR products

Table 1 Allele frequency values of the STR loci in the Hungarian Caucasian (N = 446) and Romany (N = 412) population (*BuCa* = Hungarian Caucasians residing in the Budapest area, *BaRo* = Hungarian Romanies residing in Baranya county)

Allele	D3S1358		FGA		D18S51	
	BuCa	BaRo	BuCa	BaRo	BuCa	BaRo
10					0.002	
11		0.002			0.016	
12					0.112	0.080
13	0.007				0.150	0.107
14	0.087	0.032			0.170	0.109
15	0.247	0.255			0.119	0.194
16	0.244	0.245	0.002	0.002	0.112	0.109
17	0.231	0.138			0.117	0.277
18	0.161	0.313	0.020		0.083	0.032
19	0.020	0.015	0.076	0.165	0.052	0.083
20	0.002		0.157	0.109	0.043	0.005
21			0.195	0.083	0.016	
21.2			0.002			
22			0.191	0.150	0.007	
22.2			0.011	0.022		
23			0.114	0.121		0.005
23.2			0.002	0.007		
24			0.123	0.248		
24.2			0.004	0.005		
25			0.056	0.056	0.002	
25.2			0.004	0.017		
26			0.040	0.015		

Table 2 Same legend as Table 1, but with different loci

Allele	D8S1179		D5S818		D13S317		D7S820	
	BuCa	BaRo	BuCa	BaRo	BuCa	BaRo	BuCa	BaRo
7			0.004				0.009	0.012
8	0.011	0.015	0.002		0.137	0.199	0.143	0.078
9	0.016	0.002	0.040	0.049	0.096	0.117	0.137	0.160
9.+1							0.002	
10	0.076	0.024	0.078	0.061	0.049	0.049	0.294	0.129
11	0.081	0.141	0.318	0.231	0.327	0.180	0.235	0.512
12	0.182	0.153	0.370	0.507	0.260	0.367	0.150	0.095
13	0.287	0.257	0.166	0.148	0.092	0.075	0.025	0.015
14	0.240	0.163	0.020	0.005	0.038	0.015	0.004	
15	0.083	0.194						
16	0.020	0.049						
17	0.002	0.002						
18	0.002							

Table 3 Same legend as Table 1, but a different locus

Allele	D21S11	
	BuCa	BaRo
26	0.002	
27	0.049	0.005
28	0.191	0.100
29	0.195	0.279
29.2	0.002	
30	0.217	0.143
30.2	0.049	0.068
31	0.063	0.051
31.2	0.096	0.083
32	0.011	0.002
32.2	0.076	0.214
33.2	0.040	0.039
34.2	0.007	0.017

ilar to the findings described previously, where a significantly smaller Romany population sample (135 individuals) was analysed [4, 5, 6].

In conclusion, two Hungarian population databases have been established for eight STR loci using multiplex amplification and an automated fluorescent detection system. During the analysis two intermediate-sized alleles and a sequence variant allele were observed at the D7S820 locus. Significant differences were found between the Caucasian and the Romany population database by performing G- and F-statistics. The results suggest that the possibility of population differentiation should be taken into account in the calculation of match probabilities in Hungarian forensic cases.

Table 4 Statistical values of the population genetic survey in Hungarian Caucasians and Romanies (H_o = observed heterozygosity, *PD* = power of discrimination, *PE* = power of exclusion, *BuCa* = Hungarian Caucasians residing in the Budapest area, *BaRo* = Hungarian Romanies residing in Baranya county). [a] The population pair-wise F_{ST} or Φ_{ST} value represents no statistically significant difference at the $P = 0.05$ level

Locus	H_o		Exact test (*P*)		PD		PE		Genetic structure	
	BuCa	BaRo	BuCa	BaRo	BuCa	BaRo	BuCa	BaRo	F_{ST}	Φ_{ST}
D3S1358	0.771	0.723	0.788	0.023	0.926	0.898	0.587	0.527	0.020	0.023
FGA	0.928	0.878	0.870	0.859	0.937	0.959	0.722	0.704	0.021	0.004[a]
D8S1179	0.919	0.767	0.349	0.084	0.895	0.944	0.625	0.649	0.014	0.017
D21S11	0.931	0.806	0.669	0.351	0.942	0.948	0.707	0.669	0.022	0.052
D18S51	0.917	0.830	0.875	0.295	0.951	0.952	0.764	0.678	0.022	0.006[a]
D5S818	0.863	0.655	0.646	0.736	0.843	0.841	0.491	0.425	0.017	0.001[a]
D13S317	0.890	0.714	0.808	0.244	0.892	0.917	0.589	0.570	0.022	0.007[a]
D7S820	0.879	0.621	0.314	0.011	0.896	0.850	0.599	0.467	0.068	0.012

References

1. Evett IW, Lambert JA, Buckleton JS, Weir BS (1996) Statistical analysis of a large file of data from STR profiles of British Caucasians to support forensic casework. Int J Legal Med 109: 173–177
2. Gill P, Evett I (1995) Population genetics of short tandem repeat (STR) loci. Genetica 96: 69–87
3. Evett IW, Gill PD, Lambert JA, Oldroyd N, Frazier R, Watson S, Panchal S, Connolly A, Kimpton C (1997) Statistical analysis of data for three British ethnic groups from a new STR multiplex. Int J Legal Med 110: 5–9
4. Füredi S, Angyal M, Kozma Z, Sétáló J, Woller J, Pádár Z (1997) Semi-automatic DNA profiling in a Hungarian Romany population using the STR loci HumVWA31, HumTH01, HumTPOX and HumCSF1PO. Int J Legal Med 110: 184–187
5. Füredi S, Kozma Z, Woller J, Pádár Z, Angyal M, Bajnóczky I, Nishi K (1998) Population genetic data on four STR loci in a Hungarian Romany population. Int J Legal Med 112: 72–74
6. Füredi S, Woller J, Pádár Z, Angyal M (1999) Y-STR haplotyping in two Hungarian populations. Int J Legal Med 113: 38–42
7. Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B (1997) DNA recommendations. Further report of the DNA commisson 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. Weir BS (1996) Genetic data analysis II. Sinauer Associates, Sunderland, Mass
10. Excoffier L, Smouse P, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479–491
11. Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142: 1061–1064
12. Füredi S, Budowle B, Woller J, Pádár Z (1996) Hungarian population data on six STR loci – HUMVWA31, HUMTH01, HUMCSF1PO, HUMFES/FPS, HUMTPOX, AND HUMHPRTB – derived using multiplex PCR amplification and manual typing. Int J Legal Med 109: 100–101
13. Economou EP, Bergen AW, Warren AC, Antonarakis SE (1990) The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. Proc Natl Acad Sci USA 87: 2951–2954
14. Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995) Alu repeats: a source for the genesis of primate microsatellites. Genomics 29: 136–144
15. Yandava CN, Gastier JM, Pulido JC, Brody T, Sheffield V, Murray J, Buetow K, Duyck GM (1997) Characterization of Alu repeats that are associated with trinucleotide and tetranucleotide repeat microsatellites. Genome Res 7: 716–724
16. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402
17. Bouffard GG, Iyver LM, Idol JR, Braden VV, Cunningham AF, Weintraub LA, Mohr-Tidwell RM, Peluso DC, Fulton RS, Leckie MP, Green ED (1997) A collection of 1814 human chromosome 7 specific STSs. Genome Res 7: 59–64
18. Garofano L, Pizzamiglio M, Vecchio C, Lago G, Floris T, D'Errico G, Brembilla G, Romano A, Budowle B (1998) Italian population data on thirteen short tandem repeat loci: HUMTH01, D21S11, D18S51, HUMVWA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820, D13S317, D5S818, D3S1358. Forensic Sci Int 97: 53–60
19. Kozma Z, Nagai A, Woller J, Füredi S, Sétáló J, Ohya I, Nishi K (1998) Fluorescence based co-amplification and automated detection of the STR loci HUMFIBRA and HUMD21S11 in a Hungarian Caucasian population sample. Int J Legal Med 111: 103–104
20. Szabo A, Schürenkamp M, Hühne J (1998) Hungarian population data for six STR loci. Int J Legal Med 111: 49–51
21. Entrala C, Lorente M, Lorente JA, Carlos Alvarez J, Moretti T, Budowle B, Villanueva E (1998) Fluorescent multiplex analysis of nine STR loci: Spanish population data. Forensic Sci Int 98: 179–183
22. Zupanič I, Balažic J, Komel R (1998) Analysis of nine short tandem repeat (STR) loci in the Slovenian population. Int J Legal Med 111: 248–250