Allele Frequencies for 15 Short Tandem Repeat Loci in Representative Sample of Croatian Population

Petar Projić[1], Vedrana Škaro[1], Ivana Šamija[2], Naris Pojskić[3], Adaleta Durmić-Pašić[3], Lejla Kovačević[3], Narcisa Bakal[3], Dragan Primorac[4,5], Damir Marjanović[1,3]

[1]Group for Forensic Genetics, Department of Molecular Medicine, Ruđer Bošković Institute, Zagreb, Croatia
[2]Institute of Gynecologic Cytology, University Department of Gynecology and Obstetrics, Zagreb University Hospital Center, Zagreb, Croatia
[3]Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina
[4]Department for Chemistry and Biochemistry, Osijek University School of Medicine, Osijek, Croatia
[5]Department of Medical Diagnostics, Split University School of Medicine, Split, Croatia

> **Correspondence to:**
Damir Marjanović
Kemalbegova 10
71000 Sarajevo, Bosnia and Herzegovina
marjanovd@hotmail.com

> **Received:** May 14, 2007
> **Accepted:** July 6, 2007

Aim To study the distribution of allele frequencies of 15 short tandem repeat (STR) loci in a representative sample of Croatian population.

Methods A total of 195 unrelated Caucasian individuals born in Croatia, from 14 counties and the City of Zagreb, were sampled for the analysis. All the tested individuals were voluntary donors. Buccal swab was used as the DNA source. AmpFlSTR® Identifiler® was applied to simultaneously amplify 15 STR loci. Total reaction volume was 12.5 μL. The PCR amplification was carried out in PE Gene Amp PCR System Thermal Cycler. Electrophoresis of the amplification products was preformed on an ABI PRISM 3130 Genetic Analyzer. After PCR amplification and separation by electrophoresis, raw data were compiled, analyzed, and numerical allele designations of the profiles were obtained. Deviation from Hardy-Weinberg equilibrium, observed and expected heterozygosity, power of discrimination, and power of exclusion were calculated. Bonferroni's correction was used before each comparative analysis.

Results We compared Croatian data with those obtained from geographically neighboring European populations. The significant difference (at $P<0.01$) in allele frequencies was recorded only between Croatian and Slovenian populations for vWA locus. There was no significant deviation from Hardy-Weinberg equilibrium for all the observed loci.

Conclusion Obtained population data concurred with the expected "STR data frame" for this part of Europe.

> **Croat Med J. 2007;48:473-7**

In nearly thirty years since its first formal application (1), DNA analysis has been promoted into a core method for different types of examination, from routine paternity testing and massive identification of human remains to complicated forensic casework analysis (2). Introduction of a set of short tandem repeat (STR) loci as the markers induced a significant progress in this field of science (3-5). DNA typing for forensic, identification, and paternity testing purposes is based on the same techniques that are routinely employed in a wide range of medical and genetic situations, such as diagnosis, population genetic studies, and gene mapping (6). Therefore, importance of understanding the used marker heterogeneity within different populations is constantly emphasized (2,7).

Different number and different sets of STR loci in a different number of the individuals of different geographical origin were used in previous studies of Croatian population. Namely, 8 STR loci have already been employed in a study of population of Republic of Croatia (8) and 13 CODIS STR loci in the analysis of the population of Southern Croatia (9). Here we have analyzed the distribution of allele frequencies at 15 STR loci in a representative sample of the Croatian population. Additionally, we compared these data with those obtained from geographically neighboring European populations.

Material and methods

Population

This study is based on a representative sample of Croatian population. A total of 195 unrelated individuals from 14 Croatian counties and the City of Zagreb participated in this study. Eighty per cent of tested persons were born in 5 counties (Splitsko-dalmatinska, Dubrovačko-neretvanska, Osječko-baranjska, Šibensko-kninska, Primorsko-goranska) and the City of Zagreb. Other 20% originated from Zadarska, Karlovačka, Istarska, Međimurska, Virovitičko-podravska, Bjelovarsko-bilogorska, Koprivničko-križevačka, and Krapinsko-zagorska county. All the tested individuals were voluntary donors who gave informed consent. Sample collection from voluntaries received ethical approval by the Ethical council of Zagreb. University School of Medicine.

Collected samples

Buccal swabs were used as the DNA source. All specimens were air-dried, placed in 1.5-mL tubes, and immediately transported to the DNA laboratory of the Group for Forensic Genetics, Department of Molecular Medicine, Ruđer Bošković Institute, Zagreb. The samples were stored at -80°C until DNA extraction by Qiagen Dnaeasy™ Tissue Kit (10).

DNA analysis.

AmpFlSTR®Identifiler® (Applied Biysystems, Foster City, CA, USA) was used to simultaneously amplify 15 STR loci including 13 CODIS (D3S1358, TH01, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, CSF1PO, vWA, D8S1179, TPOX, FGA) and 2 additional loci (D2S1338 and D19S433), as well as the gender determination locus Amelogenin. Similar amounts of DNA were used in all PCR reactions. Amplification was carried out as described previously (11), with modification of decreasing the total volume of each reaction to 12.5 μL. The PCR amplification was carried out in AB Gene Amp PCR System 9700 Thermal Cycler (Applied Biosystems) according to the manufacturer's recommendations. Electrophoresis of the amplification products was preformed on an ABI PRISM 3130 Genetic Analyzer (ABI). The raw data were compiled and analyzed using the accessory software – ABI Data Collection Software and GeneMapper™ 3.2 (Applied Biosystems).

Statistical analysis

Deviation from Hardy-Weinberg equilibrium (12), as well as observed and expected heterozygosity (13), were calculated using Powermarker

software (14). Power of discrimination and power of exclusion were calculated by Microsoft® Excel workbook template – PowerStats (15). Also, we performed exact test of population differentiation (16) with Bonferroni correction (17).

Results and discussion

Allele frequencies, heterozygosity (observed and expected), results of exact test, power of discrimination, and power of exclusion are shown in Tables 1 and 2. The most frequent allele among all allelic variants over all loci was the allele 8 at TPOX (0.5487). The comparison of the proportion of the calculated heterozygosity across loci showed higher value for FGA and D18S51, and lower for TPOX.

No significant deviation from Hardy-Weinberg equilibrium was found for any of the observed loci. Allele frequencies of the observed loci were compared with the data obtained from geographically close European regions, such as Slovenia (18), Bosnia and Herzegovina (19,20), Serbia (our unpublished data), and Macedonia (21).

Significance threshold was set at $P<0.05$ or $P<0.01$ after Bonferroni correction. At the level $P<0.05$, statistically significant differences in allele frequencies were noticed between Croatian and Serbian populations for D21S11 locus, as well as between Croatian and Slovenian

Table 1. Croatian allele frequencies for 8 autosomal loci (n = 195)*

Allele	D3S1358	THO1	D21S11	D18S51	D2S1338	D5S818	D13S317	D7S820
5	–	–	–	–	–	–	–	–
6	–	0.2641	–	–	–	–	–	–
7	–	0.1385	–	–	–	–	–	0.0051
8	–	0.1308	–	–	–	–	0.1718	0.2000
9	–	0.1667	–	–	–	0.0359	0.0718	0.1795
9,3	–	0.2846	–	–	–	–	–	–
10	–	0.0154	–	0.0103	–	0.0615	0.0615	0.2590
11	–	–	–	0.0461	–	0.3256	0.3308	0.1949
12	–	–	–	0.0975	–	0.3897	0.2487	0.1410
13	–	–	–	0.1461	–	0.1795	0.0821	0.0179
14	0.1051	–	–	0.1923	–	0.0051	0.0333	0.0026
15	0.2538	–	–	0.1359	0.0026	–	–	–
16	0.2410	–	–	0.1692	0.0692	–	–	–
17	0.2359	–	–	0.0897	0.2154	0.0026	–	–
18	0.1487	–	–	0.0564	0.0923	–	–	–
19	0.0103	–	–	0.0205	0.0897	–	–	–
20	0.0051	–	–	0.0179	0.1615	–	–	–
20,2	–	–	–	–	–	–	–	–
21	–	–	–	0.0154	0.0256	–	–	–
22	–	–	–	0.0026	0.0179	–	–	–
22,2	–	–	–	–	–	–	–	–
23	–	–	–	–	0.1026	–	–	–
23,2	–	–	–	–	–	–	–	–
24	–	–	–	–	0.0872	–	–	–
25	–	–	–	–	0.1256	–	–	–
26	–	–	–	–	0.0103	–	–	–
27	–	–	0.0359	–	–	–	–	–
28	–	–	0.1538	–	–	–	–	–
29	–	–	0.1641	–	–	–	–	–
29,2	–	–	0.0051	–	–	–	–	–
30	–	–	0.2821	–	–	–	–	–
30,2	–	–	0.0487	–	–	–	–	–
31	–	–	0.0590	–	–	–	–	–
31,2	–	–	0.0923	–	–	–	–	–
32	–	–	0.0077	–	–	–	–	–
32,2	–	–	0.1128	–	–	–	–	–
33,2	–	–	0.0385	–	–	–	–	–
H(ob)	0.7077	0.8051	0.8410	0.8974	0.8205	0.6769	0.7744	0.8256
H(ex)	0.7885	0.7850	0.8399	0.8706	0.8712	0.7047	0.7824	0.8025
P	0.1853	0.6202	0.6020	0.4127	0.2506	0.0997	0.9526	0.1905
Power of discrimination	0.930	0.915	0.954	0.964	0.967	0.855	0.919	0.924
Power of exclusion	0.457	0.599	0.677	0.790	0.638	0.393	0.552	0.647

*Abbreviations: H(ob) – observed heterozygosity; H(ex) – expected heterozygosity; P – Hardy-Weinberg equilibrium, exact test.

Table 2. Croatian allele frequencies for 7 autosomal loci (n = 195)*

Allele	D16S539	CSF1P0	D19S433	vWA	D8S1179	TPOX	FGA
5	–	–	–	–	–	–	–
6	–	–	–	–	–	–	–
7	–	–	–	–	–	0.0077	–
8	0.0282	0.0026	–	–	0.0179	0.5487	–
9	0.1000	0.0205	0.0026	–	0.0128	0,0846	–
9,3	–	–	–	–	–	–	–
10	0.0590	0.2538	0.0026	–	0.0718	0.0641	–
11	0.2795	0.3026	–	–	0.0564	0.2667	–
12	0.3282	0.3538	0.0872	–	0.1487	0.0308	–
13	0.1718	0.0590	0.2128	0.0051	0.3487	–	–
13,2	–	–	0.0103	–	–	–	–
14	0.0333	0.0026	0.3154	0.1051	0.2205	–	–
14,2	–	–	0.0308	–	–	–	–
15	–	0.0051	0.1744	0.1385	0.1103	–	–
15,2	–	–	0.0564	–	–	–	–
16	–	–	0.0641	0.2205	0.0128	–	0.0026
16,2	–	–	0.0385	–	–	–	–
17	–	–	0.0026	0.2256	–	–	–
18	–	–	–	0.1974	–	–	0.0128
18,2	–	–	0.0026	–	–	–	–
19	–	–	–	0.0872	–	–	0.0692
20	–	–	–	0.0154	–	–	0.1231
20,2	–	–	–	–	–	–	–
21	–	–	–	0.0051	–	–	0.1846
21,2	–	–	–	–	–	–	0.0026
22	–	–	–	–	–	–	0.1897
22,2	–	–	–	–	–	–	0.0128
23	–	–	–	–	–	–	0.1103
23,2	–	–	–	–	–	–	0.0026
24	–	–	–	–	–	–	0.1487
25	–	–	–	–	–	–	0.1000
26	–	–	–	–	–	–	0.0308
26,2		–	–	–	–	–	0.0026
27	–	–	–	–	–	–	0.0077
28	–	–	–	–	–	–	–
29	–	–	–	–	–	–	–
29,2	–	–	–	–	–	–	–
30	–	–	–	–	–	–	–
30,2	–	–	–	–	–	–	–
H(ob)	0.8000	0.7231	0.8051	0.8051	0.7385	0.6872	0.8718
H(ex)	0.7693	0.7149	0.8074	0.8234	0.7865	0.6158	0.8643
P	0.9496	0.7421	0.7034	0.3887	0.6304	0.0925	0.8806
Power of discrimination	0.913	0.862	0.935	0.942	0.927	0.782	0.963
Power of exclusion	0.599	0.465	0.609	0.599	0.490	0.416	0.738

*Abbreviations: H(ob) – observed heterozygosity; H(ex) – expected heterozygosity; P – Hardy-Weinberg equilibrium, exact test.

population for vWA, D7S820, and D18S51 loci. However, at the threshold of $P<0.01$, obvious difference in allele frequencies was recorded only between Croatian and Slovenian populations for vWA.

Previous studies (22,23) suggested that 100-150 tested individuals per population might provide sufficient likelihood calculation. Naturally, collecting information from more samples usually adds to the precision of the obtained results (2). Therefore, considering the size of recent Croatian population, the total number of unrelated individuals (n = 195) tested in this case is sufficient for reliable estimation of the frequencies of major alleles for any DNA locus.

In other words, the number of tested individuals, as well as the randomization in sample collection allows the usage of the obtained data in the calculation of different DNA forensic statistical parameters.

On the other hand, general results of this study were expected. The obtained population data completely concur with the expected "STR data frame" for this part of Europe. The high level of similarity allows combining sets of regional data to yield a larger data set (2) and increase the

precision of various "DNA forensic calculations" from a mix of cases from different countries.

Acknowledgment

This study is a part of the scientific project "Development of Scientific System for DNA Analysis of Archeological Bone Samples," supported by the grant No. 098-1300855-2738 from the Ministry of Science, Education, and Sports, Republic of Croatia.

References

1 Jeffreys AJ, Brookfield JF, Semeonoff R. Positive identification of an immigration test-case using human DNA fingerprints. Nature. 1985;317:818-9. Medline:4058586

2 Butler JM. Forensic DNA typing: biology, technology and genetics of STR markers. 2nd ed. London: Elsevier Academic Press; 2005.

3 Gill P, Kimpton CP, Urquhart A, Oldroyd N, Millican ES, Watson SK, et al. Automated short tandem repeat (STR) analysis in forensic casework – a strategy for the future. Electrophoresis. 1995;16:1543-52. Medline:8582334

4 Thomson JA, Pilotti V, Stevens P, Ayres KL, Debenham PG. Validation of short tandem repeat analysis for the investigation of cases of disputed paternity. Forensic Sci Int. 1999;100:1-16. Medline:10356771

5 Alonso A, Anđelinović S, Martin P, Sutlović D, Erceg I, Huffine E, et al. DNA typing from skeletal remains: evaluation of multiplex and megaplex STR systems on DNA isolated from bone and teeth samples. Croat Med J. 2001;42:260-6. Medline:11387635

6 National Research Council. The evaluation of forensic DNA evidence. Washington (DC): National Academy Press; 1996.

7 Cohen JE. DNA fingerprinting for forensic identification: potential effects on data interpretation of subpopulation heterogeneity and band number variability. Am J Hum Genet. 1990;46:358-68. Medline:2301401

8 Ross J, Parson W, Furać I, Kubat M, Holland M. Multiplex PCR amplification of eight STR loci in Austrian and Croatian Caucasian populations. Int J Legal Med. 2001;115:57-60. Medline:11599771

9 Schanfield M, Gabriel MN, Andelinovic S, Reynolds RL, Ladd C, Lee HC, et al. Allele frequencies for the 13 CODIS STR loci in a sample of southern Croatians. J Forensic Sci. 2002;47:669-70. Medline:12051361

10 Qiagen Companies. QIAGEN Genomic DNA Handbook. Hilden: Qiagen; 2001.

11 Applied Biosystems. AmpFlSTR® Identifiler™ PCR Amplification Kit User's Manual. Foster City (CA): Applied Biosystems; 2005.

12 Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics. 1992;48:361-72. Medline:1637966

13 Nei M. Molecular evolutionary genetics. New York (NY): Columbia University Press; 1987.

14 Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2005;21:2128-9. Medline:15705655

15 Tereba A. Tools for analysis of population statistics. Profiles in DNA. 1999;3:14-6.

16 Raymond M, Rousset F. An exact test for population differentiation. Evolution Int J Org Evolution. 1995;49:1280-3.

17 Bonferroni CE. Theory statistics of the classes and calculation of the probabilities [in Italian]. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. 1936;8:3-62.

18 Drobnic K, Pojskic N, Bakal N, Marjanovic D. Allele frequencies for 15 short tandem repeat loci in Slovenian population. J Forensic Sci. 2005;50:1505-7. Medline:16382858

19 Marjanovic D, Bakal N, Pojskic N, Kapur L, Drobnic K, Primorac D, et al. Allele frequencies for 15 short tandem repeat loci in a representative sample of Bosnians and Herzegovinians. Forensic Sci Int. 2006;156:79-81. Medline:16410155

20 Marjanovic D, Pojskic N, Davoren J, Kovacevic L, Durmic A, Bakal N, et al. Population data at two STR loci D2S1338 and D19S433 in the sample of multinational Bosnia and Herzegovina residents. J Forensic Sci. 2006;51:1219-20. Medline:17018117

21 Jakovski Z, Nikolova K, Furac I, Masic M, Janeska B, Kubat M. Allele frequencies for 15 STR loci in a population from the Republic of Macedonia. Int J Legal Med. 2006;120:53-5. Medline:16088410

22 Chakraborty R. Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. Hum Biol. 1992;64:141-59. Medline:1559686

23 Evett IW, Gill P. A discussion of the robustness of methods for assessing the evidential value of DNA single locus profiles in crime investigations. Electrophoresis. 1991;12:226-30. Medline:2040269