Genetic Diversity of 21 Autosomic STR Markers of the CODIS System in Populations of Eastern Europe

V. N. Kharkov[a], *, S. A. Kotova[b], N. A. Kolesnikov[a], E. A. Spivak[b], K. V. Vagaitseva[a], T. V. Zabavskaya[b], V. I. Rybakova[b], A. S. Parfenova[b], A. N. Verchuk[b], N. K. Yankovsky[c, d], and V. A. Stepanov[a]

[a] *Research Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, 634050 Russia*
[b] *Scientific Practical Center of the State Forensic Expertise Committee of the Republic of Belarus, Minsk, 220114 Belarus*
[c] *Vavilov Institute of General Genetics, Moscow, 119991 Russia*
[d] *Moscow State University, Moscow, 119991 Russia*
*e-mail: vladimir.kharkov@medgenetics.ru

Abstract—For the first time, genetic characteristics of 15 local populations of eight ethnic groups of Eastern Europe were obtained using an expanded panel that included 21 autosomal STR markers used in forensic practice for DNA identification and kinship establishment. The assessment of polymorphism of markers in the Russian, Belarusian, and Moldovan populations was carried out, and the levels of intra- and interpopulation genetic differentiation of the studied population groups were characterized. The parameters of the informational significance of the system of autosomal microsatellites for expert DNA identification and establishment of kinship were estimated. A database of allele frequencies of autosomal STR loci has been developed, which can be used to carry out probabilistic and statistical calculations when assessing the level of reliability of an expert study in the countries of the Union State. It was shown that population genetic structure should be taking into account in forensic DNA analysis.

INTRODUCTION

Each individual carries a unique genome, which is characterized by a nonrepeating combination of alleles of many polymorphic DNA markers. The main method of forensic identification based on DNA analysis is the study of autosomal short tandem repeats (STR), which have a high level of allelic polymorphism. Genotyping of the allelic state of 15—20 loci of such repeats has a sufficient level of information content to establish individual identity. For the analysis of autosomal STR loci, genotyping reagent kits are used, the main manufacturers of which are Thermo Fisher Scientific (United States) and Promega Corporation (United States). The reagent kits of these manufacturers are based on the CODIS marker panel (Combined DNA Index System), proposed in 2015 and approved for use for forensic DNA analysis in 2017 by the Federal Bureau of Investigation of the US Department of Justice. Despite the fact that the CODIS system was developed and operates in the United States, many other countries, including the Republic of Belarus and the Russian Federation, rely on the standard of this system when carrying out DNA identification of a person in forensic examination. In 2017, the CODIS marker panel, which previously included 13 microsatellites, was expanded. Currently, the so-called core of the system includes 20 autosomal STR markers: CSF1PO, FGA, THO1, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, D22S1045. These changes have led to the need to update the reference databases containing the information necessary for the probabilistic and statistical assessment of the results of the expert study, as well as to update the information on the level of genetic differentiation of local populations by markers of the extended CODIS system and the informational significance of the new core of the system for DNA identification in a number of local populations of Eastern Europe. Most publications devoted to this issue currently contain information on only 16 STR markers [1–14].

In this work, we characterized the allele frequencies and the level of genetic differentiation of populations and assessed the identification potential of the system of 21 autosomal STR markers (D1S1656, D2S441, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S395,

D18S51, D21S11, D22S1045, CSF1PO, FGA, Penta D, Penta E, TH01, TPOX, vWA) in population samples of eight ethnic groups of Eastern Europe—Russians, Belarusians, Moldovans, Gagauz, Poles, Komi, Mari, and Ukrainians.

MATERIALS AND METHODS

Research material consisted of DNA samples from 15 different local populations of eight peoples of Eastern Europe. The total number of samples was 1093 individuals. DNA samples were taken from unrelated individuals from the following populations living in Russia (Russians—205 samples, Komi—83, Mari—76), Moldova (Moldovans—117, Gagauz—57), and Belarus (Russians—57, Belarusians—413, Poles—32, Ukrainians—110). The material from the populations of Belarus contains six samples from the historical and ethnographic regions of the republic: Eastern Polesie ($N = 41$), Western Polesie ($N = 56$), Central region ($N = 135$), Podneprovye ($N = 80$), Poozerye ($N = 55$), Ponemanye ($N = 46$). Komi are represented by two samples: Komi-Zyryans ($N = 43$), Komi-Izhemtsy ($N = 40$).

The collection of primary biological material (venous blood) from donors was carried out in compliance with the procedure of written informed consent for the study. For each donor, a questionnaire was drawn up with his genealogy, ethnicity, and places of birth of ancestors. The study included only DNA samples from donors who, according to the results of questioning, denied the fact of crossbreeding with representatives of other ethnic groups for at least three generations. An individual was assigned to a given ethnic group based on his own ethnic identity, his parents, and his place of birth.

Genotyping of samples for autosomal STR markers was performed by multiplex polymerase chain reaction with subsequent analysis of amplification products on an automatic genetic analyzer as described previously [10]. Experimental studies were carried out on the basis of the Center for Shared Use of Scientific Research Equipment Medical Genomics of the Tomsk National Research Medical Center and the Scientific Practical Center of the State Forensic Expertise Committee of the Republic of Belarus.

The correspondence of the distributions of genotypes to the Hardy–Weinberg equilibrium, genetic variability of STR loci, and interpopulation comparison were assessed using Arlequin v. 3.11 [15]. To analyze the genetic differentiation of populations, we used the calculation of pairwise values F_{st} and analysis of molecular variance (AMOVA), using the matrix of mean square differences in the number of repeats R_{st} as a matrix of genetic distances.

The discriminatory potential of the system of 21 STR markers was assessed using standard forensic indicators, namely, the probability of random coinci-

dence of genotypes (MP, matching probability), the probability of discrimination of unrelated individuals (PD, power of discrimination), excluding ability (PE, power of exclusion), and paternity index (PI) [16]. Allele frequencies and forensic parameters were calculated using the MS Excel software package.

RESULTS AND DISCUSSION

Genetic Variability of 21 STR

Allele frequencies of 21 STR show significant diversity in the studied populations. The most frequent alleles for most loci in all studied populations are the same. The largest total number of alleles for 21 autosomal markers was found in the Belarusian population (250), a slightly smaller number of alleles was shown in the Russian and Ukrainian populations (228 and 214, respectively), 204 alleles were found in the Moldovan population. In the population samples of the Gagauz (195), Mari (192), and Poles (184), less than 200 alleles are shown. The smallest number of alleles was found in the Komi-Izhma population (163).

After applying Bonferroni correction ($p = 0.002$), no significant deviations from the Hardy–Weinberg equilibrium were observed, except for four markers in three populations. The FGA marker showed deviations in the Russian population ($p = 0.00082 \pm 0.00002$). The Penta E marker showed deviations in the total sample of Belarusians ($p = 0.00175 \pm 0.00004$). In the subsample of Eastern Polesie, a deviation from equilibrium was shown at the D1S1656 locus ($p = 0.00095 \pm 0.00003$). Finally, D16S539 ($p = 0.00095 \pm 0.00003$) showed deviations in the sample of Komi-Izhemtsy.

Analysis of genetic distances for six samples of Belarusians did not show statistically significant genetic distance between them. Thus, for a given population, there is no need to take into account the territorial origin of the individual within the country in the identification analysis. Nevertheless, in further calculations, we present the results both for the total sample of Belarusians and for the territorial samples.

The opposite situation is shown for the Komi samples. Analysis of Nei's genetic distances revealed the presence of a significant genetic distance between the Komi-Zyryans and the Komi-Izhemtsy (0.0054, $p = 0.004$). These results provide a basis for considering these samples separately from each other in the formation of reference population data during genetic examination.

Also, significant genetic distances by 21 analyzed STR markers separate the populations of Komi-Zyryans, Komi-Izhemtsy, Mari, and Moldovans from each other and from the rest of the analyzed populations. On the contrary, there is no significant genetic distance between Russians, Belarusians, Poles, and Ukrainians (Table 1).

It should be noted that the marker with the highest number of alleles is the D12S391 locus, for which 18 allelic variants in the Belarusian population and 15 alleles in the Polish population have been shown. In the rest of the analyzed populations, the largest number of alleles is noted for the Penta E locus. Allele frequency data are available upon request from the authors.

Genetic Differentiation of Populations

The assessment of the genetic differentiation of the studied samples for the totality of all STR markers was carried out using the analysis of molecular variance. When comparing the entire array of samples, the differences between the populations were 0.35%. In the case of combining samples to the level of ethnic groups, differentiation at the level of ethnic groups increased to 0.39%. When comparing two samples of Russians and Belarusians, there were no differences at all. Comparison of the populations of Russians with the Komi and Mari showed a small negative level of genetic differentiation (0.03%), indicating a slightly greater genetic similarity of individuals from different populations as compared to individuals from the same population. The differentiation of the populations of Russians, Moldavians, and Gagauz was 0.13%.

When analyzing allele frequencies for all populations from the Republic of Belarus (all samples of Belarusians, Russians, Poles, and Ukrainians), the proportion of differences between them was 0.1%. When comparing samples of Belarusians from different historical and ethnographic regions of the republic, the level of differentiation between them was 0.1%. When comparing the group of Belarusians with Russians, Poles, and Ukrainians, the difference reached 0.8%. These results are quite expected owing to the multi-allelicity of STR markers and their relatively high degree of mutation and recombination. It is also natural to increase the indicators of genetic differentiation when comparing ethnic groups of different origins. The results of analysis of variance show that there are no significant genetic differences between Russians and Belarusians, and Slavic populations in general, as well as a very low level of differences between Russians, Komi, and Mari in STR genotypes. This means that, despite the available information on significant anthropological differentiation and existing modern differences between the studied ethnoterritorial groups, part of their gene pool for autosomal STR markers is very poorly differentiated.

Genetic Relationships between Populations

Visualization of the matrix of genetic distances between population samples by the method of multidimensional scaling (Fig. 1) shows that the studied samples are distributed in accordance with their ethnic proximity and geographic location. The genetically closest to each other are samples of Russians from Russia and Belarus. The most geographically dispersed samples of Komi-Izhemtsy and Poles are the most distant from each other. The Komi-Zyryans and the Komi-Izhemtsy show significant differences in the first dimension, which is probably related to a relatively strong shift in the allele frequencies in the sample of Komi-Izhemtsy because of their relatively recent settlement on this territory and the presence of the founder effect in this subethnic group. In accordance with the geographic location, the samples of Belarusians from the Eastern and Western Polesie, Central Belarus, the Podneprovye, and Ponemanye are also close to each other. Only the Poozerye sample is knocked out, which is the most northern, but at the same time is grouped with the southern samples from Polesie. The analysis highlights the genetic affinity between Moldovans and Gagauz.

Identification Potential of the System of Autosomal STR Loci

To assess the possibility of using the marker system for DNA identification in forensic examination, the standard forensic parameters of informativeness were calculated: the probability of random coincidence of genotypes (MP, matching probability), the probability of discrimination of unrelated individuals (PD, power of discrimination), excluding ability (PE, power of exclusion), paternity index (PI, paternity index). Indicators MP and PD are used for DNA identification of a person, and values for PE and PI are calculated when determining paternity.

The forensic parameters for each locus are presented in the supplementary materials. The greatest information content of the extended set of markers for DNA identification was shown for the Komi samples (Komi-Zyryans, Komi-Izhemtsy) (Table 2). In the Russian population, the value of the probability of discrimination of unrelated individuals (PD) increased to 0.999999999994978, and the index of excluding ability (PE) changed from 0.99999999989 to 0.999999998679. The lowest identifying informational content of a set of 21 STRs is shown for the sample of Poles. The indicators of informativeness for identification and determination of paternity in all populations are several orders of magnitude higher than the values established by the regulations in force on the territory of the Russian Federation. Data on the allele frequencies of 21 STR loci in the populations studied in this work are included in the database developed by the authors. The presented frequencies can be reference (for the corresponding population or ethnic group) for a probabilistic and statistical assessment of the results of an expert study in identifying a person, establishing kinship, etc. In addition, the data can be used for population genetic comparative studies.

Thus, the frequencies of alleles and forensic parameters of 21 autosomal STR loci were established

Table 1. Nei's genetic distances

|	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Russians (1)	*													
Gagauz (2)	0.0010 (0.125)	*												
Komi-Izhemtsy (3)	0.0075 (0.000)	0.0103 (0.000)	*											
Komi-Zyryans (4)	0.0020 (0.042)	0.0051 (0.001)	0.0054 (0.004)	*										
Mari (5)	0.0156 (0.000)	0.0186 (0.000)	0.0254 (0.000)	0.0148 (0.000)	*									
Moldovans (6)	0.0017 (0.002)	0.0008 (0.182)	0.0107 (0.000)	0.0032 (0.004)	0.0169 (0.000)	*								
BELvp (7)	–0.0001 (0.505)	0.0002 (0.378)	0.0095 (0.000)	0.0018 (0.118)	0.0168 (0.000)	–0.0006 (0.594)	*							
BELzp (8)	0.0016 (0.030)	0.0029 (0.025)	0.0097 (0.000)	0.0027 (0.044)	0.0189 (0.000)	0.0032 (0.001)	–0.0017 (0.860)	*						
BELp (9)	0.0003 (0.363)	0.0002 (0.406)	0.0073 (0.000)	0.0014 (0.211)	0.0154 (0.000)	0.0008 (0.190)	–0.0007 (0.637)	–0.0011 (0.804)	*					
BELpd (10)	0.0007 (0.149)	0.0017 (0.052)	0.0099 (0.000)	0.0042 (0.003)	0.0215 (0.000)	0.0015 (0.023)	–0.0005 (0.628)	–0.0001 (0.537)	0.0003 (0.404)	*				
BELpz (11)	0.0001 (0.453)	0.0020 (0.088)	0.0093 (0.001)	0.0031 (0.037)	0.0159 (0.000)	–0.0001 (0.464)	–0.0009 (0.698)	0.0008 (0.269)	0.0012 (0.218)	0.0002 (0.460)	*			
BELc (12)	–0.0004 (0.811)	0.0017 (0.040)	0.0098 (0.000)	0.0023 (0.033)	0.0165 (0.000)	0.0017 (0.004)	0.0004 (0.337)	0.0016 (0.057)	–0.0000 (0.502)	0.0002 (0.418)	–0.0002 (0.603)	*		
Poles (13)	0.0006 (0.289)	0.0034 (0.028)	0.0075 (0.001)	0.0050 (0.011)	0.0205 (0.000)	0.0026 (0.035)	–0.0011 (0.627)	0.0002 (0.408)	–0.0010 (0.663)	0.0001 (0.462)	–0.0012 (0.748)	–0.0004 (0.583)	*	
Belarusians (14)	0.0001 (0.327)	0.0012 (0.065)	0.0092 (0.000)	0.0027 (0.006)	0.0173 (0.000)	0.0011 (0.007)	–0.0017 (0.982)	–0.0010 (0.928)	–0.0013 (0.950)	–0.0012 (0.993)	–0.0011 (0.953)	–0.0010 (0.998)	–0.0005 (0.659)	*
Ukrainians (15)	–0.0006 (0.845)	0.0019 (0.030)	0.0077 (0.000)	0.0020 (0.044)	0.0188 (0.000)	0.0011 (0.030)	–0.0000 (0.439)	–0.0001 (0.509)	–0.0006 (0.668)	–0.0002 (0.543)	–0.0006 (0.727)	–0.0007 (0.921)	0.0002 (0.393)	–0.0005 (0.901)
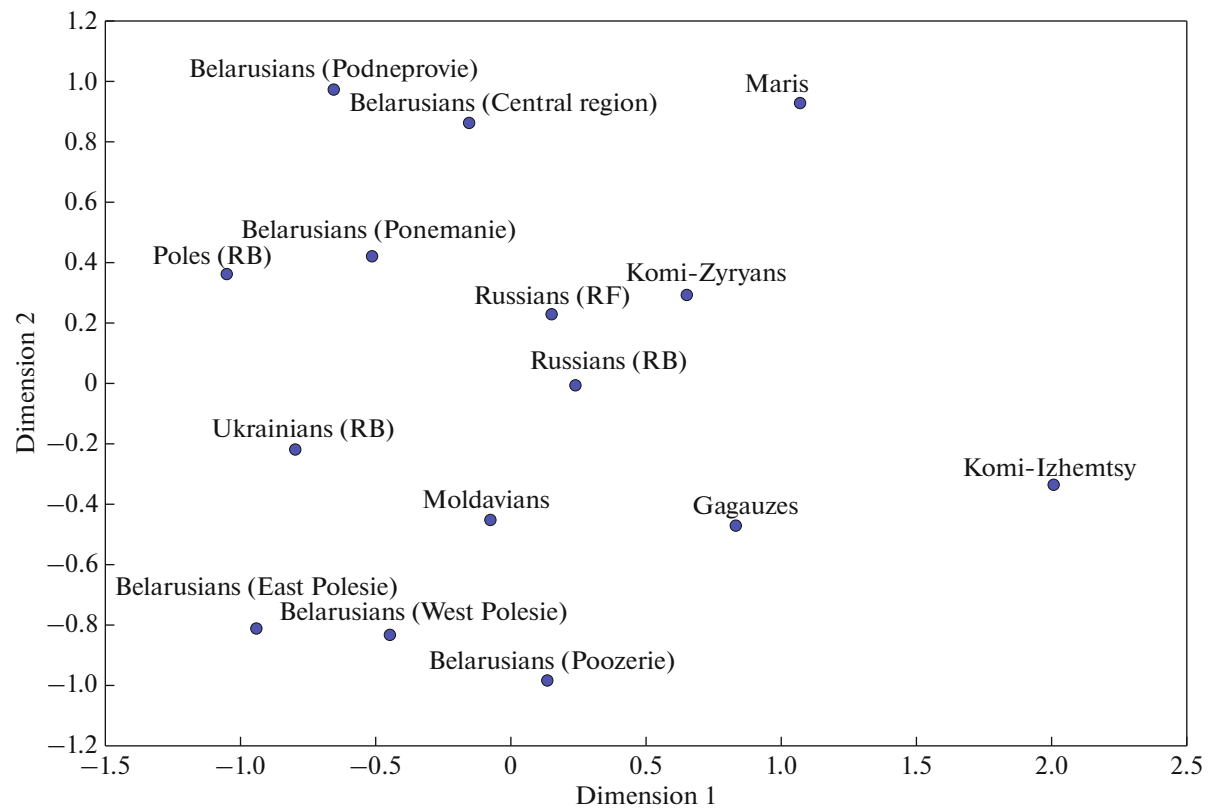
Fig. 1. Position of the studied populations in the MDS space.

for 15 populations of eight peoples of Eastern Europe. The significant discriminatory potential of the used kit for DNA identification and its effectiveness in determining kinship are shown. Assessment of the genetic differentiation of different populations revealed the need to take into account the population affiliation when creating databases of genotypes for comparing the reference allele frequencies.

Table 2. Identification informativeness of a set of 21 autosomal STR loci in populations of Eastern Europe

Population	PD	PE
Gagauz	0.999999999954539	0.999999997608839
Komi-Zyryans	1.0	0.999999999658466
Komi-Izhemtsy	1.0	0.999999999137624
Mari	0.999999999972582	0.999999998174319
Moldovans	0.999999999994137	0.999999999509155
Poles	0.999999999789335	0.999999999974522
Russians	0.999999999994978	0.999999998679004
Ukrainians	0.999999999996165	0.999999999893009
Belarusians	0.999999999997410	0.999999999263556
Eastern Polesie	0.999999999952167	0.999999999989964
Dnieper	0.999999999952167	0.999999999133377
Central	0.999999999996428	0.999999999233081
Lake Plateau	0.999999999969731	0.999999998645556
Neman	0.999999999961487	0.999999999765122
Western Polesie	0.999999999980485	0.999999999810011

FUNDING

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest. The authors declare that they have no conflict of interest.

Statement of compliance with standards of research involving humans as subjects. All procedures performed in a study involving people comply with the ethical standards of the institutional and/or national committee for research ethics and the 1964 Helsinki Declaration and its subsequent changes or comparable ethical standards.

Informed voluntary consent was obtained from each of the participants.

SUPPLEMENTARY INFORMATION

The online version contains supplementary material available at https://doi.org/10.1134/S102279542112005X.

REFERENCES

1. Szczerkowska, Z., Kapińska, E., Wysocka, J., and Cybulska, L., Northern Polish population data and forensic usefulness of 15 autosomal STR loci, *Forensic Sci. Int.*, 2004, vol. 144, no. 1, pp. 69−71.
https://doi.org/10.1016/j.forsciint.2004.02.020

2. Jacewicz, R., Jedrzejczyk, M., Ludwikowska, M., and Berent, J., Population database on 15 autosomal STR loci in 1000 unrelated individuals from the Lodz region of Poland, *Forensic Sci. Int. Genet.*, 2008, vol. 2, no. 1, pp. e1−e3.
https://doi.org/10.1016/j.fsigen.2007.08.002

3. Malyarchuk, B.A., Woźniak, M., Czarny, J., et al., Variation of 15 autosomal microsatellite DNA loci in the Russian population, *Mol. Biol.* (Moscow), 2007, vol. 41, no. 1, pp. 1−4.
https://doi.org/10.1134/S0026893307010013

4. Piatek, J., Jacewicz, R., Ossowski, A., et al., Population genetics of 15 autosomal STR loci in the population of Pomorze Zachodnie (NW Poland), *Forensic Sci. Int. Genet.*, 2008, vol. 2, no. 3, pp. e41−e43.
https://doi.org/10.1016/j.fsigen.2007.09.002

5. Rebała, K., Wysocka, J., Kapińska, E., et al., Belarusian population genetic database for 15 autosomal STR loci, *Forensic Sci. Int.*, 2007, vol. 173, nos. 2−3, pp. 235−237.
https://doi.org/10.1016/j.forsciint.2007.02.002

6. Pepiński, W., Skawrońska, M., Niemcunowicz-Janica, A., et al., Polymorphism of 11 autosomal STRs in a population sample of Belarussian minority residing in the region of Podlasie (Northeastern Poland)—an extension of the STR core set, *Arch. Med. Sadowej Kryminol.*, 2008, vol. 58, nos. 2−3, pp. 65−68.

7. Zhivotovsky, L.A., Malyarchuk, B.A., Derenko, M.V., et al., Developing STR databases on structured populations: the native South Siberian population versus the Russian population, *Forensic Sci. Int. Genet.*, 2009, vol. 3, no. 4, pp. e111−e116.
https://doi.org/10.1016/j.fsigen.2008.08.001

8. Parys-Proszek, A., Kupiec, T., Wolańska-Nowak, P., and Branicki, W., Genetic variation of 15 autosomal STR loci in a population sample from Poland, *Leg. Med.* (Tokyo), 2010, vol. 12, no. 5, pp. 246−248.
https://doi.org/10.1016/j.legalmed.2010.05.002

9. Stepanov, V.A., Melnikov, A.V., Lash-Zavada, A.Y., et al., Genetic variability of 15 autosomal STR loci in Russian populations, *Leg. Med.* (Tokyo), 2010, vol. 12, no. 5, pp. 256−258.
https://doi.org/10.1016/j.legalmed.2010.05.006

10. Stepanov, V.A., Balanovsky, O.P., Melnikov, A.V., et al., Characteristics of populations of the Russian Federation over the panel of fifteen loci used for DNA identification and in forensic medical examination, *Acta Nat.*, 2011, vol. 3, no. 2, pp. 59−71.
https://doi.org/10.32607/20758251-2011-3-2-56-67

11. Soltyszewski, I., Pepinski, W., Wolanska-Nowak, P., et al., Polish population data on 15 autosomal STRs of AmpFlSTR NGM PCR kit, *Forensic Sci. Int. Genet.*, 2014, vol. 9, pp. 142−149.
https://doi.org/10.1016/j.fsigen.2013.12.002

12. Tsybovskii, I.S., Veremeichik, V.M., Kotova, S.A., et al., Developing forensic reference database by 18 autosomal STR for DNA identification in Republic of Belarus, *Russ. J. Genet.*, 2017, vol. 53, no. 2, pp. 275−284.
https://doi.org/10.1134/S1022795417020132

13. Benvisto, A., Messina, F., Finocchio, A., et al., A genetic portrait of the South-Eastern Carpathians based on autosomal short tandem repeats loci used in forensics, *Am. J. Hum. Biol.*, 2018, vol. 30, no. 5. e23139.
https://doi.org/10.1002/ajhb.23139

14. Rębała, K., Veselinović, I., Siváková, D., et al., Northern Slavs from Serbia do not show a founder effect at autosomal and Y-chromosomal STRs and retain their paternal genetic heritage, *Forensic Sci. Int. Genet.*, 2009, vol. 3, no. 4, pp. e133−e136.
https://doi.org/10.1016/j.fsigen.2008.11.001

15. Excoffier, L., Laval, G., and Schneider, S., Arlequin ver. 3.0: an integrated software package for population genetics data analysis, *Evol. Bioinf. Online*, 2005, vol. 1, pp. 47−50.

16. Desmarais, D., Zhong, Y., Chakraborty, R., et al., Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), *J. Forensic Sci.*, 1998, vol. 43, pp. 1046−1049.