

A. SEMIKHODSKII¹, N.A. KOZUB^{2,3}, I.A. SOZINOV²

¹ Medical Genomics Ltd, London, UK

E-mail: andrei.semikhodskii@medicalgenomics.co.uk

² Institute of Plant Protection, NAAS of Ukraine, Kyiv

³ Institute of Food Biotechnology and Genomics, NAS of Ukraine, Kyiv

GENETIC DATA ON 15 STR LOCI IN THE CAUCASIAN POPULATION OF THE RUSSIAN FEDERATION



Population genetic data for the 15 STR loci included in the AmpFISTR Identifiler kit (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, D5S818 and FGA) obtained from 1118 unrelated Caucasian individuals from the Russian Federation are presented. In addition, a number of forensically useful genetic parameters are reported.

© A. SEMIKHODSKII¹, N.A. KOZUB^{2,3}, I.A. SOZINOV²,
2012

ISSN 0564–3783. Цитология и генетика. 2012. № 6

Introduction. Since it was first used more than 25 years ago for human identity and relationship testing DNA has emerged as a powerful tool in both civil and criminal justice systems. Currently, there are many commercially available STR (short tandem repeat) multiplex kits used in forensic casework and other types of DNA analysis such as paternity testing. Various criminal justice systems use different STR multiplexes: SGM Plus was adopted in the UK, The Netherlands, Austria and many European countries, and New Zealand, and CODIS was adopted by the USA, Canada and several other countries. To harmonise STR loci used in forensic work and facilitate DNA data exchange between countries and police forces, Interpol has introduced the Interpol Standard Set of Loci (ISSOL) containing the seven loci (six informative loci, one optional gender marker) which are common between major forensic DNA profiling systems.

AmpF/STR Identifiler contains all markers which are included in the SGM Plus and ISSOL systems and is the method of choice for obtaining DNA evidence in both criminal and non-criminal legal practice in Russian Federation, Ukraine and several other countries. Although allelic frequencies for STRs included in this marker system have been published for many European Slavic populations [1–3], there is paucity of information about populations from former USSR countries. In this paper we present data on genetic variability of STR markers included in the AmpF/STR Identifiler kit (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, D5S818 and FGA) for Caucasian population of the Russian Federation.

Materials and Methods. *Population.* This research is part of a wider socio-economic study of paternity in the Russian Federation. DNA samples were collected using buccal swabs from 1118 unrelated Caucasian individuals from 47 different regions of Russia (Figure). In paternity duo tests, the genotype of the child was selected for the analysis in cases of non-excluded paternity; in exclusion cases the genotypes of the alleged father and the child were used. In paternity trio tests, genotypes of both parents were selected for the study. Alleged fathers were questioned prior to the test as to whether they are biologically related to the mother. Only cases where the alleged father and the mother were known to be unrelated were used for this study.

Allelic frequencies and forensic parameters of 15 STR loci in the Caucasian population of the Russian Federation

Allele	Locus														
	CSFIPO	D2S1338*	D3S1358*	D5S818	D7S820	D8S1179*	D13S317	D16S539*	D18S51*	D19S433*	D21S11*	FGA*	TH01*	TPOX	vWA
5	-	-	-	-	-	-	-	-	-	-	-	-	0.0009	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-	0.2229	0.0004	-
7	0.0009	-	-	0.0076	0.0167	0.0009	0.0004	0.0004	-	-	-	-	0.1484	0.0013	-
8	0.0013	-	-	0.0013	0.1509	0.0197	0.1261	0.0242	-	-	-	-	0.1063	0.5739	-
9	0.0382	-	-	0.0528	0.1536	0.0228	0.0803	0.1113	-	0.0004	-	-	0.1969	0.0824	-
9.1	-	-	-	-	0.0005	-	-	-	-	-	-	-	-	-	-
9.3	-	-	-	-	-	-	0.0004	-	-	-	-	-	0.3112	-	-
10	0.2776	-	-	0.0809	0.2936	0.0815	0.0552	0.0507	0.0094	-	-	-	0.0126	0.0676	-
10.3	-	-	-	-	-	-	-	-	-	-	-	-	0.0004	-	-
11	0.2817	-	0.0013	0.3211	0.2164	0.0721	0.3407	0.2819	0.0153	0.0054	-	-	0.0004	0.2498	-
11.1	-	-	-	-	0.0018	-	-	-	-	-	-	-	-	-	-
11.2	-	-	-	-	-	-	-	-	-	0.0004	-	-	-	-	-
12	0.3167	-	-	0.3502	0.1369	0.1620	0.2531	0.2980	0.0930	0.0960	-	-	-	0.0242	-
12.2	-	-	-	-	-	-	-	-	-	0.0013	-	-	-	-	-
13	0.0674	0.0004	0.0031	0.1691	0.0266	0.3111	0.0974	0.1943	0.1096	0.2511	-	-	-	0.0004	0.0031
13.2	-	-	-	-	-	-	-	-	-	0.0179	-	-	-	-	-
14	0.0148	-	0.1168	0.0143	0.0027	0.2099	0.0444	0.0346	0.1608	0.3417	-	-	-	-	0.0932
14.2	-	-	-	-	-	-	-	-	-	0.0193	-	-	-	-	-
15	0.0013	0.0013	0.2601	0.0027	0.0005	0.0927	0.0018	0.0045	0.1824	0.1498	-	-	-	-	0.1102
15.2	-	-	-	-	-	-	-	-	-	0.0422	-	-	-	-	-
16	-	0.0436	0.2645	-	-	0.0237	-	-	0.1491	0.0390	-	0.0005	-	-	0.1850
16.2	-	-	-	-	-	-	-	-	-	0.0215	-	-	-	-	-
17	-	0.1855	0.2180	-	-	0.0031	-	-	0.1262	0.0040	-	0.0005	-	-	0.2666
17.2	-	-	-	-	-	-	-	-	-	0.0054	-	-	-	-	-
18	-	0.0854	0.1267	-	-	0.0004	-	-	0.0746	-	-	0.0150	-	-	0.2348
18.2	-	-	-	-	-	-	-	-	-	0.0040	-	-	-	-	-
19	-	0.1321	0.0081	-	-	-	-	-	0.0364	-	-	0.0865	-	-	0.0936
20	-	0.1509	0.0009	-	-	-	-	-	0.0229	-	-	0.1444	-	-	0.0134
20.2	-	-	-	-	-	-	-	-	-	-	-	0.0009	-	-	-
21	-	0.0413	0.0004	-	-	-	-	-	0.0108	-	-	0.1771	-	-	-
21.2	-	-	-	-	-	-	-	-	-	-	-	0.0018	-	-	-
21.3	-	-	-	-	-	-	-	-	-	-	-	0.0005	-	-	-
22	0.0220	0.0220	-	-	-	-	-	-	0.0049	0.0004	-	0.1753	-	-	-
22.2	-	-	-	-	-	-	-	-	-	-	-	0.0091	-	-	-
22.3	-	-	-	-	-	-	-	-	-	-	-	0.0005	-	-	-
23	0.0889	0.0889	-	-	-	-	-	-	0.0040	-	-	0.1334	-	-	-
23.2	-	-	-	-	-	-	-	-	-	-	-	0.0068	-	-	-

24	-	0.0997	-	-	-	-	-	-	0.0004	-	-	0.1389	-	-	
24.2	-	-	-	-	-	-	-	-	-	-	0.0004	0.0023	-	-	
25	-	0.1208	-	-	-	-	-	-	-	-	0.0004	0.0811	-	-	
26	-	0.0225	-	-	-	-	-	-	-	-	0.0022	0.0209	-	-	
27	-	0.0022	-	-	-	-	-	-	-	-	0.0206	0.0041	-	-	
27.2	-	-	-	-	-	-	-	-	-	-	0.0004	-	-	-	
28	-	0.0022	-	-	-	-	-	-	-	-	0.1490	-	-	-	
28.2	-	-	-	-	-	-	-	-	-	-	0.0004	-	-	-	
29	-	-	-	-	-	-	-	-	-	-	0.1939	0.0005	-	-	
29.2	-	-	-	-	-	-	-	-	-	-	0.0009	-	-	-	
30	-	0.0009	-	-	-	-	-	-	-	-	0.2401	-	-	-	
30.2	-	-	-	-	-	-	-	-	-	-	0.0682	-	-	-	
30.3	-	-	-	-	-	-	-	-	-	-	0.0004	-	-	-	
31	-	-	-	-	-	-	-	-	-	-	0.0655	-	-	-	
31.2	-	-	-	-	-	-	-	-	-	-	0.0907	-	-	-	
32	-	-	-	-	-	-	-	-	-	-	0.0103	-	-	-	
32.2	-	-	-	-	-	-	-	-	-	-	0.1136	-	-	-	
33	-	-	-	-	-	-	-	-	-	-	0.0013	-	-	-	
33.2	-	-	-	-	-	-	-	-	-	-	0.0368	-	-	-	
33.3	-	-	-	-	-	-	-	-	-	-	0.0004	-	-	-	
34.2	-	-	-	-	-	-	-	-	-	-	0.0040	-	-	-	
N_A	9	16	10	9	11	12	10	9	15	17	20	20	9	8	8
N	1113	1113	1117	1118	1107	1117	1114	1114	1113	1115	1114	1098	1115	1117	1116
H_o	0.724	0.880	0.778	0.738	0.818	0.816	0.768	0.807	0.877	0.788	0.838	0.863	0.798	0.584	0.789
H_E	0.738	0.881	0.785	0.736	0.801	0.811	0.783	0.778	0.875	0.784	0.851	0.866	0.781	0.597	0.810
P	0.609	0.080	0.667	0.231	0.086	0.438	0.419	0.009	0.310	0.480	0.606	0.024	0.795	0.146	0.469
F_{is}	0.018	0.002	0.009	-0.002	-0.020	-0.006	0.019	-0.038	-0.003	-0.005	0.015	0.002	-0.020	0.021	0.027
PIC	0.690	0.869	0.751	0.693	0.772	0.787	0.754	0.743	0.861	0.756	0.834	0.850	0.747	0.546	0.783
DP	0.885	0.973	0.920	0.886	0.929	0.941	0.926	0.913	0.970	0.923	0.962	0.965	0.917	0.792	0.939
MP	0.115	0.027	0.080	0.114	0.071	0.059	0.074	0.087	0.030	0.077	0.038	0.035	0.083	0.208	0.061
TPI	1.813	4.153	2.252	1.908	2.740	2.711	2.159	2.591	4.062	2.362	3.094	3.660	2.478	1.201	2.364
PE	0.467	0.754	0.559	0.489	0.632	0.628	0.542	0.612	0.749	0.578	0.672	0.721	0.596	0.272	0.578

Indications. N_A – number of unique alleles; N – number of individuals; H_o – observed heterozygosity; H_E – expected heterozygosity; P – Fisher exact test probability; F_{is} – coefficient of inbreeding; PIC – polymorphism information content; DP – discrimination power; MP – match probability; TPI – typical percentage index; PE – power of exclusion. * – loci included in SGM+ system.



Regional distribution of 1118 DNA samples analysed in the present study: 1 – Moscow and Moscow Region (304 samples), 2 – St Petersburg and Leningrad Region (68 samples), 3 – Vologda Region (6 samples), 4 – Ivanovo Region (1 sample), 5 – Vladimir Region (2 samples), 6 – Smolensk Region (51 samples), 7 – Kaluga Region (2 samples), 8 – Bryansk Region (63 samples), 9 – Orel Region (2 samples), 10 – Kursk Region (1 sample), 11 – Tula Region (3 samples), 12 – Belgorod Region (26 samples), 13 – Rostov Region (155 samples), 14 – Krasnodar Area (3 samples), 15 – Stavropol Area (31 samples), 16 – Voronezh Region (26 samples), 17 – Republic of Northern Ossetia (52 samples), 18 – Samara Region (4 samples), 19 – Kaliningrad Region (3 samples), 20 – Republic of Karelia (2 samples), 21 – Nizhegorod Region (1 sample), 22 – Perm Region (1 sample), 23 – Pskov Region (2 samples), 24 – Arkhangelsk (1 sample), 25 – Tver Region (1 sample), 26 – Volgograd Region (6 samples), 27 – Republic of Tataria (2 samples), 28 – Republic of Chuvashiya (3 samples), 29 – Murmansk Region (6 samples), 30 – Republic of Bashkortostan (49 samples), 31 – Orenburg Region (9 samples), 32 – Republic of Komi (1 sample), 33 – Kurgan Region (88 samples), 34 – Sverdlovsk Region (91 samples), 35 – Omsk Region (2 samples), 36 – Tyumen' Region (9 samples), 37 – Novosibirsk Region (12 samples), 38 – Altai Area (4 samples), 39 – Yamalo-Nenetskii Autonomous Area (1 sample), 40 – Republic of Khakassiya (2 samples), 41 – Krasnoyarsk Area (3 samples), 42 – Republic of Yakutia Saha (4 samples), 43 – Kamchatka Region (5 samples), 44 – Sakhalin Region (5 samples), 45 – Irkutsk Region (3 samples), 46 – Chita Region (1 sample), 47 – Amur Region (1 sample)

Individuals, who submitted DNA samples for testing, were provided with a consent form for using genetic information obtained from analysis of their DNA for scientific purposes in anonymised fashion. Only DNA samples for which consent was obtained were used in the study.

DNA analysis. DNA from buccal swabs was extracted using the QIAamp® 96 system (QIAGEN GmbH) according to manufacturer's recommendations. Polymerase chain reactions were performed in a GeneAmp PCR System 9600 (Perkin-Elmer) with AmpF/STR Identifiler kit (Applied Biosystems) in a final volume of 25 µl using 1.0–2.5 ng of target DNA. The products of amplification were analysed on an ABI PRISM® 3700 (Applied Biosystems) instrument. Genotypes were assigned using the Genotyper v.3.7

software by comparison with reference allelic ladders (Applied Biosystems).

Statistical analysis. Allelic frequencies and forensic parameters were determined by the PowerStats v 1.2 spreadsheet [4]. Testing for deviations from Hardy-Weinberg equilibrium (HWE), computations of the coefficient of inbreeding *Fis* and the sub-population correction *Fst* were performed using GDA software v 1.1 [5]. The accepted level of significance for all statistical tests was 0.05.

Data available from the author by email upon request (andrei.semikhodskii@medicalgenomics.co.uk).

Results and Discussion. Allelic frequencies and forensic parameters are given in Table. Comparisons with previously published data for a Russian population [6] reveal big similarities in

allele frequency distribution. At the same time, alleles 11 and 21 at locus D3S1358, alleles 9.1, 11.1 and 15 at locus D7S820, alleles 7 and 18 at locus D8S1179, alleles 7, 9.3 and 15 at locus D13S317, allele 24 at locus D18S51, alleles 24.2, 25, 27.3, 28.2, 30.3 and 33.3 at locus D21S11, and alleles 16, 20.2, 21.3, 22.3, 24.2 and 29 at locus FGA which were not reported in the above study were discovered in our population. Three rare alleles at loci vWA (allele 21), D7S820 (allele 6) and D18S51 (allele 14.2) were not detected, probably due to differences in ethnicity between the two populations (there is no indication of whether the Russian population described in [6] was Caucasian, Asian or mixed, while the present study was done exclusively on Caucasians). Based on RxC contingency test (G. Carmody, Carleton University, Canada), no differences with the Byelorussian minority of Poland [2] or the Russian population [6] were found. At the same time, significant differences were detected at D21S11, FGA and vWA (all $P < 0.05$) when comparing the studied population with a Polish population [1].

Deviations from HWE were tested using Fisher exact- p method. Statistically significant deviations from HWE were found for D16S539 and FGA. This probably does not indicate population stratification as such a phenomenon was not observed previously in a larger study of the Russian population [6]. Statistically significant deviation from the HWE were detected for the combinations D3S1358/FGA, D7S820/D8S1179, D8S1179/D13S317 (all $P < 0.01$), D2S1338/vWA, D3S1358/D21S11, D7S820/D13S317, and D18S51/FGA (all $P < 0.05$). The total number of alleles observed for each system ranged from 8 (TPOX and vWA) to 20 (D21S11 and FGA). As expected, there is a positive correlation between the level of polymorphism and the power of discrimination (DP) with the most polymorphic loci being generally the most discriminating. The power of discrimination ranged from 0.792 in TPOX to 0.973 in D2S1338. The combined values of power of exclusion (PE) and match probability (MP) were 0.99955 and $2.4 \cdot 10^{-18}$ respectively. The combined value of MP for the loci included in the SGM+ system was $1.7 \cdot 10^{-13}$. The inbreeding coefficient F_{is} across all the 15 loci was 0.001.

The sub-population correction F_{st} was estimated across all the 47 populations. The average value of F_{st} computed across all loci and alleles was 0.001268 thus indicating very small differentiation between the populations. The upper and lower bounds (95 % confidence interval) were computed by bootstrapping and were found to be 0.009282 and -0.007788 respectively.

There is a lack of information of frequency of forensically important STR for the population of the Russian Federation. The data presented in this paper can be used in criminal legal practice for calculating random match probability as well as in civil practice to resolve paternity and relatedness disputes. Because the data are collected in various regions of the country from general population they better represent the general Caucasian population than databases compiled using samples from particular strata of society (e.g. convicted criminals, students or soldiers). Although the estimated value of F_{st} for the Russian Caucasian population is small (0.001268) some populations studied consisted of only several individuals which could have skewed the results. The effect of sampling on estimation of F_{st} is well known. Because of this forensic scientists tend to use relatively large values for F_{st} (typically 1–4 %) thus giving a conservative estimate of a match probability which usually tends to favour the defendant.

А. Семиходский, Н.А. Козуб, И.А. Созинов

ГЕНЕТИЧЕСКИЕ ДАННЫЕ ПО 15 STR ЛОКУСАМ У ЕВРОПЕОИДНОГО НАСЕЛЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Представлены популяционно-генетические данные для 15 STR локусов, входящих в систему AmpF/STR Identifier (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, D5S818 и FGA). Локусы получены на основании анализа 1118 неродственных лиц европеоидной расы, проживающих на территории Российской Федерации. На основании этих данных рассчитаны важные генетические параметры, которые используются при криминалистическом анализе ДНК.

А. Семіходський, Н.О. Козуб, І.О. Созінов

ГЕНЕТИЧНІ ДАНІ ЗА 15 STR ЛОКУСАМИ У ЄВРОПЕОЇДНОГО НАСЕЛЕННЯ РОСІЙСЬКОЇ ФЕДЕРАЦІЇ

Наведено популяційно-генетичні дані за 15 STR локусами, що входять в систему AmpF/STR Identifier.

filer (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, D5S818 та FGA). Локуси отримано на основі аналізу 1118 неспоріднених осіб європейської раси, що проживають на території Російської Федерації. На основі цих даних розраховано важливі генетичні параметри, які використовуються при криміналістичному аналізі ДНК.

REFERENCES

1. *Jacewicz R., Berent J., Prośniak A., Kadlubek M., Szram S.* Genetic diversity and forensic evaluation of 10 STR loci in Lodz region of Poland // *Forensic Sci. Int.* – 2003. – **137**. – P. 94–96.
2. *Pepinski W., Niemcunowicz-Janica A., Skawronska M., Koc-Zorawska E., Janica J., Soltyszewski I.* Allele distribution of 15 STR loci in a population sample of Byelorussian minority residing in the north-eastern Poland // *Foren. Sci. Int.* – 2004. – **139**, № 2/3. – P. 265–267.
3. *Veselinović I., Kubat M., Furač I., Škavić J., Martinović Klarić I., Tasić M.* Allele frequencies of the 15 AmpF ISTR Identifiler loci in the population of Vojvodina Province, Serbia and Montenegro // *Int. J. Legal. Med.* – 2004. – **116**. – P. 184–186.
4. *Tereba A.* Tools for Analysis of Population Statistics, Profiles in DNA 3. Promega Corporation. – 1999 (www.promega.com/geneticidtools/powerstats).
5. *Lewis P.O., Zaykin D.* GDA 1.1. – 2002 Available at <http://en.bio-soft.net/dna/gda.html>
6. *Kornienko I.A., Vodolzhansky D.I., Ivanov P.L.* Genetic variation of the nine Profile Plus loci in Russians // *Int. J. Legal. Med.* – 2002. – **116**. – P. 309–311.
7. *Lewis P.O., Zaykin D.* Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). – 2001. Free program distributed by the authors over the internet from <http://lewis.eeb.uconn.edu/lewishome/software.html>

Received 19.03.12