ISSN 1022-7954, Russian Journal of Genetics, 2016, Vol. 52, No. 4, pp. 433–435. © Pleiades Publishing, Inc., 2016.
Original Russian Text © V.A. Stepanov, K.V. Vagaitseva, V.N. Kharkov, A.A. Cherednichenko, L.I. Minaicheva, A.V. Bocharova, 2016, published in Genetika, 2016, Vol. 52, No. 4, pp. 493–496.

SHORT COMMUNICATIONS

Variability and Identification Power of 60 X-Chromosome in Two Native Siberian Populations

V. A. Stepanov[a, b], K. V. Vagaitseva[a, b], V. N. Kharkov[a, b], A. A. Cherednichenko[a], L. I. Minaicheva[a], and A. V. Bocharova[a]

[a] *Research Institute for Medical Genetics, Tomsk, 634050 Russia*

e-mail: vadim.stepanov@medgenetics.ru

[b] *National Research Tomsk State University, Tomsk, 634050 Russia*

Received July 23, 2015

Abstract—Genetic diversity of 60 X-chromosome single nucleotide polymorphisms (XSNPid panel) in populations of Siberian Tatars and Tuvinians is described. A close spectrum of allele frequencies and a low level of their genetic differentiation ($G_{st} = 0.021$) is revealed. High discriminating power of the XSNPid panel in populations under study is demonstrated. The random matching probability (MP) of multilocus genotypes in males is 1.12×10^{-18} in Siberian Tatars and 7.77×10^{-16} in Tuvans. In females, MP is several orders of magnitude lower: 1.51×10^{-25} in Siberian Tatars and 1.83×10^{-23} in Tuvinians.

Keywords: genetic diversity, Siberian populations, X-chromosome markers, DNA identification

DOI: 10.1134/S1022795416030157

X-chromosome human genetic markers are widely used in population and medical genetics and are also highly interesting to criminology and forensic medicine. In those fields, X-chromosome markers are mainly used for testing the biological relationship in complicated cases; however, when sufficiently informative, they also make it possible to solve common tasks of individual identification. Unlike autosomal markers for DNA identification, for which the United States and Europe have generally accepted standard sets of 13–20 microsatellite (STR) loci, there are no generally accepted standards for X-chromosome systems. With the advent of new technologies for genotyping and massive parallel sequencing, STR markers are gradually giving way to the more common and easier-to-genotype single nucleotide markers (SNPs) [1, 2]. To achieve the discriminating power comparable with STR, the number of SNP markers should be three to four times higher [3]. Panels of X-chromosome SNPs (XSNPs), which are created on the basis of conventional technologies (real-time PCR, capillary gel electrophoresis), do not reach the information capacity of the systems based on STR markers because of the small multiplexing capacity of these technologies [4–7]. Recently, we developed a panel of 66 unlinked XSNPs (XSNPid) in which genotyping is carried out within two multiplexes of 36 and 30 markers by MALDI-TOF mass spectrometry (matrix-assisted laser desorption/ionization with time-of-flight measurements) [8].

The purpose of this work is to describe the genetic diversity of the XSNPid panel in two Siberian native populations (Tuvinians and Siberian Tatars) and to estimate the power of discrimination of the panel for DNA identification using these populations as an example.

The sample of Tuvinians ($N = 95$) was collected in the city of Kyzyl, and the sample of Siberian Tatars ($N = 76$) was collected in the village of Eushta in Tomsk oblast. The compositions of both the samples included only unrelated individuals not interbred in three generations and who signed an informed consent for the anonymous use of their DNA samples for research. DNA was isolated from peripheral blood lymphocytes by the standard method. Genotyping was carried out by multiplex PCR and separating DNA molecules by mass by tandem mass spectrometry using a Sequenom MassARRAY 4 analyzer (Sequenom, United States) according to the protocols described previously [9]. Statistical analysis of the frequencies of genes and genotypes was carried out in the Arlequin and Haploview software packages [10, 11]. Identification characteristics of the system of X-linked single nucleotide markers were estimated using standard indices used in forensics and legal medicine as described previously [12–14].

Six of the 66 markers were excluded from statistical analysis of allele frequencies and forensic indicators. Three of them (rs2130835, rs4825220, and rs4830049) showed low efficiency of genotyping in at least one of

Indicators of genetic diversity and power of discrimination of the system of 60 XSNP markers in populations of Tuvinians and Siberian Tatars

Indicator	Tuvinians	Siberian Tatars
Average gene diversity (H_e)	0.4354	0.4808
Random matching probability of genotypes for females (MPf)	1.83×10^{-23}	1.51×10^{-25}
Random matching probability of genotypes for males (MPm)	7.77×10^{-16}	1.12×10^{-18}
Power of discrimination for females (PDf)	0.99999999999999999998	0.999999999999999999999998
Power of discrimination for males (PDm)	0.9999999999999992	0.9999999999999999998

the populations (call rate below 85%), whereas for the other 63 markers more than 98% of genotypes were read. One of the SNPs (rs225067) showed significant, taking into account the Bonferroni correction, deviation from Hardy–Weinberg equilibrium in females in the population of Siberian Tatars (the exact test significance level $P < 0.000001$). Pairwise analysis of linkage disequilibrium using the exact test detected four pairs of markers with repeated SNPs demonstrating significant deviation from independent inheritance (rs2404797–rs7888207 in Tuvinians; rs2404797–rs5934683, rs2411976–rs5937091, rs2411976–rs916208 in Siberian Tatars). Two markers (rs2404797 and rs2411976), one of which appeared in each pair of linked SNPs, were also removed from the dataset.

The list of markers, allele frequencies, genetic diversity indices for each of the markers, and locus by locus forensic figures are given in the appendix available online (http://www.medgenetics.ru/UserFile/File/Doc/Publications/2015/XSNP-Prilogenie-2015.pdf) or on request from the authors. The average values of gene diversity and total values of discriminating power are shown in the table.

In general, the two studied native Siberian populations have a close range of allele frequencies of 60 X-chromosome SNPs. Locus by locus difference in the allele frequencies vary from 0 to 38%, the average being around 10%. The proximity of the two populations in allele frequencies is reflected in the relatively low value of the coefficient of genetic differentiation ($G_{ST} = 0.021$). Moreover, the population of Siberian Tatars demonstrates significantly greater average genetic diversity ($H_e = 0.481$) compared to Tuvinians ($H_e = 0.435$, see table).

The possibility of using the studied system of markers for DNA identification was assessed by determining the standard population statistical indicators characterizing the discriminating power of the system of markers. The most important of these parameters is the power of discrimination (PD) of unrelated individuals, which is the probability of a mismatch of genotypes in unrelated individuals. Because of the transmission pattern of the X-chromosome, the discriminating power of X-chromosome markers is significantly different in males and females; therefore the power of discrimination is calculated separately for females (PDf) and males (PDm) (see table).

Values of the discriminating power of the studied system of markers for DNA identification indicate that the XSNPid panel shows a very high power of discrimination in the studied populations. The random matching probability (MP) of multilocus genotypes in unrelated males is 1.12×10^{-18} in Siberian Tatars and 7.77×10^{-16} in Tuvinians. For females, this figure is several orders of magnitude lower: 1.51×10^{-25} in Siberian Tatars and 1.83×10^{-23} in Tuvinians.

It should be noted that the achieved levels of discrimination of individuals for the panel of 60 X-linked SNPs for males is not lower than that of the currently used American standard CODIS (Combined DNA Index System) and the European standard ESS (European Standard Set), both of which are based on STR markers, and commercial kits for their typing. Previously, we have shown that the random matching probability of genotypes in the populations of Russia for the standard set of STR loci varies from 1×10^{-16} to 5.60×10^{-17} [14, 15]. For females, the power of discrimination of the XSNPid panel is several orders of magni-

tude higher than that of standard microsatellite systems.

Thus, in this paper, we described the genetic diversity of the two native Siberian populations for the panel of X-linked SNP markers, found a high level of genetic diversity in these populations, and showed a high discriminating power of the XSNPid panel for DNA identification.

ACKNOWLEDGMENTS

This work was supported by state contract no. 14.604.21.0019 from June 17, 2014, in the framework of the Federal Target Program "Research and Development on Priority Directions of the Scientific and Technological Complex of Russia for 2014–2020."

REFERENCES

1. Yang, Y., Xie, B., and Yan, Y., Application of next-generation sequencing technology in forensic science, *Genomics, Proteomics Bioinf.*, 2014, vol. 12, pp. 190–197.
2. Churchill, J.D., Chang, J., Ge, J., et al., Blind study evaluation illustrates utility of the Ion PGM system for use in human identity DNA typing, *Croat. Med. J.*, 2015, vol. 56, pp. 218–229.
3. Krawczak, M., Informativity assessment for biallelic single nucleotide polymorphisms, *Electrophoresis*, 1999, vol. 20, pp. 1676–1681.
4. Li, C., Zhang, S., Zhao, S., et al., Analysis of 14 highly informative SNP markers on X chromosome by TaqMan SNP genotypoing assay, *Forensic Sci. Int.: Genet.*, 2010, vol. 4, pp. e145–e148.
5. Tomas, C., Sanchez, J.J., Castro, J.A., et al., Forensic usefulness of a 25 X-chromosome single-nuleotide polymorphism set, *Transfusion*, 2010, vol. 50, pp. 2258–2265.
6. Pereira, V., Tomas, C., Amorim, A., et al., Study of 25 X-chromosome SNPs in Portugu ese, *Forensic Sci. Int.: Genet.*, 2011, vol. 5, no. 4, pp. 336–338.
7. Vagaitseva, K.V., Khar'kov, V.N., Cherpinskaya, K.V., et al., Genetic variability of X-linked STR markers in Siberian populations, *Mol. Biol.* (Moscow), 2015, vol. 49, no. 2, pp. 267–274.
8. Stepanov, V.A., Vagaitseva, K.V., and Kharkov, V.N., Multiplex X-SNP system for forensic genetics, in *Program and Abstracts: 9th ISABS Conference on Forensic, Anthropologic Genetics and Mayo Clinic Lectures in Individualized Medicine*, Bol, Island of Brac, 2015, p. 86.
9. Stepanov, V.A. and Trifonova, E.A., Multiplex SNP genotyping by MALDI-TOF mass spectrometry: frequencies of 56 immune response gene SNPs in human populations, *Mol. Biol.* (Moscow), 2013, vol. 47, no. 6, pp. 852–862.
10. 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.
11. Barrett, J.C., Fry, B., Maller, J., and Daly, M.J., Haploview: analysis and visualization of LD and haplotype maps, *Bioinformatics*, 2005, vol. 21, no. 2, pp. 263–265.
12. Stepanov, V.A., *Etnogenomika naseleniya Severnoi Evrazii* (Population Ethnogenomics of Northern Eurasia), Tomsk: Pechatnaya manufaktura, 2002.
13. Stepanov, V.A., Khar'kov, V.N., Trifonova, E.A., and Marusin, A.V., *Metody statisticheskogo analiza v populyatsionnoi i evolyutsionnoi genetike cheloveka: uchebno-metodicheskoe posobie* (Methods of Statistical Analysis in Human Population and Evolutionary Genetics: Study Guide), Tomsk: Pechatnaya manufaktura, 2014.
14. 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.
15. Stepanov, V.A., Melnikov, A.V., Lash-Zavada, A.Y., et al., Genetic variability of 15 autosomal STR loci in Russian populations, *Legal Med.*, 2010, vol. 12, no. 5, pp. 256–258.

Translated by K. Lazarev