



genotyping based on multi-locus PCR and MALDI-TOF mass spectrometry, and report forensic and population genetic features of the panel of X-linked SNPs.

## 2. Materials and methods

### 2.1. Populations and samples

A total of 475 male samples from 4 populations representing Siberia (Buryat and Khakas), North Asia (Khanty) and Central Asia (Kazakh) were examined. Geographic locations of populations are shown in Fig. 1. Buryat is Siberian native population speaking the language belonging to Mongolian branch of Altaic language family and living in the area east to Baikal Lake. Buryats ( $N = 120$ ) were collected in Kurumkan village, Buryat Republic, Russia and in Aginskoe village, Chita Oblast, Russia. Khakas are native southern Siberian people speaking the language of Turkic branch of Altaic family. Population sample of Khakas ( $N = 120$ ) was collected in Askiz town in Khakas Republic, Russia. Khanty sample ( $N = 115$ ) representing native North Asian population, speaking the language of Uralic language family and living in the upper reaches of the Ob River, was collected in Kazym and Russkinskoe villages, Khanty-Mansi Autonomous Okrug, Russia. Central Asian Turkic-speaking Kazakhs ( $N = 120$ ) were collected in Almaty, Republic of Kazakhstan. This study was approved by the Ethical Review Board at the Research Institute for Medical Genetics, Tomsk, Russia.

Data on 6 unmixed population of Eurasian origin from the HapMap project (<http://hapmap.ncbi.nlm.nih.gov/>) were used in comparative analyses: CEU, Utah residents with Northern and Western European ancestry; CHB, Han Chinese in Beijing, China; CHD, Chinese in Metropolitan Denver, Colorado; GIH, Gujarati Indians in Houston, Texas; JPT, Japanese in Tokyo, Japan; TSI, Toscani in Italia.

### 2.2. Selection of SNPs

X chromosome SNPs (XSNP) were selected from the NCBI database focusing on the uniform distribution along the chromosome, minimal distance between adjacent markers of 1 mln bp, absence of linkage disequilibrium in HapMap populations, and gene diversity higher than 0.4. About 140 SNPs selected from the database were subjected to multiplex primer design using Sequenom Assay

Design software available on-line from [www.sequenom.com](http://www.sequenom.com). Final marker set (XSNPid panel) consisted of 66 XSNP in 2 multiplexes with the maximal multiplex capacities available for the set of 140 initially selected markers. The composition of multiplex 1 (W1) included 36 SNPs, while the multiplex 2 (W2) consisted of 30 SNPs. The composition of multiplexes is shown in Table 1. Primer sequences for PCR and iPLEX reaction are given in the Supplementary table.

### 2.3. Genotyping

Samples were genotyped by multiplex PCR and matrix-assisted laser desorption/ionization mass spectrometry with time of flight measurement (MALDI-TOF) using Sequenom MassARRAY 4 analyzer.

In brief, multi-locus PCR was performed separately for each multiplex in 5  $\mu$ l reaction volume containing 5 ng of genomic DNA template and primer mix for the corresponding multiplex. The cycle conditions used were 3 min of initial denaturation at 94°, then 42 cycles of 94° for 20 s, 56° for 40 s, and 72° for 60 s, and a final extension step of 5 min at 72°.

Subsequent steps of the experiment – SAP-reaction, iPLEX-reaction, transfer of the samples to spectro-chip, sample ionization and analysis of the mass spectra was carried out as previously described [10]. Samples transfer to the spectro-chip and ionization of samples was carried out on Sequenom MassARRAY 4 analyzer. Real-time analysis of mass spectra, as well as primary processing and documentation of the results were done with the MassARRAY TYPER 4.0 software (Agena Bioscience, San Diego, USA).

For several samples per each SNP verification of genotypes was performed by Sanger sequencing with fluorescently labeled terminators using capillary gel electrophoresis genetic analyzer ABI PRISM 3730 (Life Technologies). No discrepancy between MALDI-TOF and direct sequencing genotypes was found.

In order to evaluate the effectiveness of the panel in forensic tests with restricted amount of biological samples, we tested the genotyping efficiency in a series of experiments with lower amount of template using 500 pg, 1, and 3 ng of DNA in multiplex PCR reaction. PCR was performed as described above separately for each multiplex in 5  $\mu$ l reaction volume containing various amount of genomic DNA.



Fig. 1. Geographic locations of populations.

**Table 1**  
Allele frequencies and discriminating power of 62 X-chromosome SNPs in 4 North Eurasian populations.

SNP	Position	Multiplex	Ancestral allele	Ancestral allele frequency			
				Buryat	Khakas	Khanty	Kazakh
rs2694742	3127322	w1	A	0.528	0.464	0.475	0.500
rs1405303	4120689	w1	C	0.716	0.623	0.309	0.591
rs4826682	5119276	w1	C	0.591	0.600	0.526	0.652
rs5962008	6325520	w1	A	0.557	0.594	0.585	0.656
rs2404797	8795378	w1	A	0.495	0.679	0.542	0.527
rs5934683	9751474	w1	T	0.767	0.595	0.606	0.667
rs7888207	11916455	w2	G	0.576	0.642	0.304	0.532
rs952076	13946956	w2	A	0.567	0.424	0.270	0.462
rs2317327	15407061	w2	A	0.566	0.346	0.702	0.576
rs4484871	22751065	w2	A	0.221	0.441	0.371	0.316
rs1351260	26948596	w1	C	0.863	0.660	0.723	0.731
rs1389433	28128729	w2	G	0.760	0.726	0.629	0.768
rs225067	29149024	w1	A	0.041	0.532	0.395	0.516
rs4454452	30985342	w2	A	0.569	0.481	0.477	0.432
rs3005641	34029930	w1	T	0.343	0.364	0.436	0.435
rs761913	37859510	w1	G	0.630	0.636	0.577	0.677
rs5963641	39216082	w2	G	0.602	0.536	0.527	0.547
rs5917990	40387891	w2	A	0.358	0.393	0.267	0.242
rs6609159	41573566	w1	T	0.278	0.333	0.215	0.308
rs205847	42791946	w1	G	0.544	0.541	0.563	0.587
rs766117	43816206	w2	G	0.432	0.417	0.568	0.432
rs5953326	49373567	w1	C	0.389	0.577	0.284	0.548
rs5915291	50379075	w2	G	0.673	0.660	0.824	0.674
rs4131729	51697194	w1	C	0.283	0.358	0.211	0.163
rs11799030	53056518	w1	C	0.299	0.350	0.253	0.344
rs4826609	54765913	w1	G	0.618	0.418	0.505	0.500
rs6624701	63749154	w1	G	0.840	0.712	0.840	0.750
rs471205	66238317	w1	A	0.340	0.255	0.196	0.261
rs5919529	67358208	w2	T	0.270	0.407	0.345	0.284
rs5937091	70745323	w1	G	0.356	0.412	0.413	0.283
rs2207739	75644692	w2	G	0.673	0.714	0.544	0.484
rs2411976	78383858	w1	A	0.475	0.613	0.602	0.663
rs5969528	81668873	w1	A	0.814	0.757	0.821	0.731
rs5922869	83049688	w2	T	0.648	0.648	0.696	0.426
rs5968597	84946832	w2	G	0.388	0.519	0.490	0.404
rs222108	86910110	w1	C	0.500	0.500	0.500	0.500
rs1474970	90394689	w2	T	0.485	0.550	0.467	0.505
rs5941047	91431385	w1	T	0.500	0.468	0.537	0.548
rs5949581	94756278	w2	G	0.612	0.632	0.386	0.656
rs5921682	100130437	w2	A	0.388	0.455	0.341	0.426
rs4898334	101387968	w1	G	0.181	0.333	0.379	0.292
rs5945770	102594936	w1	T	0.613	0.554	0.553	0.609
rs1285715	106308416	w2	T	0.552	0.616	0.355	0.568
rs5973840	107319029	w1	C	0.863	0.821	0.800	0.785
rs5974348	112218701	w1	T	0.302	0.309	0.635	0.376
rs7058109	113228559	w1	G	0.125	0.119	0.266	0.237
rs9329406	115740985	w1	G	0.547	0.425	0.458	0.473
rs217937	118511843	w1	T	0.673	0.528	0.619	0.828
rs5909923	121978166	w1	C	0.222	0.330	0.433	0.272
rs5977571	124496989	w2	A	0.457	0.574	0.393	0.505
rs916208	127941251	w2	A	0.720	0.750	0.527	0.705
rs926640	129462353	w2	G	0.552	0.595	0.652	0.484
rs2797125	130815979	w1	A	0.574	0.552	0.510	0.467
rs17391	131910697	w2	G	0.269	0.267	0.473	0.436
rs5977991	133429960	w2	A	0.417	0.505	0.418	0.379
rs5975695	135268469	w1	C	0.467	0.424	0.607	0.449
rs4825002	140430391	w2	G	0.519	0.546	0.688	0.558
rs2869922	141490609	w2	A	0.425	0.389	0.489	0.442
rs4825213	143435373	w2	T	0.573	0.536	0.494	0.537
rs1781486	144674621	w2	T	0.480	0.491	0.538	0.442
rs2504169	146358005	w2	T	0.583	0.557	0.663	0.568
rs614511	149537834	w2	G	0.495	0.657	0.584	0.474
Average gene diversity ( $H_e$ )				0.436	0.458	0.451	0.454
AMOVA $F_{ST}$				0.02267			
MPF				$3.37 \times 10^{-24}$	$1.82 \times 10^{-25}$	$5.47 \times 10^{-25}$	$3.40 \times 10^{-25}$
MPm				$2.08 \times 10^{-16}$	$1.81 \times 10^{-17}$	$4.99 \times 10^{-17}$	$3.30 \times 10^{-17}$

Note: chromosomal position is given according to the reference nucleotide sequence of the human genome assembly 38.2 (GRCh38.p2).

#### 2.4. Statistical analysis

Standard population and forensic genetic indexes (gene diversity,  $H_e$ ; random genotype matching probability, MP; power

of discrimination, PD) were calculated using in-house software. Exact test of linkage disequilibrium (LD) between pairs of SNPs using Markov chain random walk with 10,000 dememorization steps, as well as estimation of genetic structure by analysis of



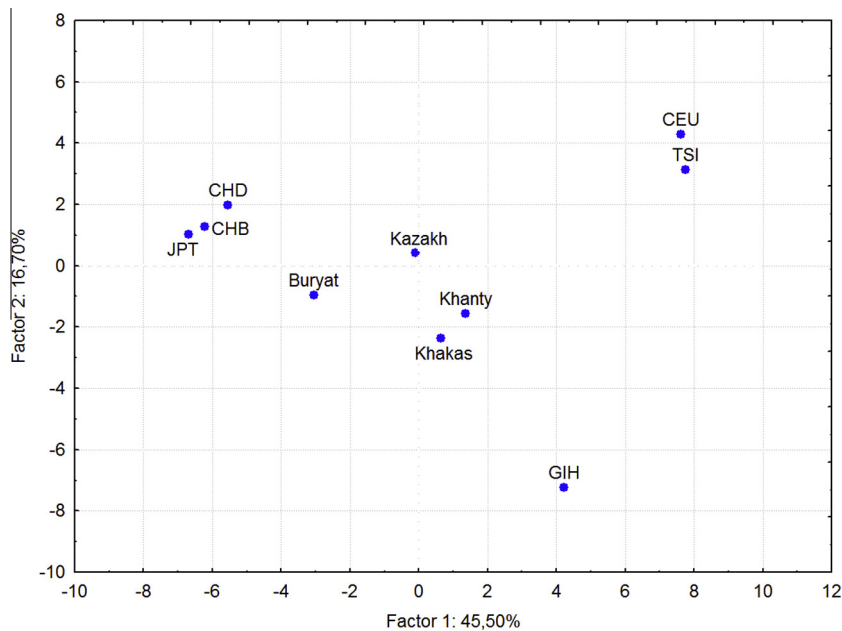


Fig. 2. Space of two first principal components of 62 XSNPs allele frequency variability.

**Table 2**  
Comparative characteristics of some DNA identification panels.

Panel	Type of markers	Number of markers*	Platform	Probability of identity (PI)	References
CODIS	aSTR	14/13	Fragment analysis/capillary gel-electrophoresis	$5.02 \times 10^{-16}$	[17]
ESS	aSTR	17/16	Fragment analysis/capillary gel-electrophoresis	$3.04 \times 10^{-16}$	[17]
PowerPlex 16	aSTR	16/15	Fragment analysis/capillary gel-electrophoresis	$2.81 \times 10^{-17}$	[15,17,19]
PowerPlex Fusion	aSTR	24/23	Fragment analysis/capillary gel-electrophoresis	$6.58 \times 10^{-29}$	[17]
SNPforID	aSNP	52	SBE/capillary gel-electrophoresis	$10^{-17}$ – $10^{-20}$	[20]
Genplex 49-plex	aSNP	49	Genplex/capillary gel-electrophoresis	$10^{-17}$ – $10^{-19}$	[21]
HID-Ion AmpliSeq Identity Panel	aSNP + YSNP	124/90	MPS	$1 \times 10^{-31}$ – $6 \times 10^{-35}$	[19]
ForenSeq DNA Signature Prep Kit	aSTR + YSTR + XSTR + SNP	230/94	MPS	n.d.	
53 XSNP	XSNP	67/53	MALDI-TOF MS	$6.9 \times 10^{-16}$ – $1 \times 10^{-19}$	[9]
XSNPId	XSNP	66/62	MALDI-TOF MS	$2.08 \times 10^{-16}$ – $1.82 \times 10^{-25}$	This work

Abbreviations: aSTR – autosomal short tandem repeats; MALDI-TOF MS – matrix-assisted laser desorption/ionization mass spectrometry with time of flight measurement; MPS – massive parallel sequencing; SBE – single base extension; XSTR – X chromosome STR; YSTR – Y chromosome STR; n.d. – no data.

\* Total number of markers/number of markers for identity testing. Some panels also include additional markers for sex determination, phenotypic characteristics and others.

ForenSeq DNA Signature Prep Kit for Illumina personal MPS machines. The cumulative matching probability for HID-Ion AmpliSeq Identity Panel varies in different populations in the range of  $1 \times 10^{-31}$ – $1 \times 10^{-33}$  [19] (Table 2).

Despite substantial progress in research of autosomal SNPs forensic panels, the niche for X-linked markers is still open. Early XSNPs panels were based on conventional typing technologies (real-time PCR, capillary gel electrophoresis) and did not reach the level of discriminating power of standard STR sets [6–8]. Among new developments the closest analog to our panel is the set of 52 XSNP markers, proposed recently by Li et al. [9] (Table 2). This panel is also based on MALDI-TOF mass spectrometry, however, has a smaller multiplexing capacity (17–18 markers per plex) and requires four separate multiplex PCR reactions and 4 iPLEX reactions, whereas our panel is assembled in two multiplexes and has power of discrimination in several orders of magnitude higher.

In conclusion, we report a promising multiplex set of X chromosome SNPs which can be a useful tool for population genetic studies as well as for forensic identity and kinship testing.

### Acknowledgements

This research was supported by State Contract No 14.604.21.0019 from the Federal Target Program “Research and development on priority directions of scientific-technological complex of Russia for 2014–2020”.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.legalmed.2015.12.008>.

## References

- [1] R. Szibor, X-chromosomal markers: past, present and future, *Forensic Sci. Int. Genet.* 1 (2007) 93–99.
- [2] P. Gill, D.J. Werrett, B. Budowle, R. Guerrieri, An assessment of whether SNPs will replace STRs in national DNA databases—joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDM), *Sci. Justice* 44 (2004) 51–53.
- [3] Y. Yang, B. Xie, J. Yan, Application of Next-generation Sequencing Technology in Forensic Science, *Genomics Proteomics Bioinf.* 12 (2014) 190–197.
- [4] J.D. Churchill, J. Chang, J. Ge, N. Rajagopalan, S.C. Wootton, C.W. Chang, R. Lagacé, W. Liao, J.L. King, B. Budowle, Blind study evaluation illustrates utility of the Ion PGM system for use in human identity DNA typing, *Croat. Med. J.* 56 (2015) 218–229.
- [5] M. Krawczak, Informativity assessment for biallelic single nucleotide polymorphisms, *Electrophoresis* 20 (1999) 1676–1681.
- [6] C. Li, S. Zhang, S. Zhao, Y. Liu, et al., Analysis of 14 highly informative SNP markers on X chromosome by TaqMan SNP genotyping assay, *Forensic Sci. Int. Genet.* 4 (2010) e145–e148.
- [7] C. Tomas, J.J. Sanchez, J.A. Castro, C. Borsting, N. Morling, Forensic usefulness of a 25 X-chromosome single-nucleotide polymorphism set, *Transfusion* 50 (2010) 2258–2265.
- [8] V. Pereira, C. Tomas, A. Amorim, N. Morling, L. Gusmão, M.J. Prata, Study of 25 X-chromosome SNPs in the Portuguese, *Forensic Sci. Int. Genet.* 5 (2010) 336–338.
- [9] Y. Li, Y. Liu, Y. Lin, Typing of 67 SNP Loci on X Chromosome by PCR and MALDI-TOF MS, *Res. Genet.* (2015) 9 374688.
- [10] V.A. Stepanov, E.A. Trifonova, Multiplex SNP genotyping by MALDI-TOF mass spectrometry: frequencies of 56 immune response gene SNPs in human populations, *Mol. Biol.* 47 (2013) 852–862.
- [11] V.N. Kharkov, K.V. Khamina, O.F. Medvedeva, K.V. Simonova, E.R. Eremina, V.A. Stepanov, Gene pool of Buryats: clinal variability and territorial subdivision based on data of Y-chromosome markers, *Rus. J. Genet.* 50 (2014) 180–190.
- [12] I.Y.U. Khitrinskaya, V.N. Kharkov, M.I. Voevoda, V.A. Stepanov, Genetic diversity and relationships of populations of Northern Eurasia by polymorphic Alu insertions, *Mol. Biol.* 48 (2014) 58–68.
- [13] I.Y.U. Khitrinskaya, V.A. Stepanov, V.P. Puzyrev, Analysis of the Alu polymorphism in the Buryat populations, *Rus. J. Genet.* 37 (2001) 1306–1311.
- [14] K.V. Vagaitseva, V.N. Kharkov, K.V. Cherpinskaya, I.Y.U. Khitrinskaya, V.A. Stepanov, Genetic variability of X-linked STR markers in Siberian populations, *Mol. Biol.* 49 (2015) 267–274.
- [15] V.A. Stepanov, O.P. Balanovsky, A.V. Melnikov, A.Y.U. Lash-Zavada, V.N. Kharkov, T.V. Tyazhelova, V.L. Akhmetova, O.V. Zhukova, Y.U.V. Shneider, I.N. Shilnikova, S.A. Borinskaya, A.V. Marusin, M.G. Spiridonova, K.V. Simonova, I.Y. U. Khitrinskaya, M.O. Radzhabov, A.G. Romanov, O.V. Shtygasheva, S.M. Koshel, E.V. Balanovskaya, A.V. Rybakova, E.K. Khusnutdinova, V.P. Puzyrev, Characterization of populations of the Russian Federation over the panel of fifteen loci used for DNA identification and in forensic medical examination, *Acta Naturae* 3 (2011) 56–67.
- [16] A.A. Cherednichenko, E.A. Trifonova, K.V. Vagaitseva, A.V. Bocharova, V.A. Stepanov, Association of the genetic polymorphism of cytokines and their receptors with climate and geographic factors in human populations, *Rus. J. Genet.* 50 (2014) 1112–1116.
- [17] PowerPlex® Fusion Systems, <<http://worldwide.promega.com/products/genetic-identity/str-analysis-for-forensic-and-paternity-testing/powerplex-fusion-str-kits/>>.
- [18] V.A. Stepanov, A.V. Melnikov, A.Y. Lash-Zavada, S.A. Borinskaya, T.V. Tyazhelova, O.V. Zhukova, Y.V. Schneider, I.N. Shil'nikova, V.P. Puzyrev, A.A. Rybakova, N.K. Yankovsky, Genetic variability of 15 autosomal STR loci in Russian populations, *Leg. Med. (Tokyo)* 12 (2010) 256–258.
- [19] Life Technologies, HID-Ion AmpliSeq™ Identity Panel, <<https://www.lifetechnologies.com/content/dam/LifeTech/global/applied-sciences/pdfs/human-identification/CO119694-HID-Ion-AmpliSeq-Identity-panel-Flyer-Americas-FLR.pdf>>.
- [20] J.J. Sanchez, C. Phillips, C. Borsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, A. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis* 27 (2006) 1713–1724.
- [21] C. Phillips, R. Fang, D. Ballard, M. Fondevila, C. Harrison, F. Hyland, E. Musgrave-Brown, C. Proff, E. Ramos-Luis, B. Sobrino, A. Carracedo, M.R. Furtado, D. Syndercombe Court, P.M. SchneiderSNPforID Consortium, Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel, *Forensic Sci. Int. Genet.* 1 (2007) 180–185.