

Взаимоотношения популяций мира по SNP маркерам X-хромосомы, входящим в состав тест-системы для ДНК-идентификации XSNPid

Вагайцева К.В.^{1,2*}, Харьков В.Н.^{1,2}, Степанов В.А.^{1,2}

1 — НИИ медицинской генетики, Томский НИМЦ, Томск; * kseniya.simonova@medgenetics.ru

2 — Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет»

В статье представлены результаты анализа генетических взаимоотношений популяций по панели 66 X-SNP маркеров X-хромосомы, входящих в состав тест-системы для ДНК-идентификации XSNPid, а также результаты оценки идентификационного потенциала данной системы для популяций мира. В исследование были включены такие популяции, как: русские, тувинцы, буряты, казахи, хакасы, сибирские татары, ханты, жители штата Юта центрально-европейского происхождения, итальянцы, мексиканцы, японцы, хань, индийцы, афроамериканцы. Анализ показал, что генетические отличия исследованных популяций демонстрируют географическую структурированность, при этом тест-система XSNPid является наиболее информативной для популяций, относящихся к европеоидному расовому типу.

Ключевые слова: популяционная генетика, ДНК-идентификация, X-хромосома, однонуклеотидные полиморфные маркеры, XSNPid

Relationships of world populations by SNP markers of the X chromosome are part of the test system for DNA-identification XSNPid

Vagaitseva K.V.^{1,2*}, Kharkov V.N.^{1,2}, Stepanov V.A.^{1,2}

1 — Research Institute of Medical Genetics, Tomsk NRMС, Tomsk, Russia; * kseniya.simonova@medgenetics.ru

2 — National Research Tomsk State University, Tomsk, Russia

This paper presents results of analysis of the genetic relationships of populations on panel 66 X-SNP markers of the X chromosome included in the test system for DNA-identification XSNPid, and the results of evaluation of identification potential of this system for the populations of the world. The study included populations such as Russians, Tuvinians, Buryats, Kazakhs, Khakas, Siberian Tatars, Khanty, Utah residents of Central European origin, Italians, Mexicans, Japanese, Han, Indians, African Americans. The analysis showed that the genetic differences of the studied populations show geographic structuring. The test system XSNPid is the most informative for populations belonging to the Caucasoid racial type.

Key words: population genetics, DNA identification, X-chromosome, single nucleotide polymorphism, XSNPid

Введение

За счет особенностей наследования, гемизиготности мужчин и низкой эффективной численности X-хромосомные маркеры являются информативным инструментом, как в популяционной генетике, так и в криминалистике [1, 2]. Одним из последних трендов ДНК-идентификации является разработка высокоинформативных систем основанных на SNP-маркерах [3–6]. Помимо того, что SNP-маркеры более продуктивны при работе с деградированной ДНК, легко мультиплексируются и генотипируются, низкий темп мутирования делает этот тип маркеров наиболее информативным для установления родства [7].

Ранее на основе данных о частотах встречаемости аллелей X-хромосомных маркеров в популяциях, представляющих различные коренные этносы населения России и ближнего зарубежья, нами была разработана тест-система для ДНК-идентификации, наиболее ин-

формативная для населения Российской Федерации [8–10].

В настоящей статье представлены результаты анализа генетических взаимоотношений мировых популяций по панели 66 X-SNP маркеров X-хромосомы входящих в состав разработанной тест-системы XSNPid и данные о информативности представленной тест-системы для мировых популяций.

Материалы и методы

В исследовании были использованы образцы ДНК индивидов относящихся к семи популяционным выборкам: русские, тувинцы, буряты, казахи, хакасы, сибирские татары, ханты. Все образцы являются частью банка ДНК НИИМГ г.Томска. Этническая принадлежность индивидов учитывалась до третьего поколения. Также для сравнительного анализа были привлечены данные о частотах аллелей из проекта HarMap: жители штата Юта

центрально-европейского происхождения (CEU), итальянцы (TSI), мексиканцы (MEX), японцы (JPT), китайцы (CHB, CHD), индийцы (GIN), афроамериканцы (ASW).

Генотипирование осуществлялось путем мультиплексной ПЦП с последующим анализом методом MALDI-TOF масс-спектрометрии на генетическом анализаторе MassARRAY Analyzer 4 (Sequenom, США) [8].

Генетические взаимоотношения популяций выявляли двумя методами: методом главных компонент и методом многомерного шкалирования. В качестве переменных, анализируемых методом главных компонент, использовали частоты аллелей. Подробное описание частот аллелей и генетического разнообразия популяций Российской Федерации и ближнего зарубежья представлено в ранних публикациях [8–10]. Для оценки идентификационной информативности использовали стандартные популяционно-статистические показатели, характеризующие идентификационный потенциал системы маркеров: вероятность дискриминации неродственных индивидов (PD, Power of Discrimination), исключающая способность (PE, Power of Exclusion), индекс отцовства (PI, Paternity Index) [11].

Результаты и обсуждение

Анализ генетических взаимоотношений пятнадцати мировых популяций по 66 X-SNP маркерам показал корреляцию генетических и географических параметров.

Метод главных компонент. Две первые главные компоненты объясняют 65,07% вариабельности аллельных частот в анализируемых популяциях (рис. 1). На долю первой главной компоненты приходится 40,61%, второй главной компоненты — 24,46%. Как показано на рисунке, первая компонента отделяет афроамериканцев. Вторая главная компонента, вероятно, ранжирует выборки по восточно-евразийской генетической составляющей. В самой верхней части графика расположены европеоидные популяции, в эту группу вошли русские, европейцы Соединенных Штатов Америки и итальянцы. В нижней части графика расположился кластер монголоидных

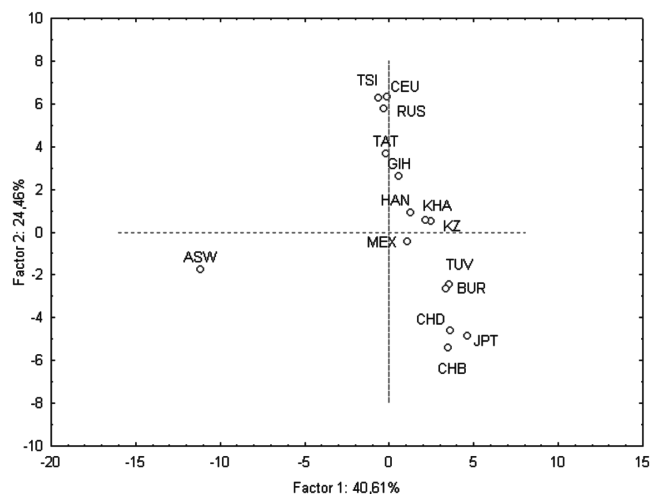


Рис. 1. Положение популяций в пространстве главных компонент аллельных частот SNP-маркеров:

здесь и далее: RUS — русские, CEU — жители штата Юта центрально-европейского происхождения, TSI — итальянцы, TAT — сибирские татары, HAN — ханты, KHA — хакасы, KZ — казахи, BUR — буряты, TUV — тувинцы, JPT — японцы, CHB — китайцы, MEX — мексиканцы, CHD — китайцы США, ASW — афроамериканцы, GIN — гуджаратские индийцы проживающие в США.

популяций: китайцы, китайцы проживающие на территории США, японцы. Если развернуть график на 90°, можно отметить согласованность расположения выборок в пространстве ГК с географической картой. При анализе корреляции главных компонент с географическими координатами первая главная компонента показала отрицательную корреляцию с долготой ($-0,821$, $r = 0,023$) и положительную с широтой ($0,786$, $r = 0,036$).

Многомерное шкалирование. Для расчета генетических расстояний между популяциями данные полученные из базы НарМар были исключены. На основе частот аллелей была построена матрица генетических расстояний по методу Нея [12] (табл. 1). Анализ полученной матрицы расстояний методом многомерного шкалирования представлен на рис. 2.

Таблица 1

Матрица генетических расстояний (Fst)

	Русские	Тувинцы	Буряты	Казахи	Хакасы	Ханты	Татары
Русские	0						
Тувинцы	0,074	0					
Буряты	0,077	0,015	0				
Казахи	0,034	0,021	0,019	0			
Хакасы	0,050	0,022	0,024	0,018	0		
Ханты	0,045	0,040	0,036	0,028	0,037	0	
Татары	0,013	0,049	0,052	0,025	0,031	0,033	0

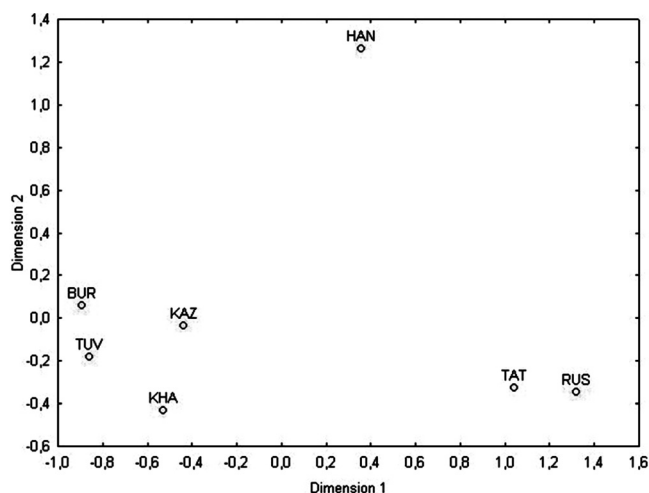


Рис. 2. Положение популяций в пространстве двух первых размерностей многомерного шкалирования матрицы генетических расстояний по частотам аллелей X-SNP маркеров.

Согласно графику, исследуемые выборки можно разделить на три группы наиболее близких популяций. Наиболее многочисленная группа представлена выборками алтайской лингвистической семьи: бурятами, тувинцами, казахами, хакасами. Популяция хантов оказалась сильно дистанцирована от основного пула популяций. Третью группу образовали выборки татар Томска и русских, генетическая близость этих популяций также была показана и по микросателлитным маркерам X-хромосомы [13]. При анализе корреляции двух первых размерностей с географическими координатами, значимые значения корреляции были показаны для первой размерности и долготы ($-0,893$, $p = 0,007$).

Аналогичная географическая структурированность популяций наблюдалась и по другим системам нейтральных генетических маркеров [14–16], что свидетельствует о том, что представленная панель X-хромосомных SNP является адекватным инструментом для популяционно-генетических исследований.

В результате проведенного корреляционного анализ матриц отражающих генетические расстояния между популяциями с географическими расстояниями была выявлена значимая корреляция матрицы генетических расстояний по частотам X-SNP маркеров с географическими ($0,529$, $p = 0,020$). Доля вариативности частот аллелей, объясняемая географией составила 28%.

Идентификационный потенциал. Расчеты криминалистических показателей тест-системы проводились без учета сцепления, поскольку в предыдущих работах по системе XSPid сцепление было обнаружено лишь в единичных случаях, а также по причине отсутствия данных о многоаллельных индивидуальных генотипах для некоторых из мировых популяций. Показатель вероятности дискриминации неродственных индивидов (PD) отражает информативность панели маркеров в идентификации индивида (табл. 2). Информативность системы маркеров для тестирования родства и в частности отцовства, отражают показатели исключательной способности (PE) (табл. 2). Как видно из таблицы, система 66 SNP-маркеров показала достаточную информативность для популяций, относящихся к европеоидному и монголоидному расовым типам, и низкую информативность для афроамериканцев. При этом наибольшая информативность показана для европеоидов.

Полученные результаты свидетельствуют о том, что рассматриваемая в данной статье тест-система XSNPid является новым информативным инструментом, как для ДНК-идентификации, так и для популяционной генетики.

Авторы подтвердили отсутствие конфликта интересов.

Работа получила финансовую поддержку Госконтракта № 14.604.21.0019 от 17 июня 2014 года в рамках ФЦП "Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы".

Таблица 2

Криминалистические показатели тест-системы XSNPid для мировых популяций

	RUS	KZ	TUV	KHA	HAN	BUR	TAT	
PDf	$1-8,6 \cdot 10^{-27}$	$1-8,9 \cdot 10^{-27}$	$1-4,8 \cdot 10^{-25}$	$1-1,7 \cdot 10^{-26}$	$1-1,3 \cdot 10^{-26}$	$1-8,6 \cdot 10^{-26}$	$1-1,3 \cdot 10^{-27}$	
PDm	$1-2,1 \cdot 10^{-18}$	$1-2,9 \cdot 10^{-18}$	$1-3,4 \cdot 10^{-17}$	$1-4,2 \cdot 10^{-18}$	$1-3,8 \cdot 10^{-18}$	$1-1,8 \cdot 10^{-17}$	$1-3,2 \cdot 10^{-19}$	
PE	0,999989	0,999986	0,999956	0,999985	0,999985	0,999972	0,999995	
	CEU	TSI	MEX	JPT	CHB	CHD	GIH	ASW
PDf	$1-6,5 \cdot 10^{-27}$	$1-5,7 \cdot 10^{-27}$	$1-1,9 \cdot 10^{-25}$	$1-5 \cdot 10^{-24}$	$1-1,9 \cdot 10^{-24}$	$1-7,7 \cdot 10^{-25}$	$1-2,7 \cdot 10^{-26}$	$1-3,4 \cdot 10^{-22}$
PDm	$1-1,9 \cdot 10^{-18}$	$1-1,3 \cdot 10^{-18}$	$1-4,4 \cdot 10^{-17}$	$1-3,6 \cdot 10^{-16}$	$1-1,7 \cdot 10^{-16}$	$1-8,2 \cdot 10^{-17}$	$1-8,8 \cdot 10^{-18}$	$1-1,6 \cdot 10^{-14}$
PE	0,999989	0,999991	0,999957	0,999925	0,999942	0,999956	0,999977	0,999662
Примечание. PDf, PDm — вероятность дискриминации неродственных индивидов для женщин и мужчин соответственно, PE — исключаяющая способность								

References

1. Schaffner SF. The X chromosome in population genetics. *Nat Rev Genet.* 2004; 5(1):43-51
2. Szibor R. X-chromosomal markers: past, present and future. *Forensic Sci Int Genet.* 2007; 1(2):93-99
3. Petkovski E, Keyser-Tracqui C, Hienne R et al. SNPs and MALDI-TOF MS: tools for DNA typing in forensic paternity testing and anthropology. *J. Forensic. Sci.* 2005; 50(3):535-541.
4. Lin Y, Li L, Liu Y et al. SNP typing of formalin-fixed tissue in individual identification. *Fa. Yi. Xue. Za. Zhi.* 2010; 26(6):446-448.
5. Li L, Liu Y, Lin Y et al. Genotyping and linkage disequilibrium analysis of 67 SNP loci on X chromosome. *Fa. Yi. Xue. Za. Zhi.* 2011; 27(5):337-341.
6. Li Y, Liu Y, Lin Y. Typing of 67 SNP Loci on X Chromosome by PCR and MALDI-TOF MS. *Research in Genetics.* 2015; Article ID 374688: 9.
7. Butler JM, Coble MD, Vallone PM. STRs vs SNP: thoughts on the future of forensic DNA testing. *Forensic. Sci. Med. Pathol.* 2007; 3(3):200-205.
8. Степанов ВА, Вагайцева КВ, Харьков ВН и др. Панель однонуклеотидных, сцепленных с X-хромосомой, полиморфных маркеров для ДНК-идентификации (XSNPid) на основе мультиплексного генотипирования с использованием методов многолокусной ПЦР и масс-спектрометрии MALDI-TOF. *Молекулярная биология.* 2016; 50(3):1-12.
9. Степанов ВА, Вагайцева КВ, Харьков ВН и др. Вариативность и идентификационный потенциал 60 X-хромосомных SNPs в двух популяциях коренного населения Сибири. *Генетика.* 2016; 52(4):493-496.
10. Stepanov V, Vagaitseva K, Kharkov V, et al. Forensic and population genetic characteristics of 62 X chromosome SNPs revealed by multiplex PCR and MALDI-TOF mass spectrometry genotyping in 4 North Eurasian populations. *Legal Medicine.* 2016; 18:66-71.
11. Степанов В.А., Харьков В.Н., Трифонова Е.А., Марусин А.В. Методы статистического анализа в популяционной и эволюционной генетике человека. Томск. 2014. 99с.
12. Nei M. *Molecular evolutionary genetics.* New York: Columbia Univ. Press. 1987. 512 p.
13. Вагайцева КВ., Харьков ВН., Черпинская КВ и др. Генетическая вариативность X-сцепленных STR-маркеров в популяциях Сибири. *Молекулярная биология.* 2015; 49(2): 305-312.
14. Степанов ВА. Геномы, популяции, болезни: этническая геномика и персонализированная медицина. *Acta Naturae.* 2010; 2(4):18-34.
15. Elhaik E, Tatarinova T, Chebotarev et al. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun.* 2014; 5:3513.
16. Yunusbayev B, Metspalu M, Jarve M, et al. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol Biol Evol.* 2012; 29(1): 359-65.