FOOTPRINTS OF INTERACTION AMONG FINNIC-SPEAKING, SLAVIC, AND TURKIC-SPEAKING POPULATIONS IN MODERN GENE POOL AND THEIR REFLECTION IN PHARMACOGENETICS

Balanovska EV[1,2,3] ✉, Gorin IO[1], Ponomarev GY[2], Pylev VY[1,3], Petrushenko VS[1], Markina NV[2], Mamaeva AD[2], Larin AK[4], Agdzhoyan AT[2]

[1] Bochkov Research Centre of Medical Genetics, Moscow, Russia

[2] Vavilov Institute of General Genetics, Moscow, Russia

[3] Biobank of North Eurasia, Moscow, Russia

[4] Federal Research and Clinical Center of Physical-Chemical Medicine, Moscow, Russia

Genetic contribution of pre-Slavic populations to gene pools of modern Russia is increasingly relevant, along with genetic footprints of the Golden Horde invasion. The novel genome-wide approaches enable advanced solutions in this field. The study aimed at searching for the footprints of genetic interaction among Finnic-speaking, Slavic and Turkic-speaking populations of Central Russia and Volga Region and their reflection in pharmacogenetic landscape. Modeling ancestral components by ADMIXTURE software and their mapping involved genome-wide genotyping data for 248 individual genomes representing 47 populations of 9 ethnic groups. Of specific ancestral components identified in each of the Finnic-speaking peoples, only Mordovian ancestral components are common for all populations within the studied geographic area, regardless of their linguistic affiliation. Gene pools of Russian populations include 80% of intrinsic component, 19% contribution from Finnic-speaking peoples, and 1% of Central Asian influence. The Tatar gene pool combines all identified ancestral components, including 81% contribution from Finnic-speaking peoples and only 12% of Central Asian influence, which prevents using it as a reference for the assessment of Golden Horde footprints in Russian gene pools. A map of genetic distances from Ryazan Russians based on a panel of 42 pharmacogenetic markers reveals a landscape strikingly independent from the selectively neutral ancestral genomic patterns. For instance, populations of Mordovia, Kaluga, Smolensk, and Kostroma regions are the closest to Ryazan Russians in pharmacogenetic status, whereas populations of Ryazan and Nizhny Novgorod regions have strikingly divergent pharmacogenetic status despite the similarity of the selectively neutral ancestral genomic patterns. These findings confirm the relevance of targeted pharmacogenetic characterization for gene pools of Russia.

Keywords: gene pool, gene geography, ancestral component, ADMIXTURE, genome-wide panel, pharmacogenetics, European part of Russia

СЛЕДЫ ВЗАИМОДЕЙСТВИЯ ФИННОЯЗЫЧНОГО, СЛАВЯНСКОГО И ТЮРКОЯЗЫЧНОГО НАСЕЛЕНИЯ В СОВРЕМЕННОМ ГЕНОФОНДЕ И ИХ ОТРАЖЕНИЕ В ФАРМАКОГЕНЕТИКЕ

Е. В. Балановская[1,2,3] ✉, И. О. Горин[1], Г. Ю. Пономарев[2], В. Ю. Пылёв[1,3], В. С. Петрушенко[1], Н. В. Маркина[2], А. Д. Мамаева[2], А. К. Ларин[4], А. Т. Агджоян[2]

[1] Медико-генетический научный центр имени Н. П. Бочкова, Москва, Россия

[2] Институт общей генетики имени Н. И. Вавилова, Москва, Россия

[3] Биобанк Северной Евразии, Москва, Россия

[4] Федеральный научно-клинический центр физико-химической медицины Федерального медико-биологического агентства, Москва, Россия

Актуальность проблемы генетического вклада дославянского населения в генофонд русских популяций и генетического следа вторжения Золотой орды со временем лишь возрастает. Включение в арсенал полногеномных данных о широком круге популяций позволяет искать наиболее корректные решения этой проблемы. Целью работы был поиск следов взаимодействия генофондов финноязычных, славянских и тюркоязычных народов Центральной России и Поволжья и их отражения в фармакогенетическом ландшафте по данным о 248 геномах представителей 47 популяций 9 этносов с помощью моделирования и картографирования предковых компонент ADMIXTURE. Выявлены специфичные компоненты для каждого из финноязычных народов, но лишь предковые компоненты мордвы распространены во всех популяциях региона независимо от их языковой принадлежности. Генофонды русских популяций включают 80% собственной компоненты, 19% вклада финноязычных народов, 1% центральноазиатского влияния. Генофонд татар является комбинацией всех выявленных предковых компонент, включая 81% вклада финноязычных народов и лишь 12% центральноазиатского вклада, что затрудняет оценку их влияния на русский генофонд. Карта генетических расстояний от русских Рязанской области по панели 42 фармакогенетических маркеров выявила ландшафт, резко отличный от селективно-нейтрального ландшафта предковых компонент. Наиболее близки по фармакогенетическому статусу к рязанским русским популяции Мордовии, Калужской, Смоленской и Костромской областей. Схожие по селективно-нейтральным геномам рязанские и нижегородские популяции резко различаются по фармакогенетическому статусу. Это подтверждает необходимость прицельно исследовать фармакогенетические особенности популяций России.

Ключевые слова: генофонд, геногеография, предковые компоненты, ADMIXTURE, полногеномные панели, фармакогенетика, европейская часть России

Genetic history of the Russian people involves contributions from pre-Slavic populations and genetic footprints of the Golden Horde invasion. Gene pools of modern Russians are thought to result from interactions of three ethnic layers: pre-Slavic (Finnic-speaking), Slavic, and Golden Horde (Turkic-speaking). In the perspective of population genetics, these interactions have diverse projections including Y-chromosome phylogeography, selectively neutral ancestral components of autosomal genomes, and selectively relevant pharmacogenetic landscapes of DNA markers that determine drug sensitivity. However, the degree of interaction significantly varies within the indigenous geographical range of Russians [1]. For informative analysis, it is reasonable to focus on a nodal territory with the highest possible degree of interpenetration of the three genetic influences [2]. An excellent candidate territory for this role is the Volga-Oka interfluve in general and Ryazan region in particular.

In the second half of the 1st millennium AD, Slavic tribes started to penetrate into these lands, inhabited by Finnic-speaking and partly Baltic tribes, and the vectors of their migration were diverse. According to existing evidence, Slavic tribes initially arriving from southwestern territories were subsequently joined by Slavs from the northwest of Eastern Europe at the beginning of the 2nd millennium [1, 3–6]. In the early 11th century, the Murom principality was established, incorporating Ryazan lands [4, 7]. In the mid 12th century, the Murom principality splits into two, with capitals in Murom and Staraya Ryazan. In 1237, the Ryazan principality becomes the first casualty of the Mongol invasion led by Batu; since then, raids and devastation of Ryazan lands continue for over 350 years. In 1521, the Ryazan principality experiencing critical loss of its territories ultimately comes under control of Moscow sovereigns, but, even with subordination to Moscow, the ruin of Ryazan lands by Tatar raids continues until 1594. Taking into account the early military encounters of Ryazan people with the neighboring Volga Bulgaria (Ryazan campaigns against them in 1172 and 1183 are documented), the interaction of Ryazan people with the Turkic world, located at its borders, can be dated to the 12th century or earlier. In addition, the Ryazan region was, in a sense, an outpost that bordered on the Wild Field (rus. Dikoe Pole; the vast steppes sparsely populated by nomadic groups). It is reasonable therefore to view the Ryazan region as the major hub of interpenetration between gene pools of Slavic and Turkic-speaking populations, with corresponding genetic footprints in its modern Russian populations. The interaction between Slavic and Finnic-speaking tribes has an even longer history. Overall, the "nodal" territory of the Volga-Oka interfluve and Ryazan lands provides arguably the best model for studying genetic footprints of Finnic-speaking, Slavic, and Turkic-speaking tribes and peoples.

The modern methods of DNA analysis allow reconstruction of ancient genomes from excavated human remains [8–14]. However, the number of ancient genomes suitable for analysis is limited, especially for populations that practiced, like the Slavs, cremation of the dead. An alternative important source of information on population history is provided by modern genomes subject to genome-wide genotyping or sequencing [15–19]. The most appropriate bioinformatics handling for such data is provided by the autosomal genome ancestral component modeling tool ADMIXTURE [20].

Genetic interactions among peoples of Indo-European, Uralic, and Altaic language families have been considered in a number of studies applying genome-wide analysis to the modern gene pool of Northern Eurasia [21–26]. For instance, a genome-wide panel-assisted reconstruction of gene pools for Balto-Slavic populations [21] reveals the genetic proximity of the Balts (Lithuanians, Latvians) to the Volga group of Finno-Ugric peoples and especially to Mordovians. The Slavs, both Eastern and Western, absorbed the local pre-Slavic Eastern European genetic substratum. A genome-wide study of modern ethnic groups populating the East European Plain [22] reveals the "East Asian" ancestral component contributing 20% to gene pool of Bashkirs and 5% to gene pools of Chuvashs and Volga Tatars. Another genome-wide study identifies a specific ancestral component shared by peoples of the Uralic language family, including Finnic-speaking Karelians, Mordovians, Mari, and Udmurts, and defining the degree of their genetic relationship [23]. A genome-wide genetic study of North Eurasian populations reveals three clines stretching from west to east [24]. Subsequent analysis shows that gene pools of Turkic-speaking and Uralic-speaking populations in Povolzhye are highly similar, although the Uralic-speaking populations genetically gravitate towards Trans-Ural Ugrians. Comparison of autosomal genome data between Novgorod region and a wide range of populations in the European part of Russia and the Urals produced a hypothesis on considerable preservation of the local pre-Slavic population legacy in gene pools of the Novgorod region, which turned out to be closer to the eastern Finnic-speaking groups (Volga and Perm) than to the western (Baltic) [25]. Another important line of evidence is provided by pharmacogenetic studies, which enable creation of cartographic atlases of subcontinents, but consider local variants as well. For instance, Besermyans and Udmurts are pharmacogenetically close to indigenous populations of Volga Region, Urals, and Southern Urals, but distant from inhabitants of more remote regions [26].

This study aimed at modeling of ancestral components in order to reveal genetic footprints of interactions among Finnic-speaking, Slavic and Turkic-speaking ethnic groups in the autosomal gene pool of modern Russian populations inhabiting the nodal region of the Volga-Oka interfluve. The second, more applied, goal of this study was to create maps of pharmacogenetic DNA markers and contemplate pharmacogenetic landscape of the studied geographic area.

METHODS

Methodological and bioinformatics aspects of the analysis of autosomal gene pools using genome-wide panels have been described in detail previously [27]. The genotyping for a genome-wide panel of 4.5 million SNP markers was performed using Infinium OmniExome BeadChip Kit (Illumina; USA) with an iScan system (Illumina; USA). Primary analysis and quality assessment of the data was carried out in the GenomeStudio v2011.1 software at a CallRate of at least 0.99.

The population genetic analysis for small panels of autosomal markers requires samples of at least 50 individuals. By contrast, genome-wide panels comprising millions of DNA markers afford a reliable output on much smaller samples of 5–10 individuals. Since the reduced sample size implies ultimate tightening of the sampling criteria, we emphasize that all genomes included in this study were selected in accordance with internationally recognized criteria [28]. In particular, genealogies of all participants, traced at least three generations backward, proved their origin from a given population and identification with a given ethnic group.

The "nodal" Ryazan region was represented by 20 genomes from 4 ethnic Russian populations (Mikhailovsky, Spassky, Sapozhkovsky, and Saraevsky districts), with Russian populations in Tver, Kostroma, Smolensk, Kaluga, Oryol, Tambov, and Nizhny Novgorod regions included for comparison.

Table. Estimated contribution of each of the ten ancestral components identified by ADMIXTURE at $k = 3$ and $k = 7$ to genomes in each of the studied populations

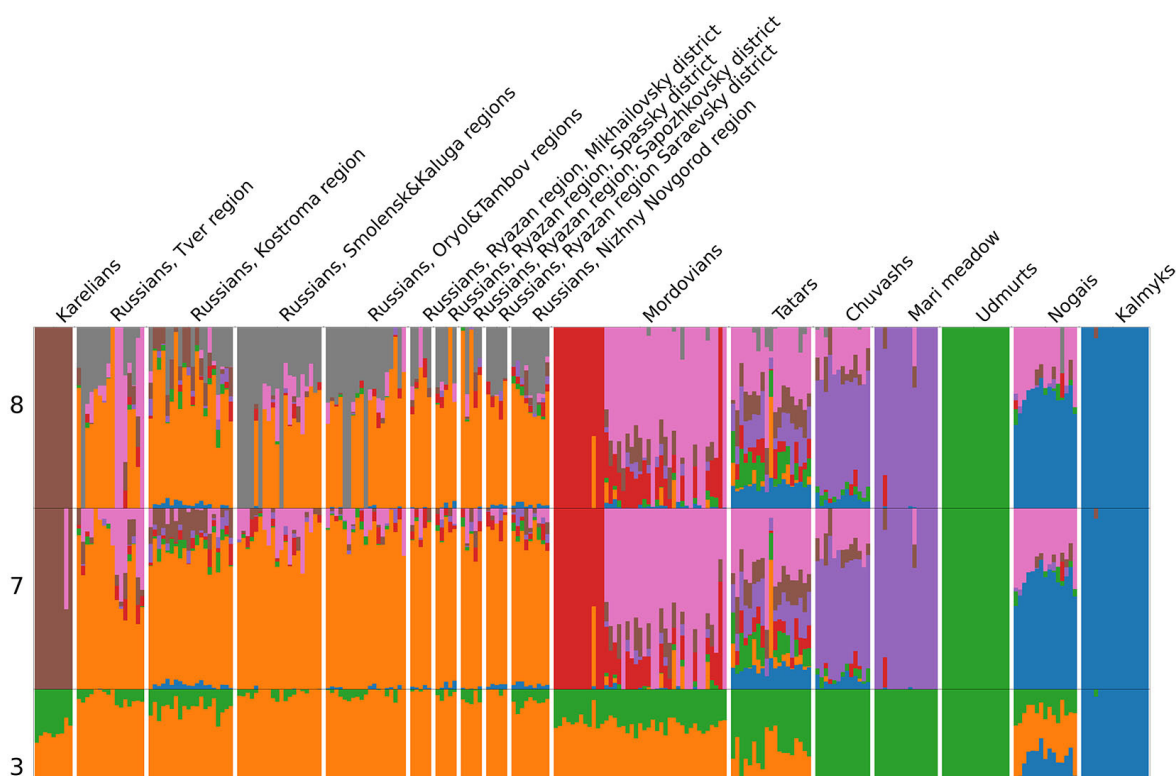| Ethnic groups | Populations | Population number | Genome number | ANCESTRAL COMPONENT CONTRIBUTION, AVERAGE (%) | | | | | | | | | | Total contribution of Finnic-speaking ethnic groups (%) | |
| | | | | $k = 3$ | | | $k = 7$ | | | | | | | $k = 7$ | |
				"Western"	"Ural"	"Eastern"	"Karelian"	"Slavic"	"Mordovian-1"	"Mordovian-2"	"Mari"	"Udmurt"	"Kalmyk"	except "Karelian"	"Karelian" included
Karelians	South	1	9	75	24	2	94	0	6	0	0	0	0	6	100
Russians	Tver region	2	16	95	5	0	6	70	19	3	1	1	0	24	30
	Kostroma region	1	20	87	10	3	11	73	3	3	4	4	3	14	25
	Smolensk and Kaluga regions	6	20	97	2	0	2	82	12	2	1	1	0	16	18
	Oryol and Tambov regions	3	19	97	3	1	1	87	7	2	1	1	1	11	12
	Ryazan region, Mikhailovsky district	1	5	98	0	2	2	86	5	5	0	1	1	11	13
	Ryazan region, Spassky district	1	5	95	3	3	1	86	3	5	2	1	2	11	12
	Ryazan region, Sapozhkovsky district	1	5	93	6	1	1	87	3	5	1	0	2	9	10
	Ryazan region, Saraevsky district	1	5	96	1	3	4	81	6	4	3	2	1	15	19
	Ryazan region totally	*4*	*20*	*95*	*2*	*2*	*2*	*85*	*4*	*5*	*1*	*1*	*2*	*11*	*13*
	Nizhny Novgorod region	4	9	90	8	3	3	81	4	4	3	3	3	14	17
Mordovians	Moksha, Erzya, Shoksha	3	41	78	19	3	5	3	53	37	1	1	0	92	97
Tatars	Kazan	5	19	52	34	14	13	6	36	7	15	10	12	68	81
Chuvashs	Chuvashian	4	13	23	67	9	7	0	20	2	62	3	5	87	94
Mari	Meadow	1	15	4	92	5	2	0	1	1	96	0	0	98	100
Udmurts	Udmurtian	7	16	0	99	1	0	0	0	0	0	100	0	100	100
Nogais	Astrakhan, Stavropol	2	15	25	13	62	2	0	33	1	1	2	61	37	39
Kalmyks	Six tribal groups	6	16	0	0	100	0	0	0	0	0	0	100	0	0
	TOTALLY	49	248												

Fig. 1. Contributions of the ancestral components identified by ADMIXTURE at $k = 3$, $k = 7$, and $k = 8$ to individual genomes in the studied populations. Each ancestral component is indicated by specific color; each vertical line represents individual genome with a palette of ancestral component contributions; the populations are demarcated by *white* lines; the raw data are given in the table

The Finnic-speaking populations of the Volga-Ural region were represented by Mordovians (Erzya, Moksha, Shoksha), Mari, and Udmurts, whereas southern Karelians were included as the most geographically close representative of the western branch of the Finnic-speaking peoples. The Turkic-speaking populations of Volga Region and Ural were represented by Kazan Tatars and Chuvashs, with Astrakhan and Stavropol Nogais included for comparison. Identification of genetic footprints of Mongolic-speaking peoples involved genome-wide data for six tribal groups of Kalmyks.

The analysis of ancestral components was carried out using the ADMIXTURE bioinformatics tool for 248 individual genomes representing 47 populations of 9 ethnic groups (Table), including 104 genomes from Russian populations, 81 genomes from four Finnic-speaking peoples, 47 genomes from three Turkic-speaking peoples, and 16 genomes of Mongolic-speaking Kalmyks. The ADMIXTURE tool affords quantitative assessment for the contributions of different ancestral components to each individual genome [20, 29]. The ancestral components are modeled for the same uploaded set of genomes, with each level of modeling carried out independently. The number of ancestral components k is the only parameter specified by the user. At $k = 2$, contributions of two ancestral components are modeled for each genome; at $k = 3$, the tool presents the same genomes with three ancestral components; at $k = 20$, the tool reconstructs contributions of twenty ancestral components for the same set of genomes, etc.; as the k increases, the patterns become increasingly elaborated. The contribution of particular ancestral component to a gene pool is estimated by averaging its contributions to individual genomes.

A series of pharmacogenetic maps were built to assess the interactions among Finnic-speaking, Slavic, and Turkic-speaking ethnic groups in pharmacogenetic perspective and estimate their impact on the modern pharmacogenetic landscape.

The mapping employed data on 42 key pharmacogenetic markers (the absorption, distribution, metabolism, and excretion (ADME) genes; pharmacological target-encoding genes; and hemostasis system genes) derived from the same genome-wide genotyping datasets previously used in the ADMIXTURE processing [26]. The incidence matrix for the studied 42 pharmacogenetic DNA markers comprised data for 16 pooled populations (to increase sample size). Calculation of Nei's genetic distances (*d*) based on this matrix produced 42 partial maps showing the extent of pharmacogenetic similarity between Ryazan and other regions for each of the studied markers. The averaging of partial maps produced the map of average pharmacogenetic distances from Ryazan, reflecting its pharmacogenetic status with regard to other subjects within the studied geographic area.

All maps of pharmacogenetic landscapes and ancestral components were built using the original GeneGeo mapping package [30] using the weighted average interpolation method with an influence radius of 400 km and a weight function value of 3. The genogeographic technology has been described in detail elsewhere [2, 31].

RESULTS

Modeling of ancestral components for the studied scope of 47 populations was carried out at 13 levels of k, obtained by sequentially incrementing k by 1, from 2 to 14 inclusive. Two models turned out to be the most informative for solving the main problem: at $k = 3$ and $k = 7$ (Table). The level of $k = 3$ reveals three ancestral components conventionally defined as "Western", "Ural" and "Eastern". At the level of $k = 7$, the ancestral components of the western and eastern Finnic-speaking peoples become separated for the first time, which allows differentiating their contributions. The estimated contributions to individual genomes for each of the identified

ancestral components are given in Table. Contributions to individual genomes for each ancestral component at $k = 3$, $k = 7$, and $k = 8$ are presented in Fig. 1. The level of $k = 8$ preserves the contributions of all components identified at the previous steps of analysis, while the new eighth component further elaborates the structure of Russian populations.

To validate the observed trends, modeling at each level from $k = 2$ to $k = 14$ was run in 10 repeats (yielding a total of 130 models). At $k = 3$, all models were virtually identical; at $k = 7$, six of ten runs revealed stable ancestral components (these are described in the text of the article). In the remaining four runs, one of the ancestral components was replaced by an alternative, and each of these runs presented with a higher simulation error value.

DISCUSSION

Modeling of three ancestral components

The analysis at $k = 3$ revealed three ancestral components conventionally designated "Western", "Ural", and "Eastern". Most notably, the identified ancestral components poorly fit into the framework of three language families (Slavic, Finnic, and Turkic) (Fig. 1, Table).

"Western" ancestral component

This component prevails in all Russian populations (95%), but also in Finnic-speaking populations of Karelians (75%) and Mordovians (78%) (Fig. 2A, Table). Moreover, it constitutes a significant portion of gene pools in Turkic-speaking peoples: more than a half in Kazan Tatars (52%) and about a quarter in both Nogais (25%) and Chuvashs (23%).

"Ural" ancestral component

This component dominates in gene pools of Udmurts (99%) and Mari (91%) (Fig. 2B, Table). It is also prominent in Turkic-speaking peoples, accounting for two-thirds of Chuvash (67%) and a third of Tatar (34%) gene pools. A smaller but still substantive contribution of the "Ural" ancestral component is found in Karelians (24%) and Mordovians (19%). The average contribution of this component to Russian populations is small (4%) with the maxima in Kostroma and Nizhny Novgorod regions.

"Eastern" ancestral component

This component totally prevails (100%) in all six Kalmyk tribal groups included in this study, so it provides a suitable measure of the Central Asian influence on European gene pools (Fig. 2C, Table). This component is also prominent in Nogais (62%), which confirms its "Central Asian" status. Among Volga Region peoples, the highest Central Asian influence is observed in Kazan Tatars (14%) and Chuvashs (9%). In other studied gene pools, the "Eastern" influence is small, 5% (in Mari) or lower. Its average contribution to Russian gene pools is 1% (up to 3% in eastern districts of the Nizhny Novgorod and Ryazan regions).

Kazan Tatars

The "composite" nature of gene pool in Kazan Tatars, represented by five populations, should be discussed in detail. The subtle interpopulation divergence is due to variable "Western" (48–60%) and "Ural" (26–38%) contributions
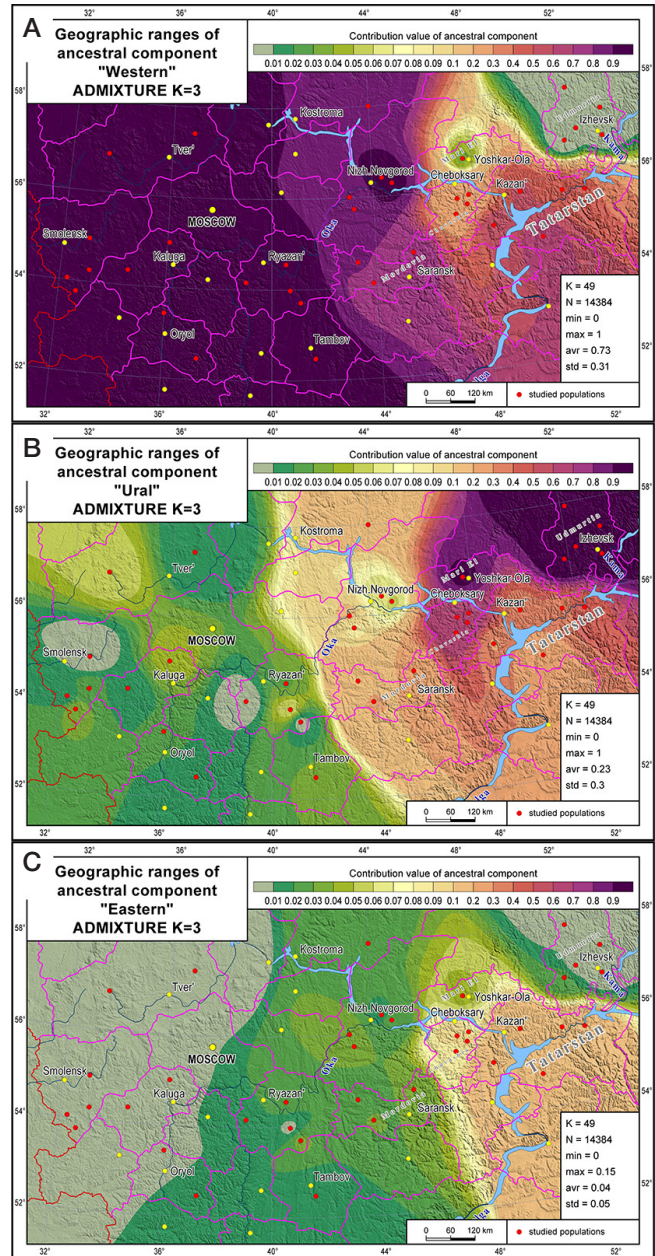

Fig. 2. Geographic ranges of three ancestral components identified by ADMIXTURE at $k = 3$: "Western" (**A**), "Ural" (**B**), and "Eastern" (**C**). High prevalence rates are colored *red-to-purple*, low values are green, the scale of transitions is given in the map; studied populations are indicated by *red* circles

accompanied by similar "Eastern" contributions (14–15%). The dominant "Western" component (a half or more) was followed by "Ural" (roughly one third) and "Eastern" (14%) components in all studied populations. Increasing the resolution of analysis by incrementing k revealed some minor ancestral components, but these were shared with other ethnic groups. The analysis identified no singular ancestral component for Kazan Tatars; the "composite" structure preserved at higher levels of k prevents using their gene pool for evaluation of the "tatar" influence on the neighboring Russian populations.

Modeling of seven ancestral components

Analysis at $k = 7$ yielded four new ancestral components, though not as a result of sheer branching of those identified at previous steps of the analysis. The picture is much more complex: the new components mosaically absorb the elements

of "Western" and "Ural" components revealed at $k = 3$. It should be emphasized that the components were attributed with "ethnic" trivial names only for the sake of brevity.

"Karelian" ancestral component

This component, which marks the contribution of Western Finnic-speaking ethnic groups, accounts for 94% of Karelian gene pools and is minor in other populations (Table), with secondary maxima in gene pools of Kazan Tatars and Kostroma Russians (11%).

"Slavic" ancestral component

This component dominates in all ethnic Russian populations (81% on average, within the total range of 70–87%) (Fig. 3A, Table) and is virtually absent in other gene pools with the exception of Kazan Tatars (6%). The accentuated presence of the "Slavic" component in the Tatar gene pool cannot be explained genealogically, given its 80% prevalence in individual genomes. Although this component is also detectable in Mordovian gene pools (3%), it is present in only 17% of genomes in the northwest of Mordovia.

"Mordovian-1" ancestral component

This component shares the second largest geographic range with "Mordovian-2" (Fig. 3B, Table). Reaching maximum (53%) in gene pools of Mordovia, it is also ubiquitously found in other populations. Its prominent contributions are characteristic of Turkic-speaking peoples: 36%, 35%, and 20% in gene pools of Kazan Tatars, Astrakhan Nogais, and Chuvashs. Importantly, the "Mordovian-1" component shows almost total prevalence in these ethnic groups, contributing to almost all individual genomes (Table 2), which indicates its historical significance in gene pools of the Turkic-speaking peoples of Volga Region.

Contribution of the "Mordovian-1" ancestral component to gene pools of ethnic Russians is modest (7%) despite rather high prevalence (60% of individual genomes). The maxima are encountered in Tver (19%) and Kaluga (16%) regions, with a very high prevalence (80–90% of individual genomes; Fig. 4); in other regions, the prevalence is lower (45–65% of genomes). Overall, the "Mordovian-1" ancestral component is ubiquitously found in gene pools of almost all Slavic-speaking, Turkic-speaking and Finnic-speaking populations within the studied geographic area.

"Mordovian-2" ancestral component

The component shown in Fig. 3C and Table 1, has a more distinguished authenticity: it is already identifiable at $k = 4$, whereas the "Mordovian-1" component arrives at $k = 7$. Genomes of Mordovia show distinct clusterization (Fig. 1): one-fifth of them are 100% "Mordovian-1" and another one-fifth are 100% "Mordovian-2". This component is found in gene pools of all studied populations (except Udmurts). In none of them, however, its contribution exceeds 5%, apart from, again, Kazan Tatars: with the average contribution of 6%, the "Mordovian-2" component is 90% prevalent in Tatars (in Chuvashs, it is found in 40% of genomes only).

In gene pools of ethnic Russians, the "Mordovian-2" component is relatively weak (3% on average) but ubiquitous. Moreover, it is encountered in 60% of individual Russian genomes, most commonly in eastern regions (Kostroma, Nizhny Novgorod, and Ryazan) (Figs. 3C and 4).
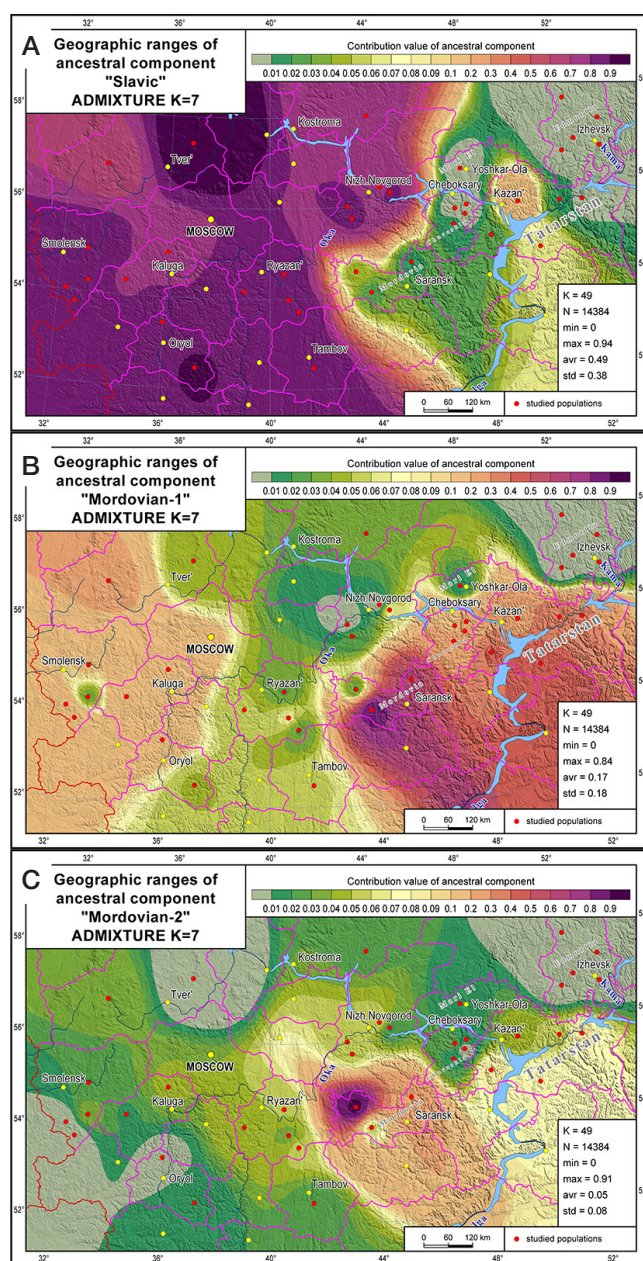

Fig. 3. Geographic ranges of three ancestral components identified by ADMIXTURE at $k = 7$: "Slavic" (**A**), "Mordovian-1" (**B**), and "Mordovian-2" (**C**). High prevalence rates are colored *red-to-purple*, low values are green, the scale of transitions is given in the map; studied populations are indicated by *red* circles

"Mari" ancestral component

This component firstly arrives at the level of five ancestral components ($k = 5$) and almost totally prevails in the meadow Mari gene pool (96%). It also accounts for two-thirds of the Chuvash gene pool (62%), with similarly high levels in all Chuvash populations (57–65%) (Table). Of other ethnic groups, the most significant contribution of the "Mari" component is encountered in Kazan Tatars (15% on average, with 100% prevalence in individual genomes). In other studied populations, contributions of the "Mari" component never exceeds 4% (Table).

"Udmurt" ancestral component

This component firstly arrives at $k = 3$ and has been already described by us as "Ural" (Table, Fig. 2B). At all higher levels it
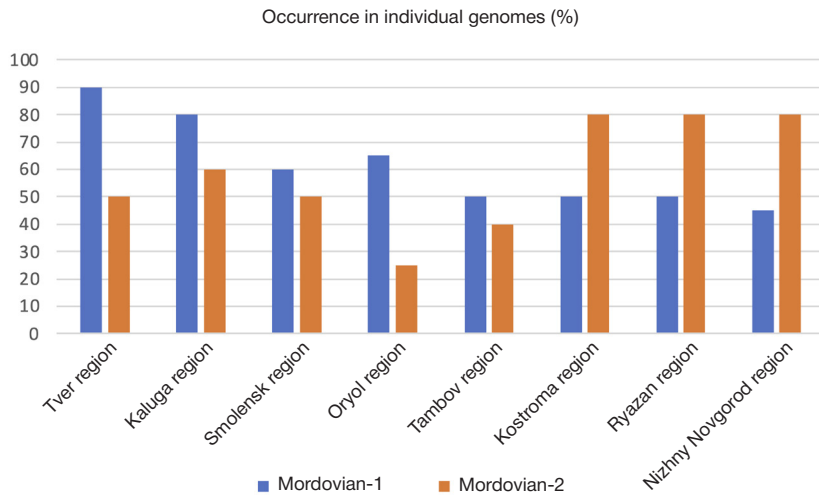
Occurrence in individual genomes (%)

Fig. 4. Fractions of individual genomes (%) harboring "Mordovian-1" and "Mordovian-2" ancestral components in ethnic Russian populations

accounts for 100% of the Udmurt gene pool, while being minor (within 4%) in gene pools of other peoples. The only exception is, again, Kazan Tatars: the "Udmurt" component accounts for 10% of the gene pool and is present in almost all individual Tatar genomes with a maximal contribution of 21%.

"Kalmyk" ancestral component

This component fixes the "breath" of Central Asia; it firstly arrives at $k = 2$ and has been already described by us as "Eastern" (Fig. 2C). Among the studied populations, it is only prominent in gene pools of Kalmyks (100%) and Nogais (61%). Of other ethnic groups, it is present at highest in Kazan Tatars (12%). Noteworthy, the "Kalmyk" component was found in all individual genomes of Tatars, constituting 7–17%. In other studied gene pools, contributions of the "Kalmyk" component never exceeded 5% (Table).

Ryazan gene pool

Four modern populations of ethnic Russians (Figs. 1–3, Table) provided a relevant model for the assessment of the mutual

genetic influence of pre-Slavic, Slavic, and Turkic-speaking populations in the "nodal" Ryazan region. We picked one district (Mikhailovsky) at the very west of Ryazan region and three districts (Spassky, Sapozhkovsky, and Saraevsky) located on the same transect from north to south, with Saraevsky being borderline. The analysis indicates similar genetic constitution of the four gene pools, with certain differences in contributions of Finnic-speaking peoples: 19% in the borderline Saraevsky district and as low as 10–13% in the other three districts (Table). Given the equally small Central Asian influence in all four populations (1–2%), this difference could not be directly related to the Golden Horde invasion, nor attributed to the influence of any known pre-Slavic tribe. The only suggestion to explain the authenticity of gene pools in the southeastern Ryazan lands is the higher influence of the Wild Field in this borderline area.

Pharmacogenetic status of Ryazan Russians

Analysis of genetic markers associated with pharmacologic phenotypes is a prerequisite in the transition to personalized medicine in terms of optimal drug choice and medication regimen adjustment. However, the majority of studies in this
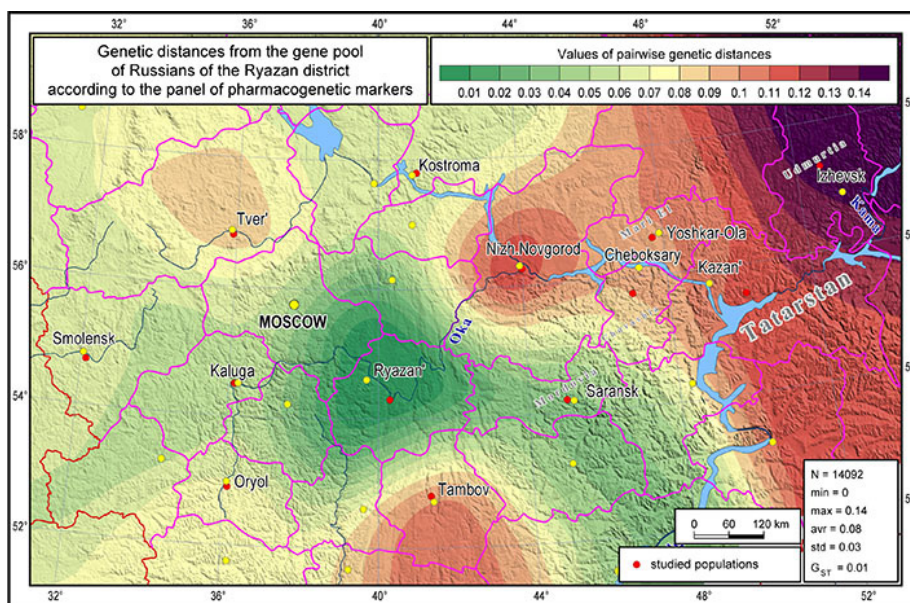

Fig. 5. Pharmacogenetic landscape of the studied geographic area, representing genetic distances from Ryazan Russians according to pharmacogenetic markers. Small genetic distances (indicating pharmacogenetic proximity to Ryazan Russians) are shown in *green*; large genetic distances (indicating pharmacogenetic divergence from Ryazan Russians) are shown in *red-to-purple*; the scale of transitions is given in the map

field have been focused on Western Europe and the results have little application to populations of Russia with their huge genetic diversity [31].

To assess the uniqueness of pharmacogenetic landscape within the studied geographic area, a map of genetic distances (d) from Ryazan Russians was created using an extensive panel of pharmacogenetic markers (Fig. 5). In contrast with the ancestral component maps based on selectively neutral DNA markers (Figs. 2 and 3), pharmacogenetic mapping revealed the highest proximity of Ryazan Russians to their Finnic-speaking neighbor — Mordovian populations (0.03 < d < 0.04). One step more distant from Ryazan Russians in terms of pharmacogenetic status were Russian populations of Kaluga, Smolensk, and Kostroma regions (0.05 < d < 0.07), followed by Russians in Oryol and Tver regions (0.08 < d < 0.09). The third most similar to Ryazan Russians were Tambov Russians and their eastern neighbors — the Finnic-speaking Mari and the Turkic-speaking Chuvash peoples (0.09 < d < 0.10). Pharmacogenetic portraits of Tatar and Udmurt peoples were expectedly divergent from those of Ryazan Russians (0.11 < d < 0.15). The highest pharmacogenetics divergency from Ryazan Russians was most unexpectedly revealed by Russians of the adjacent Nizhny Novgorod region (0.11 < d < 0.12) despite the substantive similarity of selectively neutral genomic patterns between the two regions (Figs. 1–4).

Overall, comparing pharamacogenetic landscape and selectively neutral genomic pattern maps demonstrates that optimization of healthcare programs at the regional level should not be based on averaged genetic status of the target populations, but requires specific assessment of local pharmacogenetic landscapes.

CONCLUSIONS

Modeling of ancestral components for the autosomal gene pool of modern populations in the nodal region of interaction between Finnic-speaking, Slavic, and Turkic-speaking peoples revealed that (1) the Finnic-speaking ethnic groups of

Volga Region (Udmurts, Mari, and Mordovians) lack common ancestral components; instead, each of these groups has its own characteristic ancestral components. Remarkably, two ancestral components identified in the Mordovian gene pool can be traced in almost all populations within the studied geographic area, regardless of their linguistic affiliation, which allows us to suggest that genetic portrait of pre-Slavic population of the studied geographic area included two main "colors" preserved in the modern gene pool of Mordovia. (2) The contribution of Finnic-speaking ethnicities to gene pools of modern Turkic-speaking peoples in Volga Region is enormous: ancestral components associated with Finnic-speaking ethnic groups constitute 81% and 94% in gene pools of Kazan Tatars and Chuvashs, respectively. (3) Gene pool of Kazan Tatars is the most "composite" of all gene pools in this study, most organically incorporating all ancestral components observed in other gene pools within the studied geographic area. Although the Central Asian influence is most pronounced among the Kazan Tatars, its contribution is low (12%), seven times lower than the total contribution of the Finnish-speaking peoples (81%), which makes it a poor basis for evaluation of the "Tatar" influence. (4) Gene pools of the studied Russian populations represent a single cluster, which can be basically (80%) described by characteristic ancestral component. At the same time, gene pools of modern ethnic Russians incorporate all other ancestral components found in genetic landscapes of other ethnic groups within the studied geographic area. The "nodal" Ryazan region is fully archetypal in terms of common features of the Slavic cluster with a somewhat increased total contribution of Finnish-speaking groups towards its southeastern border. (5) The picture of diversity based on selectively neutral ancestral components is complemented by unique pharmacogenetic landscapes revealed with a custom panel within the same genome-wide genotyping platform. The knowledge of pharmacogenetic parameters for a given population is essential for the future of personalized medicine and proper logistics of pharmaceuticals at the regional level with regard to genetic diversity of modern Russian populations.

References

1. Sedov VV. Osvoenie slavyanami Vostochnoevropejskoj ravniny. V knige: Vostochnye slavyane. Antropologiya i ehtnicheskaya istoriya. M.: Nauchnyj mir, 1999; s. 153–160. Russian.

2. Balanovskaya EV, Balanovskij OP. Russkij genofond na Russkoj ravnine. M.: Luch, 2007. Russian.

3. Leontev AE, Ryabinin EA. Ehtapy i formy assimilyacii letopisnoj meri (postanovka voprosa). Sovetskaya arxeologiya. 1980; (2): 67–79. Russian.

4. Bejlekchi VV. Drevnosti letopisnoj muromy (pogrebal'nyj obryad i poseleniya). M.: Murom, 2005; 278 s. Russian.

5. Alekseeva TI. Ehtnogenez vostochnyx slavyan po dannym antropologii. M.: Nauka, 1973; 329 s. Russian.

6. Sedov VV. Drevnerusskaya narodnost': Istoriko-arxeologicheskoe issledovanie. M.: Yazyki russkoj kul'tury, 1999; 316 s. Russian.

7. Goryunova EI. Ehtnicheskaya istoriya Volgo-Okskogo mezhdurech'ya. M.: Izd-vo AN SSSR, 1961; 264 c. Russian.

8. Rizzi E, Lari M, Gigli E, De Bellis G, Caramelli D. Ancient DNA studies: new perspectives on old samples. Genet Sel Evol. 2012 Jul 6; 44 (1): 21. DOI: 10.1186/1297-9686-44-21.

9. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014; 513: 409–413. DOI: 10.1038/nature13673.

10. Haak W, Lazaridis I, Patterson N, Rohland N, Mallik S, Llamas B et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015 Jun 11; 522 (7555): 207–11. DOI: 10.1038/nature14317.

11. Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J et al. Population genomics of Bronze Age Eurasia. Nature. 2015 Jun 11; 522 (7555): 167–72. DOI: 10.1038/nature14507.

12. Jones ER, Zarina G, Moiseyev V, et al. The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. Curr Biol. 2017 Feb 20; 27 (4): 576–82. DOI: 10.1016/j.cub.2016.12.060.

13. Mittnik A, Wang CC, Pfrengle S, Daubaras M, Zarina G, Hallgren F et al. The genetic prehistory of the Baltic Sea region. Nat Commun. 2018 Jan 30; 9 (1): 442. DOI: 10.1038/s41467-018-02825-9.

14. Lamnidis TC, Majander K, Jeong C, Salmela E, Wessman A, Moiseyev V et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. Nat Commun. 2018 Nov 27; 9 (5018). DOI: 10.1038/s41467-018-07483-5.

15. Li JZ, Absher DM, Tang H, et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 2008; 319 (5866): 1100–4. DOI: 10.1126/science.1153717.

16. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko A, Auton A, et

al. Genes mirror geography within Europe. Nature. 2008 Aug 31; 456: 98–101. DOI: 10.1038/nature07331.

17. Pagani L, Lawson D, Jagoda E, Morseburg A, Eriksson A, Mitt M, et al. Genomic analyses inform on migration events during the peopling of Eurasia. Nature. 2016 Oct 13; 538 (7624): 238–42. DOI: 10.1038/nature19792.

18. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016 Oct 13; 538 (7624): 201–6. DOI: 10.1038/nature18964.

19. Malaspinas AS, Westaway M, Muller C, Sousa V, Lao O, Alves I et al. A genomic history of Aboriginal Australia. Nature. 2016 Oct 13; 538 (7624): 207–14. DOI: 10.1038/nature18299.

20. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011; 12 (1): 246. DOI: 10.1186/1471-2105-12-246.

21. Kushniarevich A, Utevska O, Chuhryaeva M, Agdzhoyan A, Dibirova K, Uktverite I et al. Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data. PLoS ONE. 2015 Sep 2; 10 (9). PubMed PMID: 26332464. DOI: 10.1371/journal.pone.0135820.

22. Triska P, Chekanov N, Stepanov V, Khusnutdinova E, Arun Kumar GP, Akhmetova V. Between Lake Baikal and the Baltic Sea: genomic history of the gateway to Europe. BMC Genetics. 2017 Dec 28; 18 (1): 5–20. PubMed PMID: 29297395. DOI: 10.1186/s12863-017-0578-3.

23. Tambets K, Yunusbayev B, Hudjashov G, Ilumäe A-M, Rootsi S, Honkola T. Genes reveal traces of common recent demographic history for most of the Uralicspeaking populations. BMC Genome Biology. 2018 Sep 21; 19 (1): 1–20. DOI: 10.1186/s13059-018-1522-1.

24. Jeong C, Balanovsky O, Lukianova E, Kahbatkyzy N, Flegontov P, Zaporozhchenko V, et al. The genetic history of admixture across inner Eurasia. Nature ecology & evolution. 2019 Jun; 3 (6): 966–76. PubMed PMID: 31036896. DOI: 10.1038/s41559-019-0878-2.

25. Balanovska EV, Chernevsky DK, Balanovsky OP. Svoeobrazie Novgorodskogo genofonda v kontekste narodonaseleniya evropejskoj chasti Rossii. Vestnik Novgorodskogo gosudarstvennogo universiteta. Ser.: Medicinskie nauki. 2021; 124 (3): 51–57. DOI: 10.34680/2076-8052.2021. Russian.

26. Balanovska EV, Petrushenko VS, Koshel SM, Pocheshhova EA, Chernevsky DK, Mirzaev KB, et al. Cartographic atlas of frequency variation for 45 pharmacogenetic markers in populations of Russia and its neighbor states. Vestnik RGMU. 2020; (6): 38–50. DOI: 10.24075/brsmu.2020.080.

27. Balanovsky OP, Gorin IO, Zapisetskaya YS, Golubeva AA, Kostryukova ES, Balanovska EV. Interactions between gene pools of russian and finnish-speaking populations from Tver region: analysis of 4 million SNP markers. Vestnik RGMU. 2020; (6): 15–22. DOI: 10.24075/brsmu.2020.072.

28. Balanovska EV, Zhabagin MK, Agdjoyan AT, Chuhryaeva MI, Markina NV, Balaganskaya OA, et al. Populyacionnye biobanki: principy organizacii i perspektivy primeneniya v genogeografii i personalizirovannoj medicine. Genetika. 2016; 52 (12): 1371–87. DOI: 10.7868/S001667581612002X. Russian.

29. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009 Sep; 19 (9): 1655–64. DOI: 10.1101/gr.094052.109.

30. Koshel SM. Geoinformacionnye texnologii v genogeografii. V knige: Sovremennaya geograficheskaya kartografiya. M: Data+, 2012; s. 158–166. Russian.

31. Balanovsky OP. Genofond Evropy. M.: Tovarishhestvo nauchnyx izdanij KMK, 2015. Russian.

Литература

1. Седов В. В. Освоение славянами Восточноевропейской равнины. В книге: Восточные славяне. Антропология и этническая история. М.: Научный мир, 1999; с. 153–160.

2. Балановская Е. В., Балановский О. П. Русский генофонд на Русской равнине. М.: Луч, 2007.

3. Леонтьев А. Е., Рябинин Е. А. Этапы и формы ассимиляции летописной мери (постановка вопроса). Советская археология. 1980; (2): 67–79.

4. Бейлекчи В. В. Древности летописной муромы (погребальный обряд и поселения). М.: Муром, 2005; 278 с.

5. Алексеева Т. И. Этногенез восточных славян по данным антропологии. М.: Наука, 1973; 329 с.

6. Седов В. В. Древнерусская народность: Историко-археологическое исследование. М.: Языки русской культуры, 1999; 316 с.

7. Горюнова Е. И. Этническая история Волго-Окского междуречья. М.: Изд-во АН СССР, 1961; 264 с.

8. Rizzi E, Lari M, Gigli E, De Bellis G, Caramelli D. Ancient DNA studies: new perspectives on old samples. Genet Sel Evol. 2012 Jul 6; 44 (1): 21. DOI: 10.1186/1297-9686-44-21.

9. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014; 513: 409–413. DOI: 10.1038/nature13673.

10. Haak W, Lazaridis I, Patterson N, Rohland N, Mallik S, Llamas B et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015 Jun 11; 522 (7555): 207–11. DOI: 10.1038/nature14317.

11. Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J et al. Population genomics of Bronze Age Eurasia. Nature. 2015 Jun 11; 522 (7555): 167–72. DOI: 10.1038/nature14507.

12. Jones ER, Zarina G, Moiseyev V, et al. The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. Curr Biol. 2017 Feb 20; 27 (4): 576–82. DOI: 10.1016/j.cub.2016.12.060.

13. Mittnik A, Wang CC, Pfrengle S, Daubaras M, Zarina G, Hallgren F et al. The genetic prehistory of the Baltic Sea region. Nat Commun. 2018 Jan 30; 9 (1): 442. DOI: 10.1038/s41467-018-02825-9.

14. Lamnidis TC, Majander K, Jeong C, Salmela E, Wessman A, Moiseyev V et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. Nat Commun. 2018 Nov 27; 9 (5018): 1. DOI: 10.1038/s41467-018-07483-5.

15. Li JZ, Absher DM, Tang H, et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 2008; 319 (5866): 1100–4. DOI: 10.1126/science.1153717.

16. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko A, Auton A, et al. Genes mirror geography within Europe. Nature. 2008 Aug 31; 456: 98–101. DOI: 10.1038/nature07331.

17. Pagani L, Lawson D, Jagoda E, Morseburg A, Eriksson A, Mitt M, et al. Genomic analyses inform on migration events during the peopling of Eurasia. Nature. 2016 Oct 13; 538 (7624): 238–42. DOI: 10.1038/nature19792.

18. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016 Oct 13; 538 (7624): 201–6. DOI: 10.1038/nature18964.

19. Malaspinas AS, Westaway M, Muller C, Sousa V, Lao O, Alves I et al. A genomic history of Aboriginal Australia. Nature. 2016 Oct 13; 538 (7624): 207–14. DOI: 10.1038/nature18299.

20. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011; 12 (1): 246. DOI: 10.1186/1471-2105-12-246.

21. Kushniarevich A, Utevska O, Chuhryaeva M, Agdzhoyan A, Dibirova K, Uktverite I et al. Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data. PLoS ONE. 2015 Sep 2; 10 (9). PubMed PMID: 26332464. DOI: 10.1371/journal.pone.0135820.

22. Triska P, Chekanov N, Stepanov V, Khusnutdinova E, Arun Kumar GP, Akhmetova V. Between Lake Baikal and the Baltic Sea: genomic history of the gateway to Europe. BMC Genetics. 2017 Dec 28; 18 (1): 5–20. PubMed PMID: 29297395. DOI: 10.1186/s12863-017-0578-3.

23. Tambets K, Yunusbayev B, Hudjashov G, Ilumäe A-M, Rootsi S,

Honkola T. Genes reveal traces of common recent demographic history for most of the Uralicspeaking populations. BMC Genome Biology. 2018 Sep 21; 19 (1): 1–20. DOI: 10.1186/s13059-018-1522-1.

24. Jeong C, Balanovsky O, Lukianova E, Kahbatkyzy N, Flegontov P, Zaporozhchenko V, et al. The genetic history of admixture across inner Eurasia. Nature ecology & evolution. 2019 Jun; 3 (6): 966–76. PubMed PMID: 31036896. DOI: 10.1038/s41559-019-0878-2.

25. Балановская Е. В., Черневский Д. К., Балановский О. П. Своеобразие Новгородского генофонда в контексте народонаселения европейской части России. Вестник Новгородского государственного университета. Сер.: Медицинские науки. 2021; 124 (3): 51–57. DOI: 10.34680/2076-8052.2021.

26. Балановская Е. В., Петрушенко В. С., Кошель С. М., Почешхова Э. А., Черневский Д. К., Мирзаев К. Б. и др. Картографический атлас распространения 45 фармакогенетических маркеров в народонаселении России и сопредельных стран. Вестник РГМУ. 2020; (6): 39–52. DOI:

27. Балановский О. П., Горин И. О., Записецкая Ю. С., Голубева А. А., Кострюкова Е. С., Балановская Е. В. Взаимодействие генофондов русского и финноязычного населения Тверской области: анализ 4 млн SNP-маркеров. Вестник РГМУ. 2020; (6): 15–22. DOI: 10.24075/vrgmu.2020.072.

28. Балановская Е. В., Жабагин М. К., Агджоян А. Т., Чухряева М. И., Маркина Н. В., Балаганская О. А. и др. Популяционные биобанки: принципы организации и перспективы применения в геногеографии и персонализированной медицине. Генетика. 2016; 52 (12): 1371–87. DOI: 10.7868/S001667581612002X.

29. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009 Sep; 19 (9): 1655–64. DOI: 10.1101/gr.094052.109.

30. Кошель С. М. Геоинформационные технологии в геногеографии. В книге: Современная географическая картография. М: Дата+, 2012; с. 158–166.

31. Балановский О. П. Генофонд Европы. М.: Товарищество научных изданий КМК, 2015.