

INTERACTIONS BETWEEN GENE POOLS OF RUSSIAN AND FINNISH-SPEAKING POPULATIONS FROM TVER REGION: ANALYSIS OF 4 MILLION SNP MARKERS

Balanovsky OP^{1,2,3} ✉, Gorin IO^{1,2}, Zapisetskaya YuS², Golubeva AA², Kostryukova EV⁴, Balanovska EV^{2,3}

¹ Vavilov Institute of General Genetics, Moscow, Russia

² Research Centre for Medical Genetics, Moscow, Russia

³ Biobank of North Eurasia, Moscow, Russia

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

This study explored the gene pools of Russian and Karelian populations of Tver region. Forty-one samples representing Tver Karels ($n = 11$) and Russians residing in the Western, Central and Eastern districts of Tver region ($n = 30$) were genotyped using a genome-wide panel of 4,559,465 SNPs. In order to investigate the phenomenon of genetic admixture between Slavic and Finnish-speaking populations, the obtained results were compared to the data on the Russian populations inhabiting the neighboring territories, Karels from Karelia and other North Eastern Europeans. Studying the gene pools of Russian populations with a genome-wide SNP panel is essential for cataloging their genetic diversity and identifying the distinct features of regional gene pools; in addition, it provides valuable data for practical pharmacogenomics and forensics. Using the principal component analysis, the ADMIXTURE method and D- and f3-statistics, we demonstrated that the gene pool of Tver Karels is closest to the gene pool of Karelian Karels, despite a long (300 to 500 years) history of living among the larger Russian population and the twentyfold population decline during the 20th century. At the same time, the gene pool of Tver Karels exhibits more pronounced similarity to the gene pool of the studied Russian populations than does any other Karelian population. The genetic admixture between Tver Russians and Tver Karels occurred due to a more intense gene flow from Russians to Karels whereas the gene flow from Karels to Russians was much weaker: Tver Russians turned out to be as genetically different from Karels as Pskov Russians. The genetic similarity of Tver Karels to Karelian Karels assessed with the autosomal SNP panel exhibits a slight shift towards the Russian gene pool and is consistent with the previously published analysis of Y-chromosome lineages in these populations that detected no admixture between Tver Karels and Russians.

Keywords: genome-wide genotyping, SNP, Illumina array, gene pool, Karelians, Russians, Tver region, Central Russia

Acknowledgement: we thank all the donors who took part in this study, the Biobank of North Eurasia for DNA collections and Napolskikh VV, the corresponding member of RAS, for his contribution to data interpretation.

Funding: the study was supported by the Russian Ministry and Science and Higher Education (Government Contact # 011–17 dated September 26, 2017). Genotyping and manuscript preparation were done under the *DNA-based identification* Research and Technology Project of the Union State. Bioinformatic analysis and interpretation of the obtained results were carried out under the State Assignment of the Russian Ministry of Science and Higher Education for Bochkov Research Centre for Medical Genetics.

✉ **Correspondence should be addressed:** Oleg P. Balanovsky
Gubkina, 3, Moscow, 119991; balanovsky@inbox.ru

Received: 10.10.2020 **Accepted:** 27.10.2020 **Published online:** 25.11.2020

DOI: 10.24075/brsmu.2020.072

ВЗАИМОДЕЙСТВИЕ ГЕНОФОНДОВ РУССКОГО И ФИННОЯЗЫЧНОГО НАСЕЛЕНИЯ ТВЕРСКОЙ ОБЛАСТИ: АНАЛИЗ 4 МЛН SNP-МАРКЕРОВ

О. П. Балановский^{1,2,3} ✉, И. О. Горин^{1,2}, Ю. С. Записецкая², А. А. Голубева², Е. С. Кострюкова⁴, Е. В. Балановская^{2,3}

¹ Институт общей генетики имени Н. И. Вавилова, Москва, Россия

² Медико-генетический научный центр, Москва, Россия

³ Биобанк Северной Евразии, Москва, Россия

⁴ Федеральный научно-клинический центр физико-химической медицины, Москва, Россия

Генофонды популяций Тверской области (русских и карел) изучены по широкогеномной панели из 4 млн аутомомных SNP-маркеров, типированной на суммарной выборке из 41 образца. Эти данные по популяции тверских карел ($n = 11$) и русских из западных, центральных и восточных районов Тверской области ($n = 30$) проанализированы на широком фоне русских популяций соседних областей, карел Карелии и других популяций Северо-Восточной Европы с целью изучения феномена взаимопроникновения генофондов славянского и финноязычного населения. Такое изучение генофондов населения России по наиболее обширной из существующих широкогеномных панелей важно для каталогизации геномного разнообразия населения России и характеристики региональных генофондов и имеет практическое применение в фармакогеномике и судебной медицине. Методами главных компонент, ADMIXTURE, d- и f3-статистик показано, что генофонд тверских карел, несмотря на их проживание среди преобладающего русского населения в течение 3–5 веков и 20-кратное сокращение численности в течение последнего столетия, сохраняет наибольшую близость к генофонду карел Карелии. Но при этом генофонд тверских карел более сходен с русским генофондом, чем генофонд других карельских популяций. Сближение генофондов русских и карел Тверской области происходит за счет более интенсивного потока генов от русских к карелам и при малозаметном потоке генов от карел к русским: тверские русские оказались столь же генетически отличны от карел, как, например, псковские. Сходство тверских карел с карелами Карелии по аутомомным маркерам (при небольшом смещении в сторону русского генофонда) согласуется с опубликованными данными по Y-хромосоме (отсутствие детектированного смещения тверских карел с русскими).

Ключевые слова: широкогеномные панели, SNP-маркер, чип Illumina, генофонд, карелы, русские, Тверская область, центральная Россия

Благодарности: мы благодарим всех доноров образцов, которые принимали участие в данном исследовании, АНО «Биобанк Северной Евразии» за предоставление коллекций ДНК и члена-корреспондента РАН В. В. Напольских за консультации при интерпретации результатов.

Финансирование: исследование выполнено при финансовой поддержке Министерства науки и образования РФ (Госконтракт # 011–17 от 26.09.2017) в рамках научно-технической программы Союзного государства «ДНК-идентификация» (работы по генотипированию), Государственного задания Министерства науки и высшего образования РФ для Медико-генетического научного центра им. Н. П. Бочкова (биоинформатический анализ данных), гранта Российского фонда фундаментальных исследований № 20-09-00479 а (анализ генеалогической информации, интерпретация результатов, написание текста).

✉ **Для корреспонденции:** Олег Павлович Балановский
ул. Губкина, д. 3, г. Москва, 119991; balanovsky@inbox.ru

Статья получена: 10.10.2020 **Статья принята к печати:** 27.10.2020 **Опубликована онлайн:** 25.11.2020

DOI: 10.24075/vrgmu.2020.072

The city of Tver and the adjacent territories situated at the border between Central and Northwest Russia played an important role in the country's history in general and the interactions between Russian and Western Finnish-speaking populations in particular. Before Slavic colonization, which started around the middle of the 1st millennium, this area was inhabited by Finno-Ugric tribes, predominantly by the Merya. In the early 12th century, a settlement of merchants and craftsmen, which came to be known as Tver, emerged in the estuary of the Tvertsa river. In the middle of the 13th century, Tver rose as one of the 3 Grand Russian Principalities of the Mongol invasion period. For two centuries, Tver was vying with Moscow for the right to unify Russian lands under its rule, maintaining its status as a center of attraction for human resources.

The 15–16th centuries marked the beginning of Karelian migration from the Karelian Isthmus and the areas adjacent to Lake Ladoga lying to the North-East of today's Tver region. In the wake of the Russo-Swedish war, the migration intensified dramatically. By 1670, as many as 25,000 to 30,000 Orthodox Karelians had fled to the lands of Tver. The refugees settled in the areas devastated by famine and chaos during the Time of Troubles. They started their own closely built settlements away from Russian villages. The subsequent waves of Karelian migration were not so massive [1, 2]. Thus, the exodus of Karels from their homeland produced an ethnographic group of Tver Karels who maintained their native Karelian language (the Finnic subgroup of Finno-Ugric languages) throughout centuries. In 1937, the Karelian national district was established with an administrative center in Likhoslavl. Two years later, in 1939, it was abolished, and the activists of the Karelian movement were arrested. This might have driven some Tver Karels to rethink their ethnic self-identification. According to the censuses, the Karelian population shrank in half during the 20th century, declining from 150,000 people in 1930 (of whom 95% spoke Karelian) to 7,000 in 2010 [3]; still, Karels remained within the borders of their habitat [4].

The fact that 2 ethnic groups, Tver Russians and Tver Karels have been living side by side for over 3 centuries raises the question of possible genetic admixture between these two populations. This question was partially answered in our previous publication on the analysis of the Tver Karelian gene pool, which we conducted using a panel of 49 lineage-informative Y-chromosome SNPs for Eastern European populations [5]. We convincingly demonstrated the genetic similarity of Y-chromosomes between Tver Karels and the indigenous populations of Northeast Europe, especially South Karels and Karelian Veps. The study showed that Tver Karels retained their ancestral Y-chromosomal gene pool throughout more than 10 generations in spite of the dramatic twentyfold population decline and years of mingling with the Russian population. The massive population decline might be explained by a change in the self-identification of Tver Karels and their assimilation by the Russian population. If this explanation is valid, it would be natural to expect that the genome of today's Tver Russians will contain an increased proportion of Y-chromosomal variants typical of Northeastern European populations in general and Karels in particular. It is known that interethnic marriages between neighboring ethnic groups produce a more stable Y-chromosomal gene pool compared to the autosomal gene pool because the majority of such marriages are patrilocal (a woman moves into her husband's home village), i.e. resulting in the geographical migration of mitochondrial DNA and autosomes and no geographical migration of Y chromosomes. Both of these factors might be the reason why the autosomal gene pools of Tver Karels and

Tver Russians were hugely mutually influential and became more homogenous than the Y gene pools.

Studying the gene pools of indigenous peoples with a genome-wide SNP panel is essential for cataloging the genetic diversity of the Russian population and identifying the distinct features of regional gene pools. These data are important for pharmacogenomics and forensics. The majority of existing pharmacogenetic protocols have been designed for European populations and may not produce a satisfactory result for the Russian populations which carry other allelic variants; besides, the frequencies of well-studied alleles differ significantly between the ethnic groups living in Russia, similarly to the populations of East Asia and Africa [6, 7]. The studies investigating the frequencies of pharmacogenetic markers in Russian populations have been summarized in a recent review [8]. Data on the gene pools are instrumental in forensic analysis in cases when there is a need to identify the origin of a person using only trace amounts of DNA. Currently, there are a few systems for DNA-based identification, and a few others are still in development, but the key thing is the availability of genetic data on ancestral populations [9, 10].

The aim of this study was to characterize the gene pools of Tver Karels and Tver Russians using a genome-wide panel of 4 million autosomal SNPs and to analyze the gene flow between these 2 populations. Conducted on a large dataset of samples from European Russia, the analysis will serve a more general purpose of exploring the interactions between Slavic and Finnish-speaking populations.

METHODS

This field study of the Russian and Karelian populations inhabiting Tver region followed the method detailed in [11]. The study included only unrelated individuals who did not share a common grandparent (according to the information they provided in the questionnaire) and whose ancestors from at least 2 previous generations had been born in Tver region, self-identified as Russian or Karelian and had no memories of other ethnicities in their ancestry.

Participants were eligible for the study if 1) both of their grandmothers and both of their grandfathers identified as Russian or Karelian; 2) they were willing to give informed consent to participate.

The following exclusion criteria were applied: poor DNA quality or insufficient DNA amount for whole-genome genotyping.

The population of Tver Karels was represented by 11 individuals whose ancestors came from the central part of Tver Karels' habitat, including Likhoslavl district ($n = 4$), Maksatikha district ($n = 1$), Spirovo district ($n = 2$), and Rameshki ($n = 4$) district. In 1930, the Karelian population of these 4 regions numbered 88,000, amounting to 58% of the total population of Tver Karels (the distribution was as follows: 15% resided in Likhoslavl, 19% in Maksatikha, 8% in Spirovo, and 16% in Rameshki districts). In 2010, there were only 5,000 Karels living in these 4 districts, constituting 78% of the total population of Tver Karels (36% in Likhoslavl, 13% in Maksatikha, 15% in Spirovo, and 14% in Rameshki districts).

Tver Russians were represented by 30 individuals. Since we aimed to study interactions between the Russian and Karelian gene pools, the plan was to compile the Russian subset in such a way that it would represent the areas that did not overlap geographically with the habitat of Tver Karels but were in the vicinity to it. This strategy appears to be optimal for determining the intensity of gene flow from Russians to

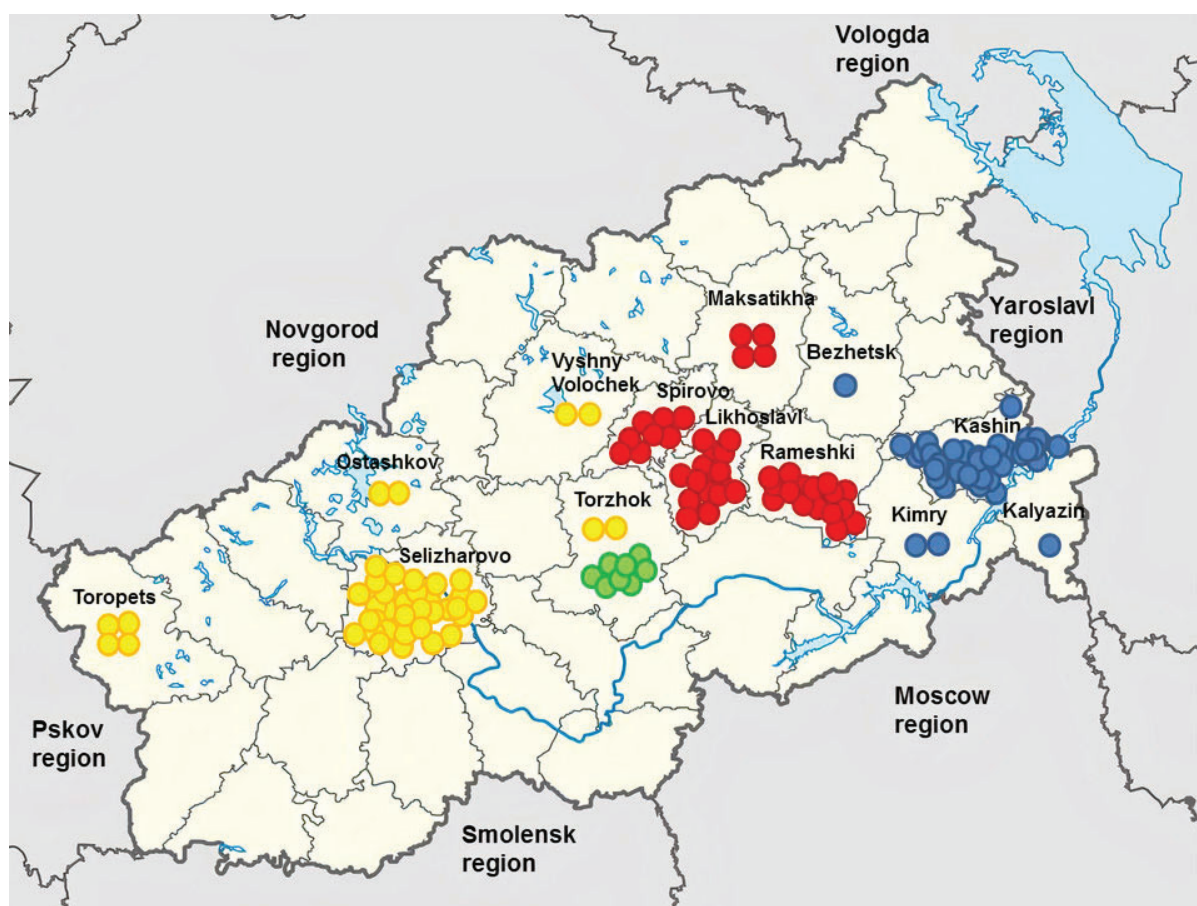


Fig. 1. The geographical map of the studied Tver region populations. Circles represent places of birth for each of 4 ancestors (2 grandmothers and 2 grandfathers) of each participant; Tver Karels are shown in red; East Tver Russians are shown in blue; South Tver Russians are shown in green; West Tver Russians are shown in yellow

Karels because the degree of genetic variation between Tver Karels and the populations of remote Russian settlements without a past history of direct contact with Karelians might turn out to be too high due to the genetic differences existing between Russian populations, whereas the degree of genetic variation between Karels and Russians living in Karelian villages might be too low as the Russian villagers might be actually the descendants of the Karels who once started to self-identify as a different ethnicity. For extra control, we studied several Russian populations (instead of one) living at various distances from the habitat of Tver Karels. The Eastern population of Tver Russians occupies the area neighboring the habitat of Tver Karels (Fig. 1). In the autosomal gene pool analysis, this population was represented by 13 individuals, all born in the Kashin district of Tver region. The Western population of Tver Russians was selected in such a way that geographically it was at a greater distance from the habitat of Tver Karels than the Eastern Russian population. The Western subset comprised 15 individuals born in the Selizharovo district of Tver region. Two individuals from the Torzhok district to the south of Likhoslavl, the administrative center of Tver Karels, were allocated to a separate Southern group. Thus, a total of 41 samples collected from the residents of Tver region were genotyped using a genome-wide SNP panel. Fig. 1 shows the places of origin for each of 4 grandparents of every participant.

The gene pool of Tver Russians and Karels was compared to the gene pools of the Russian populations from neighboring territories (Archangelsk, Vologda, Voronezh, Kursk, Pskov, Novgorod, Smolensk, and Yaroslavl) and South and North Karels residing in Karelia ($n = 16$). In total, 27 Karelian, 100 Russian and a number of other East European genomes

(Belarusian, Vepsian, Votian, Izhora Ingrian, Lithuanian, and Ukrainian) were analyzed using the same genome-wide SNP panel. The majority of the listed populations were previously studied using the panels of Y-chromosome markers [5, 12, 13].

All DNA samples, including those collected in Tver region and those representing the group of comparison, were genotyped using a panel by Illumina consisting of 4.5 million SNPs, an Infinium Omni5Exome-4 v1.3 BeadChip Kit (Illumina; USA) and an iScan genotyping system (Illumina; USA). Primary data analysis and quality control were carried out in GenomeStudio v2011.1 (Illumina; USA). For all the studied samples, the CallRate value was at least 0.99. Thus, genotypes were generated for 4, 559 465, SNP markers.

The obtained genotypes were uploaded to the GG-base [14] and are now available for downloading (RussiansTverKashin, RussiansTverSelizharovo, RussiansTverTorzhok, TverKarelians).

Primary data analysis was performed using the classic principal component analysis, which allowed us to identify the overall structure of the studied gene pools. Genetic drift between the studied populations was measured with f_3 -statistics. The D -statistic was employed to identify the direction of gene flow between the studied populations.

Data filtering was done in PLINK 1.9 [15, 16]. The applied filters are described below.

Prior to PCA, we filtered out SNPs with the genotyping rate of $< 95\%$ (geno 0.05) and the minor allele frequency of $< 1\%$ (maf 0.01); we also excluded samples with $> 10\%$ missing genotype rates (mind 0.1); SNPs that were in high linkage disequilibrium with each other ($r^2 > 0.2$) were pruned using a sliding window of 1,500 SNPs shifting 150 SNPs at a time (indep-pairwise 1500 150 0.2). The output files contained

274,036 SNPs and 126 (of the initial 131) samples. Principal components were computed in EIGENSTRAT smartpca [17, 18] with 5 outlier removal iterations. The results generated by smartpca were visualized using Python 3, pandas [19, 20], matplotlib [21] and seaborn [22] libraries.

To prepare the data for ADMIXTURE analysis, the same filters were applied (mind 0.1, geno 0.05, maf 0.01). After that, SNPs pairs with $r^2 > 0.2$ were pruned. The resultant dataset was analyzed in ADMIXTURE v1.3.0 [23]; cross validation errors were calculated for each k .

F3 statistics measure the genetic drift between two populations, i.e. the degree of their genetic ancestry relative to an outgroup. F3 statistics were computed in qp3Pop (AdmixTools) [24] using a Yoruba population from the 1000 Genomes Project as an outgroup [25]. Apart from the Yoruba dataset, the analysis covered 668 samples genotyped for 3,757,004 markers. The following filters were applied: mind 0.1, geno 0.05, maf 0.01; SNP pairs with $r^2 > 0.5$ were excluded from the dataset. The resultant dataset included 1,144,136 SNPs in a total of 635 samples.

The D-statistic is a tool for detecting genetic admixtures between 4 populations. In its classic version, the most genetically distant population (an African one) serves as an outgroup, and the test identifies the direction of gene flow between 3 remaining populations. The calculations were performed in qpDstat (AdmixTools) using a Yoruba population as an outgroup. In total, 748 samples and 3,757,004 SNPs were analyzed. The following filters were applied: mind 0.05; geno 0.2; maf 0.01; $r^2 > 0.6$. The resultant dataset included 1,355,253 SNPs in 633 samples.

RESULTS

The position of Tver Russians and Tver Karels in the PCA space which was constructed based on the genome-wide panel of 4,500,000 SNPs is shown in Fig. 3. The Tver Karelian sample is closer to the samples of Karelian Karels and at some distance from the analyzed Russian populations (Tver, Novgorod, Vologda and Yaroslavl). Only one sample of Tver Karels is genetically close to Vologda Russians. All other samples of Tver Karels cluster together, demonstrating little genetic variation. This clustering is consistent with the results of our previous study which analyzed Y-chromosome lineages [5] and concluded that the community of Tver Karels retained its ancestral gene pool.

At the same time, the analysis of the autosomal markers included in the panel reveals a shorter genetic distance between Tver Karels and Russians. Fig. 2 shows that samples of North Karels, South Karels, Tver Karels, and Russians together form a clinal gradient. The closest to Karels are Russians from Vologda region; Tver, Pskov and Central Russian populations constitute a single genetic "cloud". Genetic differences between the Western and Eastern groups of Tver Russians are slight yet pronounced and consistent with their geography: the Western population of Tver Russians shares its genetic space with Pskov samples, which is seen in the PC plot, whereas the Eastern population of Tver Russians (Kashin district) remains on the periphery. Remarkably, two samples from the Eastern population join the Novgorod-Yaroslavl group, which the second principal component differentiates from the rest of the Russian populations (Fig. 2.)

According to PCA, the highest degree of similarity exists between Tver and Karelian Karels but not between Tver Karels and the studied Russian populations. Still, PCA results encourage a hypothesis that the genetic pools of Russian

and Tver Karels might be characterized by a slight degree of admixture between each other. Of 3 studied Karelian populations, only Tver Karels shifted towards Russians, whereas Tver Russians, similarly to other Russian populations analyzed in this paper, keep their genetic distance from any of the studied Karelian populations. This suggests that the most intense gene flow occurred from Russians to Karels and not the other way around. F3 statistics clarify the degree of genetic similarity between Tver Karels and Eastern European populations: the closest to Tver Karels are Baltic populations, including the Izhora (Inger), the Vote, South Karels, Veps, Lithuanians, and North Karels (listed in the descending order). Russian populations are more genetically distant from Tver Karels and can be arranged in the following descending order based on the degree of similarity: Pskov, Novgorod, West Tver, Smolensk, Kursk, East Tver, Yaroslavl, Vologda, Voronezh, and North East Arkhangelsk. Notably, the genetic similarity between Tver Russians and Tver Karels is far from being pronounced.

The ADMIXTURE analysis can qualitatively and quantitatively assess the contribution of ancestral populations to a studied genetic pool. With ADMIXTURE, it is possible to vary the number of populations k to detect common ancestral components with various degrees of fractionality.

At $k = 5$ (Fig. 3; Table), significant contribution is visible for only 2 components. Component A is shown in blue; its contribution is the greatest in the speakers of Uralic languages. Component B (Lithuanians, Belarusians, Ukrainians and most Russian populations) is shown in ochre. Component A prevails in Karelian Karels (85%; see Table). By contrast, component B is observed in a few individual samples representing this group but found in every sample of Tver Karels, comprising 41% of their genomes (see Table). Component B occurs twice as frequently in Tver Russians, making up 80% of their genomes. Thus, the results of the ADMIXTURE analysis at $k = 5$ do not contradict the hypothesis about partial gene flow from Russians to Tver Karels.

At $k = 6$ (Fig. 3) the picture becomes more detailed and complex, now showing the contribution of the bright yellow

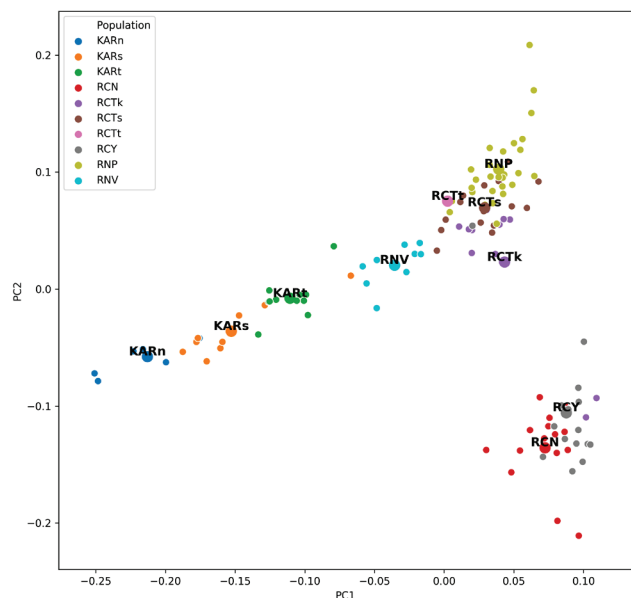


Fig. 2. The principal component plot showing genetic variance in the studied populations. KARn — North Karels, KARs — South Karels, KARt — Tver Karels, RCN — Novgorod Russians, RCTk — Tver Russians of Kashin (East), RCTt — Tver Russians of Torzhok (South), RCTs — Tver Russians of Selizharovo (West), RCY — Yaroslavl Russians, RNP — Pskov Russian, RNV — Vologda Russians. Individual samples are marked by small circles; centroids (centers of gravity for each population) are marked by larger circles of the same color

component C (Karelian genomes; see Table). In Karelian Karels, its contribution reaches 100%; it is twice as rare (52%) in Tver Karels and very rare in Tver Russians (8%) and Pskov Russians (4%), indicating that gene flow from these two groups to Karels was either insignificant or zero. Component C is present in Russians because all Eastern European populations share common ancestry. Surprisingly, component C is detected in other populations inhabiting the territories that neighbor Tver region, including the Russians of Novgorod (39%), Yaroslavl (30%) and Vologda (20%).

At $k = 8$ (see Fig. 3), the chart reflects the differentiation of component C. The bright yellow component (arbitrarily termed “West Finnish”) still looks influential in Karels and Vologda Russians (96% in Karelian Karels, 53% in Tver Karels, 20% in Vologda Russians; this contribution is designated as component E). However, its contribution to the genome of other Russian populations is minimal. Perhaps, the presence of this component in the genomes of Russian populations does not reflect their recent intermixing with Karels, but is the evidence of historically distant events like the origin of Russian populations from Slavs mingling with autochthonous Finnish-speaking tribes.

Thus, at $k = 8$ only Vologda Russians are characterized by a prominent (one-fifth of the genome) contribution of the Western Finnish component E. In other Russian populations, where component E is absent, a different component (shown in light gray, I) is observed. Its contribution is the greatest in Novgorod (91%) and Yaroslavl (90%) Russians, accounting for almost entire genome. Component I also constitutes over one-third of the genome in Tver (39%), Pskov (36%) and Vologda (34%) Russians. This “Novgorod” component also occurs in other studied populations of the Central and Southern Russia, making up at least 38% of their genomes.

Based on what proportion of the genome is represented by components I and K (arbitrarily termed “South Russian”), 2 groups of Tver Russians can be identified. Interestingly, these groups are not in accord with their geography genetically: component K is dominant relative to component I in the Western part of Tver region bordering on Novgorod region ($K/I = 63/27$), whereas in the Eastern population of Tver Russians the contributions of both components are equal ($K/I = 42/42$); in Central Tver, the “Novgorod” component I comprises the entire gene pool (100%).

DISCUSSION

This study was conducted using a genome-wide panel of autosomal SNPs. Its findings support the conclusion of our

previous study, in which we used a panel of Y-chromosome markers: the genetic distance between Tver Karels and Karelian Karels is closer than between Tver Karels and their Russian neighbors residing in Tver region [5]. Importantly, it was not only descriptive statistics (PCA, ADMIXTURE) but also D-statistics that underpinned this conclusion. Classically, the D-statistic (the f_4 -statistic) employs one African population as an outgroup. The method helps to understand the direction of gene flow between the 3 remaining populations; its results are considered reliable at $|Z| > 3$. The Z scores generated by the D-statistic (Yoruba, TverKarelians; SouthKarelians, TverRussians) for the Eastern and Western populations of Tver Russians were -6.9 and -5.0 , respectively. This proves that the gene pool of Tver Karels is closer to the gene pool of Karelian Karels than to the gene pool of Tver Russians. At the same time, there is more pronounced genetic similarity between Tver Karels and the studied Russian populations than between the studied Russian populations and Karels from South (and certainly North) Karelia. Z scores for Tver Karels become statistically significant if the analysis includes more southern (relative to Tver) populations of Russians. For example, the D statistic (Yoruba, RussiansSmolensk; TverKarelians, Karelians) produces $Z = -3.4$ for the Russian populations inhabiting the South of Smolensk region. This indicates that the gene pool of Tver Karels, which, on the whole, is similar to that of Karelian Karels, is reliably close to the gene pools of Smolensk Russians and other Russian populations.

To sum up, assuming that initially the ancestors of Tver Karels and Karelian Karels existed as a single population [1, 2, 4], D-statistics prove that later the ancestors of Tver Karels accepted the genetic contribution of populations inhabiting the southern territories of the East European plain. Eastern Europe has witnessed a lot of complex migration patterns, so the source of southern admixture in Tver Karels cannot be identified with absolute certainty; however, history suggests that the best candidates here are the Russian populations of Tver and the neighboring regions.

CONCLUSION

We have studied the gene pools of Tver Karels and Tver Russians using a panel of 4,500,000 autosomal SNPs and compared them to the samples of Karelian Karels and the inhabitants of Russian regions bordering on Tver (Pskov, Novgorod, Vologda, Yaroslavl). The applied statistical methods (PCA, ADMIXTURE, D- and f_3 -statistics) generated consistent results.

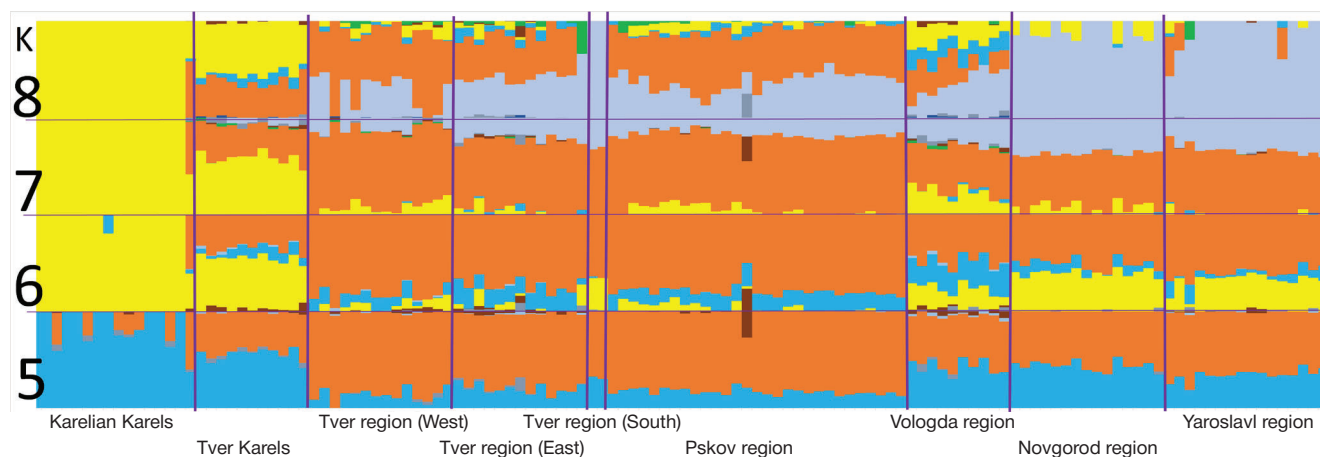


Fig. 3. The ADMIXTURE chart (contributions of ancestral components to the studied populations (at different k values). Profiles of individual samples are provided in separate columns; individual samples representing different populations are separated by vertical bars

Table. The contributions of ancestral ADMIXTURE components to the studied populations (at different *k* values)

K	Fractionality level	Component.	Ethnicity	Karels									Russians																	
				Karelia			Tver			Tver (west)			Tver (east)			Tver (south)			Pskov			Vologda			Novgorod			Yaroslavl		
				Sample size			n = 11			n = 14			n = 13			n = 2			n = 29			n = 10			n = 15			n = 16		
				Color code for Fig.3			MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX	MEAN	MIN	MAX
			%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	
5	A	Blue	85	37	100	55	41	63	13	0	24	20	10	29	31	30	33	17	11	25	42	30	52	43	34	53	34	18	45	
	B	Ochre	13	0	60	41	35	50	85	71	96	75	65	90	69	67	70	83	52	89	51	42	61	56	47	65	65	55	77	
6	C	Bright yellow	95	36	100	52	38	60	6	0	17	5	0	22	33	32	33	4	0	15	20	8	31	39	29	46	30	3	37	
	D	Blue	2	0	20	9	3	14	10	0	18	18	6	27	1	1	2	16	8	24	28	20	35	9	0	14	8	2	20	
	E	Ochre	3	0	56	35	27	45	81	68	97	73	62	84	65	64	66	79	50	86	47	37	60	51	41	62	60	49	72	
8	F	Bright yellow	96	38	100	53	40	59	7	0	19	3	0	16	0	0	0	5	0	13	20	10	30	9	0	23	1	0	13	
	G	Blue	0	0	3	8	2	13	2	0	7	6	0	15	0	0	0	3	0	11	18	12	22	0	0	4	1	0	7	
	H	Light gray	0	0	0	0	0	0	27	0	47	42	31	66	100	100	100	36	13	57	34	8	55	91	72	100	90	43	100	
	K	Ochre	3	0	58	37	29	47	63	36	98	42	0	54	0	0	0	54	36	70	24	0	38	0	0	0	6	0	41	

The gene pool of Tver Karles retains its similarity to the gene pool of Karelian Karels despite a long (300 to 500 years) history of living among the larger Russian population and the twentyfold population decline during the 20th century. At the same time, the gene pool of Tver Karels exhibits greater similarity to the Russian gene pool, in comparison with other analyzed Karelian populations. Having compared the findings of the analysis of autosomal SNP markers (a partial shift towards the Russian gene pool) and the previously obtained results of genotyping for Y-chromosome markers (no detected admixture between Tver Karels and Russians), we conclude that gene flow between

Russians and Tver Karels was predominantly determined by marriages between Karelian men and Russian women.

Demographic data (the sharp decline in the Tver Karelian population) and historical events suggest that Tver Karels changed their ethnic self-identification and were assimilated by the Russian population. Therefore, it would be logical to hypothesize that the genome of Tver Russian descendants of Karels who once changed their ethnic self-identification contains a greater Karelian genetic component. This, however, is not the case: Tver Russians are as genetically distant from Karels as Pskov Russians.

References

- Savinova AI, Stepanova UV. Karelskaya diaspora uzhnih rayonov Tverskogo Povolgya: istoria formirovaniya i istoricheskaya sud'ba. CARELiCA. 2018; 1 (19): 26–37. Russian.
- Savinova AI, Stepanova UV. Rasselenie karel v Verhnevolzhie v seredine — vtoroy polovine XVII v.: opit izucheniya s primeneniem gis-technologiy. Istoricheskaya informatika. 2018; 4: 57–72. Russian.
- Vishnevskiy AG, editor. Perepisi naseleniya Rossiyskoy Imperii, SSSR, 15 novykh nezavisimyykh gosudarstv. Demoskop Weekly. [Internet]. [cited 2020 Oct 9]. Available from: http://www.demoscope.ru/weekly/ssp/census_types.php?ct=6.
- Golovkin AN. Istoriya Tverskoy Karelii. Tver': Studiya-S, 2008; 432 s. Russian.
- Agdzhoyan AT, Daragan DM, Skhalyakho RA, Reutov PP, Balanovskiy OP, Balanovskaya EV, et al. Vozmozhnost' sokhraneniya genofonda v diaspore na primere tverskikh karel. Genetika. 2018; 54 (Application S): 91–94. DOI: 10.1134/S0016675818130027. Russian.
- Rajman I, Knapp L, Morgan T, Masimirembwa C. African Genetic Diversity: Implications for Cytochrome P450-mediated Drug Metabolism and Drug Development. EBioMedicine. 2017; 17: 67–74.

7. Jing L, Haiyi L, Xiong Y, Dongsheng L, Shilin L, Jin L, et al. Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy. *Journal of Medical Genetics*. 2014; 51 (9): 614–22.
8. Mirzaev KB, Fedorinov DS, Ivashchenko DV, Sychev DA. ADME pharmacogenetics: future outlook for Russia. *Pharmacogenomics*. 2019; 20 (11): 847–65.
9. Kidd KK, Speed WC, Pakstis AJ, Furtado MR, Fang R, Madbouly A, et al. Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Science International: Genetics*. 2014; 10: 23–32.
10. Nassir R, Kosoy R, Tian C, White PA, Butler LM, Silva G, et al. An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. *BMC Genetics*. 2009; 10: 39.
11. Balanovskaya EV, Agdzhoyan AT, Chukhryaeva MI, Markina NV, Balaganskaya OA, Balanovskiy OP, et al. Populyatsionnye biobanki: printsipy organizatsii i perspektivy primeneniya v genogeografii i personalizirovannoy meditsine. *Genetika*. 2016; (12): 1371–87. Russian.
12. Agdzhoyan AT, Skhalyakho RA, Balaganskaya OA, Kozlov SA, Palipana SD, Balanovskiy OP, et al. Genofond novgorodtsev: mezhdru severom i yugom. *Genetika*. 2017; 53 (11): 1338–1348. Russian.
13. Chuhriaev MI, Pavlova ES, Napolskih VV, Garin EV, Balanovsky OP, Balanovska EV, et al. Sohranilis' li sledy finno-ugorskogo vliyaniya v genofonde russkogo naseleniya Yaroslavskoy oblasti? Svidetel'stva Y-chromosomi. *Genetika*. 2017; 53 (3): 378–89.
14. GG-base [Internet]. [cited 2020 Oct 9]. Available from: <https://www.gg-base.org/>.
15. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*. 2007; 559–75.
16. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. 2009; (19): 1655–64.
17. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015 Feb 25; 4 (7). Available from: DOI: 10.1186/s13742-015-0047-8.
18. Price A, Patterson N, Plenge R. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006; 38: 904–9.
19. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *PLOS Genetics* [Internet]. 2006; 2 (12): e190 [cited 2020 Oct 9]. Available from: <https://doi.org/10.1371/journal.pgen.0020190>.
20. McKinney W. Data structures for statistical computing in Python. *SciPy 2010: Proceedings of the 9th Python in Science Conference*; 2010 Jun 28 – Jul 3. Austin, Texas. Available from: <https://conference.scipy.org/proceedings/scipy2010/mckinney.html>.
21. Reback J, McKinney W, jbrockmendel, Augspurger T, Cloud F, Mehyar M, et al. pandas-dev/pandas: Pandas 1.0.3. Version 1.0.3 [software]. Zenodo. 2020 Mar 18 [cited 2020 Oct 9]. Available from: <http://doi.org/10.5281/zenodo.3715232>.
22. Hunter JD. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*. IEEE Xplore. 2007; 9 (3): 90–95.
23. Waskom M, Botvinnik O, O'Kane D, Hobson P, Lukauskas S, Qalieh A, et al. mwaskom/seaborn: v0.8.1 (September 2017). Version 0.8.1 [software]. Zenodo. 2017 Sep 3 [cited 2020 Oct 9]. Available from: <http://doi.org/10.5281/zenodo.883859>.
24. Alexander DH, Novembre J, Lange K. ADMIXTURE Software. Version 1.3.0 [software]. 2020 May 3 [cited 2020 Oct 9]. Available from: <http://dalexander.github.io/admixture/index.html>.
25. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient Admixture in Human History. *GENETICS*. 2012; 192 (3): 1065–93.
26. Auton A, Abecasis G, Altshuler D, et al. A global reference for human genetic variation. *Nature*. 2015; (526): 68–74.

Литература

1. Савинова А. И., Степанова Ю. В. Карельская диаспора южных районов Тверского Поволжья: история формирования и историческая судьба. *CARELiCA*. 2018; 1 (19): 26–37.
2. Степанова Ю. В., Савинова А. И. Расселение карел в Верхневолжье в середине — второй половине XVII в.: опыт изучения с применением гис-технологий. *Историческая информатика*. 2018; 4: 57–72.
3. Вишнеvский А. Г, редактор. Переписи населения Российской Империи, СССР, 15 новых независимых государств. *Демоскоп Weekly* [Internet]. [cited 2020 Oct 9]. Available from: http://www.demoscope.ru/weekly/ssp/census_types.php?ot=6.
4. Головкин А. Н. История Тверской Карелии. Тверь: Студия-С, 2008; 432 с.
5. Агджоян А. Т., Дараган Д. М., Схаляхо Р. А., Реутов П. П., Балановский О. П., Балановская Е. В. и др. Возможность сохранения генофонда в диаспоре на примере тверских карел. *Генетика*. 2018; 54: 91–94.
6. Rajman I, Knapp L, Morgan T, Masimirembwa C. African Genetic Diversity: Implications for Cytochrome P450-mediated Drug Metabolism and Drug Development. *EBioMedicine*. 2017; 17: 67–74.
7. Jing L, Haiyi L, Xiong Y, Dongsheng L, Shilin L, Jin L, et al. Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy. *Journal of Medical Genetics*. 2014; 51 (9): 614–22.
8. Mirzaev KB, Fedorinov DS, Ivashchenko DV, Sychev DA. ADME pharmacogenetics: future outlook for Russia. *Pharmacogenomics*. 2019; 20 (11): 847–65.
9. Kidd KK, Speed WC, Pakstis AJ, Furtado MR, Fang R, Madbouly A, et al. Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Science International: Genetics*. 2014; 10: 23–32.
10. Nassir R, Kosoy R, Tian C, White PA, Butler LM, Silva G, et al. An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. *BMC Genetics*. 2009; 10: 39.
11. Балановская Е. В., Агджоян А. Т., Чухряева М. И., Маркина Н. В., Балаганская О. А. Балановский О. П. и др. Популяционные биобанки: принципы организации и перспективы применения в генеогеографии и персонализированной медицине. *Генетика*. 2016; (12): 1371–87.
12. Агджоян А. Т., Схаляхо Р. А., Балаганская О. А., Козлов С. А., Палипана С. Д., Балановский О. П. и др. Генофонд новгородцев: между севером и югом. *Генетика*. 2017; 53 (11): 1338–48.
13. Чухряева М. И., Павлова Е. С, Напольских В. В., Гарин Э. В., Балановский О. П., Балановская Е. В. и др. Сохранились ли следы финно-угорского влияния в генофонде русского населения Ярославской области? Свидетельства Y-хромосомы. *Генетика*. 2017; 53 (3): 378–89.
14. GG-base [Internet]. [cited 2020 Oct 9]. Available from: <https://www.gg-base.org/>.
15. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*. 2007; 559–75.
16. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. 2009; (19): 1655–64.
17. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015 Feb 25; 4 (7). Available from: DOI: 10.1186/s13742-015-0047-8.
18. Price A, Patterson N, Plenge R. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006; 38: 904–9.
19. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *PLOS Genetics* [Internet]. 2006; 2 (12): e190 [cited 2020 Oct 9]. Available from: <https://doi.org/10.1371/journal.pgen.0020190>.

- pgen.0020190.
20. McKinney W. Data structures for statistical computing in Python. SciPy 2010: Proceedings of the 9th Python in Science Conference; 2010 Jun 28 – Jul 3. Austin, Texas. Available from: <https://conference.scipy.org/proceedings/scipy2010/mckinney.html>.
 21. Reback J, McKinney W, jbrockmendel, Augspurger T, Cloud F, Mehyar M, et al. pandas-dev/pandas: Pandas 1.0.3. Version 1.0.3 [software]. Zenodo. 2020 Mar 18 [cited 2020 Oct 9]. Available from: <http://doi.org/10.5281/zenodo.3715232>.
 22. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. IEEE Xplore. 2007; 9 (3): 90–95.
 23. Waskom M, Botvinnik O, O’Kane D, Hobson P, Lukauskas S, Qalieh A, et al. mwaskom/seaborn: v0.8.1 (September 2017). Version 0.8.1 [software]. Zenodo. 2017 Sep 3 [cited 2020 Oct 9]. Available from: <http://doi.org/10.5281/zenodo.883859>.
 24. Alexander DH, Novembre J, Lange K. ADMIXTURE Software. Version 1.3.0 [software]. 2020 May 3 [cited 2020 Oct 9]. Available from: <http://dalexander.github.io/admixture/index.html>.
 25. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient Admixture in Human History. GENETICS. 2012; 192 (3): 1065–93.
 26. Auton A, Abecasis G, Altshuler D, et al. A global reference for human genetic variation. Nature. 2015; (526): 68–74.