The Gene Pool of the Belgorod Oblast Population: Description of the "Genetic Landscape" of 22 District Populations

I. N. Sorokina[a], M. I. Churnosov[a], and E. V. Balanovska[b]

[a] *Belgorod State University, Belgorod, 308015 Russia;*
e-mail: sorokina@bsu.edu.ru

[b] *Medical Genetic Research Center, Russian Academy of Medical Sciences, Moscow, 115478 Russia*

Abstract—Data on the frequencies of all (50412) surnames in a total population of 849 399 people have been treated by various methods of multivariate statistics (cluster analysis, multidimensional scaling, and factor analysis) to show that 22 district populations of the Central Chernozem region of Russia form a definite, ordered system of population groups. All raions (administrative districts) of Belgorod oblast (administrative region) have been grouped into four clusters corresponding to the actual geographic locations of the populations. Districts of the eastern cluster are characterized by very specific spectrum and frequencies of surnames; districts of the western cluster differ in a high proportion of Ukrainian surnames

INTRODUCTION

Different approaches allowing the population subdivision and genetic relationship between populations to be analyzed are used in studying the genetic structure of human populations. The approaches based on the estimation of genetic distances between populations [1, 2] followed by analysis using multivariate statistical methods [3] are among the most widely used ones. Matrices of genetic distances serve as the basis for cluster analysis (a very flexible method [1]), whose results are used to draw dendrograms. To get a clearer idea of the relationships between populations, the "genetic landscape" of a region is drawn on the basis of the corresponding dendrograms [4]. The landscape is a cartographic diagram where elementary populations belonging to the same clusters are connected with lines forming equidistant figures. The method of describing the genetic structure of populations with the use of equidistant figures successively uniting populations according to their genetic distances from one another, thus forming the "genetic landscape" of the population studied may be efficiently used both for determining the boundaries and sizes of elementary populations [4] and for estimating the effect of external factors, such as shift of administrative territorial boundaries and unusual immigration flows [4, 5], on the population genetic structure.

Data on genetic distances treated by multivariate statistical methods have been used for describing genetic relationships between populations in the Volga–Ural region [6]; Kostroma [3] and Kirov [7] oblasts (administrative regions); republics of Adygea [5, 7, 8], Marii El [5], Chuvashia [5]; and other regions [5, 8].

Here, we report the results of taxonomic analysis of the total population of 20 raions (administrative districts) constituting Belgorod oblast, Rep'evka raion of Voronezh oblast, and Pristen' raion of Kursk oblast at the level of elementary populations (district) on the basis of anthroponymic data.

MATERIALS AND METHODS

Estimation of genetic distances and taxonomic analysis were carried out at the level of districts (an elementary population in the 1990s [9]). Twenty-two district populations of southern central Russia, including all the 20 districts of Belgorod oblast (except for Belgorod raion, which was found [10] not to be an elementary population), Pristen' raion of Kursk oblast, and Rep'evka raion of Voronezh oblast. The sample included the populations of two districts adjacent to Belgorod oblast (Pristen' and Rep'evka raions) because our earlier detailed analysis of the marriage-migration characteristics of these two district populations and several district populations of Belgorod oblast with the use of Malecot's isolation-by-distance model showed considerable differences between the time courses of their population characteristics: the genetic distances between the populations of Prokhorovka raion of Belgorod oblast and Pristen' raion of Kursk oblast and between those of Krasnoe raion of Belgorod oblast and Rep'evka raion of Voronezh oblast have increased and decreased, respectively, during the past 50 years. Therefore, we used another type of markers (surnames) and another level of population organization (all districts of Belgorod oblast) to test (1) whether these differences would still hold and, if so, how considerable they would be and (2) how these characteristics of dif-

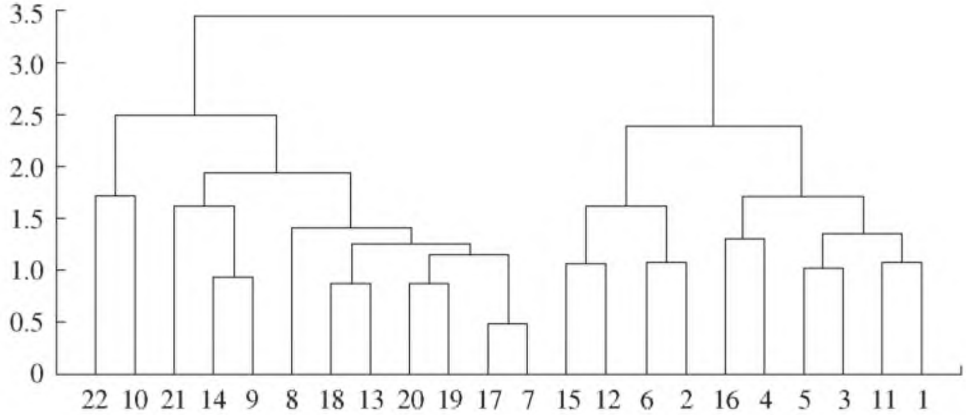

Fig. 1. The dendrogram of genetic relationships between 20 raions (districts) of Belgorod oblast, Pristen' raion of Kursk oblast, and Rep'evka raion of Voronezh oblast constructed by Ward's method. Raions: 1, Alekseevka; 2, Borisovka; 3, Valuiki; 4, Veidelevka; 5, Volokonovka; 6, Graivoron; 7, Gubkin; 8, Ivnya; 9, Korocha; 10, Krasnoe; 11, Krasnogvardeiskoe; 12, Krasnaya Yaruga; 13, Novyi Oskol; 14, Prokhorovka; 15, Rakitnoe; 16, Roven'ki; 17, Staryi Oskol; 18, Chernyanka; 19, Shebekino; 20, Yakovlevskii (Belgorod oblast); 21, Pristen' (Kursk oblast); 22, Rep'evka (Voronezh oblast).

ferentiation of local populations would fit the "genetic landscape" of a large regional population.

The material of the study consisted of data on the distribution of all surnames (50412) in a total population of 849 399 people. The frequencies of all surnames were used to calculate the matrices of genetic distances by means of the DJ Genetic software (version 0.03 beta) developed by Yu.A. Seregin and E.V. Balanovska in the Medical Genetic Research Center of the Russian Academy of Medical Sciences. The distances between elementary populations were calculated using Nei's method of comparison of populations with respect to the frequencies of the alleles of polymorphic markers [12, 13], which is generally accepted in population genetics. In the given case, surnames (quasi-genetic markers) were analogs of alleles.

The resultant matrix of genetic distances was used for cluster analysis by means of the Statistica 6.0 software. A hierarchical agglomerative procedure was used to form the clusters. Dendrograms were drawn by two methods: weighted-average linkage and Ward's method [1]. We used a dendrogram corresponding to the results obtained by other methods of multivariate analysis, namely, multidimensional scaling and facto analysis.

The results of cluster analysis (the dendrogram) were used to construct the diagram of the genetic landscape of the populations studied. The graphic diagram of this landscape was a projection of the dendrogram based on the genetic distance matrix where elementary populations belonging to the same clusters were grouped to form equidistant figures. Equidistant lines were drawn through a specific genetic distance unit in such a way that the number of grouping levels was no greater than 10, because a larger number of levels makes the diagram too cumbersome for efficient analysis [8].

The matrices of genetic distances were used for multidimensional scaling. To estimate the quality of its results, we used the stress index (S_0 (≤ 0.10)), the alienation coefficient (K_0), and Shepard's diagram [1]. The calculated correlation matrices served as the basis for factor analysis using the principal factor method. To determine the number of significant principal factors, we used Kaiser's criterion and Kettel's rejection criterion. To make the data more illustrative, they were presented graphically, with the populations studied positioned in a two- or three-dimensional space of the detected principal factors.

To estimate the role of geographic distances in genetic structural subdivision of populations, we carried out correlation analysis of the matrices of genetic and geographic distances (using Spearman's rank-order correlation coefficient). For this purpose, a map of Belgorod oblast (scale, 1 : 400000) was used to calculate the geographic distances between the district populations studied.

RESULTS AND DISCUSSION

Estimation of Genetic Distances and Taxonomic Analysis of Populations

We calculated the matrix of genetic distances from the frequencies of all surnames and then used cluster analysis to construct a dendrogram characterizing the genetic relationships between 22 district populations studied (Fig. 1).

Analysis of the dendrogram showed that the populations formed four groups of clusters. The first cluster of district populations was the largest; it comprised ten districts (Yakovlevskii, Shebekino, Prokhorovka, Korocha, Ivnya, Chernyanka, Novyi Oskol, Staryi Oskol, and Gubkin raions of Belgorod oblast and Pristen' raion of Kursk oblast). These districts have common

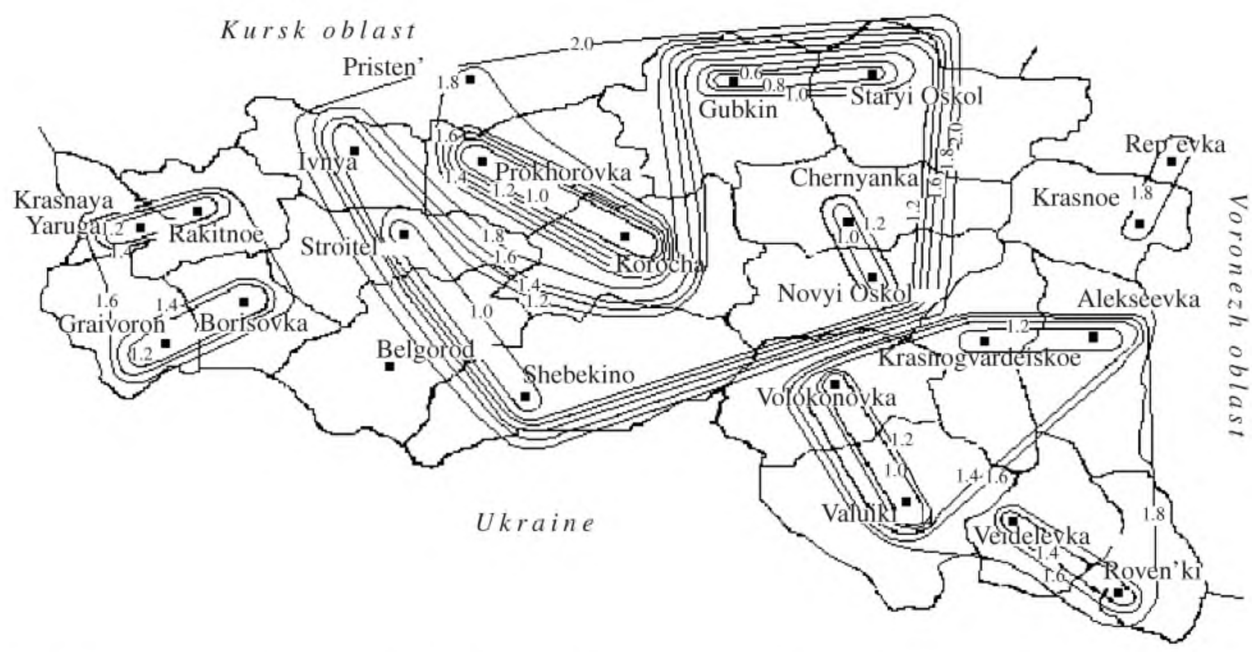

Fig. 2. The diagram of the genetic relationships between district populations of Belgorod oblast (according to the data on the distribution of all surnames).

boundaries and are located mostly in the center of Belgorod oblast. The second largest cluster comprised Volokonovka, Valuiki, Krasnogvardeiskoe, Alekseevka, Roven'ki, and Veidelevka raions located in southeastern Belgorod oblast, which also had common boundaries. The third cluster comprised western districts of Belgorod oblast (Krasnaya Yaruga, Rakitnoe, Graivoron, and Borisovka raions), also with common boundaries. The population of Krasnoe raion of Belgorod oblast was genetically close to Rep'evka raion of Voronezh oblast (they also have a common boundary), which was confirmed by their clustering into a separate group (the fourth cluster).

The results of cluster analysis were used to construct a diagram of the genetic landscape of all analyzed populations (Fig. 2). Equidistant lines were drawn at intervals of 0.2 units of genetic distance. The spatial distribution of the results of cluster analysis showed that district populations of the Central Chernozem region of Russia form a definite, ordered system of population groups. All the districts were divided into four relatively independent clusters. Note that the equidistant figures based on the results of cluster analysis of 22 districts did not intersect and described the actual geographic positions of the district populations studied. This is also evidenced by the significant positive Spearman's correlation coefficient between the matrices of genetic and geographic distances ($\rho = 0.86, p < 0.001$).

Other methods of multivariate statistics, namely, multidimensional scaling (based on the genetic distance matrix) (Fig. 3) and factor analysis using the principal component method (based on the correlation matrix) (Fig. 4), yielded almost identical results that entirely agreed with the results of cluster analysis. All analyzed district populations in three-dimensional space (Figs. 3, 4) clustered into four groups. In the plots showing the results of multidimensional scaling (Fig. 3) and factor analysis (Fig. 4), the populations clustering into separate groups are outlined with ovals.

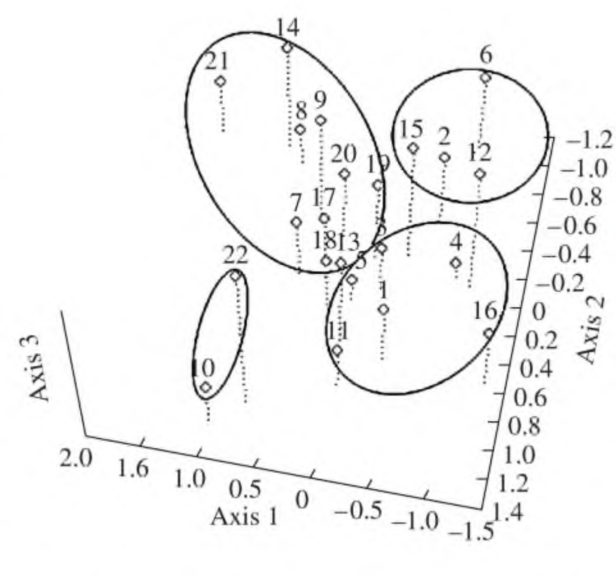

Fig. 3. The plot of three-dimensional scaling with the spatial positions of the district populations studied (48 iterations, the stress is $S_0 = 0.134$; the alienation coefficient is $K = 0.15$). Designations 1–22 are the same as in Fig. 1.

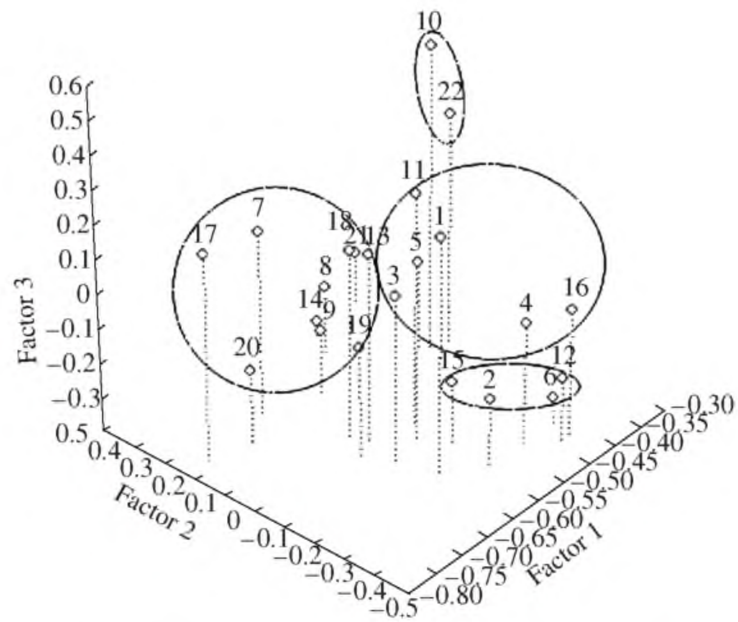

Fig. 4. The positions of the district populations studied in the space of three principal factors. Designations 1–22 are the same as in Fig. 1.

Our data on the genetic relationships between 22 district populations of the Central Chernozem region of Russia do not agree with the results obtained by G.I. El'chinova [8] suggesting that the use of surnames as a genetic marker in Russian rural populations allows an unbiased estimation of the genetic relationship between populations to be obtained only for one or, at the most, two or three rural districts, if their common boundary is very long. At the same time, our results indicate that surnames can be used for correct description of the "genetic landscapes" of large regional populations (more than 1.5×10^6 people) consisting of many (more than 20) district populations.

Note that the characteristics of the genetic relationships between four districts of the Central Chernozem region that were found earlier [11] with the use of Malecot's isolation-by-distance model entirely agree with the results of this study obtained using markers of a different type (quasi-genetic markers), which evidences for their reliability. Anthroponymic data show that Krasnoe raion of Belgorod oblast is the closest genetically to the Rep'evka raion of Voronezh oblast, these two districts forming a separate cluster (a decrease in the genetic distance between the populations of these districts during the past 50 years was found in [11]). In addition, according to the observed distribution of quasi-genetic markers, the genetic distance between Prokhorovka and Korocha raions of Belgorod oblast are the shortest ($d = 1.0$). These two districts form a primary cluster, and only at a level of $d = 1.8$ do they form a common cluster with Pristen' raion of Kursk oblast and at a level of $d = 2.0$, with the other seven districts of the central cluster (we found an increase in the genetic distance between Prokhorovka raion of Belgorod oblast and Pristen' raion of Kursk oblast in the period from the 1950s to the 2000s [11]).

Thus, different multivariate statistic analyses (cluster analyses, multidimensional scaling, and factor analysis) of the matrices of genetic distances and correlations calculated from the frequencies of all surnames have demonstrated that district populations of Belgorod oblast form a definite, ordered system of four population clusters corresponding to their actual geographic locations.

Surname Distributions in Four Population Clusters

We used the frequency distributions of all surnames to study the characteristics of "family name portraits" in each of the four groups of districts.

The mean surname frequency for each group of districts was calculated as an unweighted average of "district" surname frequencies (i.e., without the population sizes taken into account). Thus, the data were averaged over ten districts in group I (central), six districts in group II (southeastern), four districts in group III (western), and two districts in group IV (eastern). After that, the surnames in each group were arranged in the order of decreasing frequency, and the 50 most frequent surnames were selected (table). In the table, the surname frequencies (P_i) are indicated in the form $P_i \times 10^2$, and each surname is assigned an ordinal number according to its place in the total regional list of surnames where the frequencies were calculated as unweighted averages over the four groups of districts. This allowed us to evaluate the similarity of the "family name portrait" of each group to the total regional spectrum of the 50 most frequent surnames and the contribution of sur-

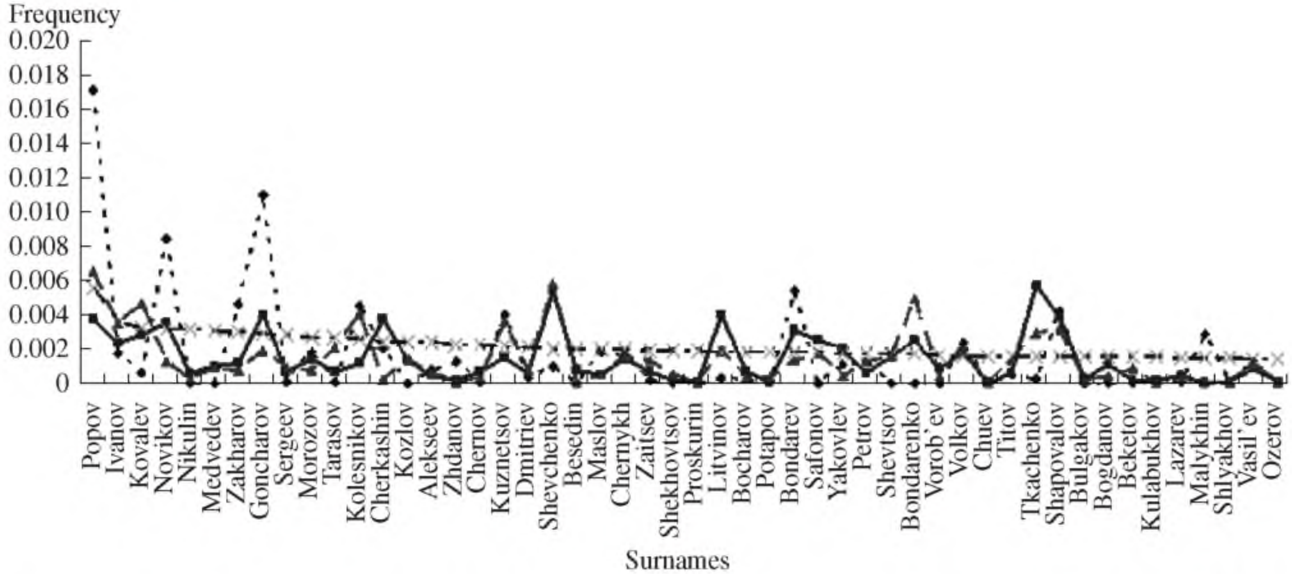

Fig. 5. Comparative analysis of the 50 most frequent surnames in group I of districts and the frequencies of these surnames in the other three groups. The 50 most frequent surnames in group I are listed in the order of decreasing frequency (crosses). The frequencies of the same surnames in the other three groups are shown: group II, triangles; group III, squares; group IV, diamonds.

names from each group to the total regional list, as well as to detect surnames characteristic of each of the four group and common surnames. The table also shows the regional mean surname frequencies calculated as unweighted averages for all the 22 districts.

Analysis of common surnames in different groups of districts showed that the first and second groups had 16 common surnames, and the first and third groups, as well as the second and third groups, had 15 common surnames (out of the 50 most frequent surnames). The number of surnames (among the 50 most frequent ones) that were common with any other group was the smallest in the fourth group of district populations; it had seven, six and five common surnames with the first, second, and third groups, respectively.

To study the main characteristics of the distributions of surnames in the analyzed groups of districts, we drew the plots where the frequencies of 50 surnames that were the most frequent in each particular group (Figs. 5–8) were compared with those in the remaining three groups. Figure 5 shows the plot and the list of the 50 surnames that were the most frequent in group I (ten district populations of central Belgorod oblast) arranged in the order of decreasing frequency. The other three curves show the frequencies of the same 50 surnames in the other three groups (the southeastern, western, and eastern ones or groups II, III, and IV, respectively). In a similar way, we plotted the curves of the surnames that were the most common in groups II (Fig. 6), III (Fig. 7), and IV (Fig. 8).

The curve of frequent surnames of group I (Fig. 5) has a plateau. The plots for the other three groups are markedly broken lines with drastic peaks and troughs reflecting the specificity of each group. The most frequent surnames in group I were Popov, Ivanov, Kovalev, Novikov, Nikulin, Medvedev, Zakharov, Goncharov, Sergeev, and Morozov. The surnames Nikulin, Medvedev, Sergeev, and Morozov were most frequently found in group I; Kovalev, in group II; and Popov, Novikov, Zakharov, and Goncharov, in group IV. Analysis of the plot shows that some surnames were almost equally frequent in all the four groups. There were only three such surnames out of the 50 most frequent ones: Chernykh, Volkov, and Vasil'ev.

Figure 6 shows the comparison of 50 surnames that were the most frequent in group II of districts (six southeastern districts) with these frequencies in the other three groups. Here, the most frequent surnames were Popopv, Shevchenko, Bondarenko, Kovalev, Kolesnikov, Kravtsov, Kuznetsov, Ivanov, Klimenko, and Shapovalov. Note that, like the curve of frequent surnames in group I (Fig. 5), the curve for group II plateaus almost immediately, whereas the curves for the other three groups are broken lines reflecting the specificity of each group. The surnames Bondarenko, Kovalev, Kravtsov, and Klimenko were more frequent in group II; and the surname Popov, in group IV. As in group I, some surnames in group II had close frequencies in all the four groups. Here, these were also three out of 50 most frequent surnames: Volkov, Andreev, and Chernykh.

Figure 7 shows the comparison of the frequencies of the 50 most frequent surnames in group III of districts (four western districts) with their frequencies in the other three groups. The 10 most frequent surnames of group III were Tkachenko, Shevchenko, Kravchenko, Mishenin, Litvinov, Goncharov, Shapovalov, Cherkashin, Popov, and Vasilenko. Like the curves of frequent surnames of the

Distributions of the 50 most frequent surnames in four groups of districts (surname frequencies, $P_i \times 10^2$)

No.	Common surnames	Group IV	Group III	Group II	Group I	Mean frequency in Belgorod oblast	
						calculated from the data on four groups of districts	calculated from the data on 22 districts
1.	Popov	1.717	0.376	0.652	0.560	0.826	0.657
2.	Goncharov	1.105	0.402	0.195	0.294	0.499	0.360
3.	Shevchenko	0.099	0.543	0.584	0.202	0.357	0.359
4.	Kovalev	0.059	0.285	0.471	0.330	0.286	0.336
5.	Novikov	0.844	0.356	0.123	0.321	0.411	0.321
6.	Ivanov	0.180	0.238	0.349	0.360	0.282	0.319
7.	Kolesnikov	0.451	0.120	0.411	0.261	0.311	0.294
8.	Shapovalov	0.424	0.385	0.322	0.160	0.323	0.269
9.	Tkachenko	0.027	0.579	0.295	0.161	0.265	0.261
10.	Kuznetsov	0.403	0.155	0.356	0.217	0.283	0.261
11.	Bondarenko	0.004	0.255	0.495	0.172	0.232	0.260
12.	Bondarev	0.544	0.314	0.132	0.181	0.293	0.225
13.	Zakharov	0.461	0.119	0.077	0.298	0.239	0.220
14.	Litvinov	0.035	0.403	0.192	0.185	0.204	0.213
15.	Kravchenko	0.008	0.407	0.314	0.113	0.210	0.212
16.	Cherkashin	0.197	0.378	0.028	0.245	0.212	0.206
17.	Golovin	1.242	0.077	0.146	0.078	0.386	0.202
18.	Kapustin	1.087	0.044	0.163	0.099	0.348	0.196
19.	Tarasov	0.008	0.065	0.211	0.264	0.137	0.190
20.	Morozov	0.183	0.145	0.078	0.264	0.168	0.185
21.	Medvedev	0.004	0.101	0.091	0.309	0.126	0.184
22.	Volkov	0.233	0.186	0.196	0.162	0.194	0.182
23.	Safonov	0.004	0.251	0.179	0.181	0.154	0.177
24.	Sorokin	0.810	0.190	0.066	0.111	0.294	0.177
25.	Kravtsov	0.075	0.166	0.372	0.083	0.174	0.176
26.	Kozlov	0.004	0.131	0.145	0.244	0.131	0.175
27.	Chernykh	0.165	0.145	0.156	0.199	0.166	0.174
28.	Kovalenko	0.016	0.309	0.261	0.089	0.169	0.169
29.	Sergeev	0.008	0.064	0.089	0.286	0.112	0.167
30.	Semenov	0.811	0.038	0.134	0.104	0.272	0.165
31.	Nikulin	0.000	0.058	0.025	0.315	0.099	0.160
32.	Shevtsov	0.000	0.156	0.158	0.177	0.123	0.152
33.	Klimenko	0.037	0.115	0.336	0.074	0.141	0.150
34.	Stepanov	0.231	0.029	0.220	0.138	0.155	0.149
35.	Miroshnichenko	0.016	0.296	0.212	0.076	0.150	0.148
36.	Ushakov	0.738	0.031	0.063	0.123	0.239	0.146
37.	Mikhailov	0.027	0.122	0.210	0.136	0.124	0.144
38.	Evsyukov	0.098	0.335	0.195	0.043	0.168	0.143
39.	Kalashnikov	0.055	0.125	0.189	0.137	0.127	0.142
40.	Alekseev	0.079	0.038	0.059	0.242	0.105	0.140
41.	Yakovlev	0.105	0.201	0.040	0.180	0.132	0.139
42.	Petrov	0.123	0.057	0.130	0.178	0.122	0.138
43.	Zaitsev	0.016	0.068	0.133	0.191	0.102	0.137
44.	Tkachev	0.008	0.178	0.143	0.134	0.116	0.133
45.	Zhukov	0.130	0.131	0.132	0.134	0.132	0.133
46.	Lysenko	0.162	0.133	0.142	0.121	0.140	0.133
47.	Sokolov	0.000	0.132	0.199	0.119	0.113	0.132
48.	Maslov	0.183	0.050	0.051	0.200	0.121	0.131
49.	Dmitriev	0.031	0.058	0.074	0.214	0.094	0.131
50.	Belyaev	0.303	0.217	0.033	0.117	0.167	0.129
Population over 18 years of age		23407	78314	208423	539255	849399	849399
Number of surnames		1550	8902	16067	40064	50412	50412

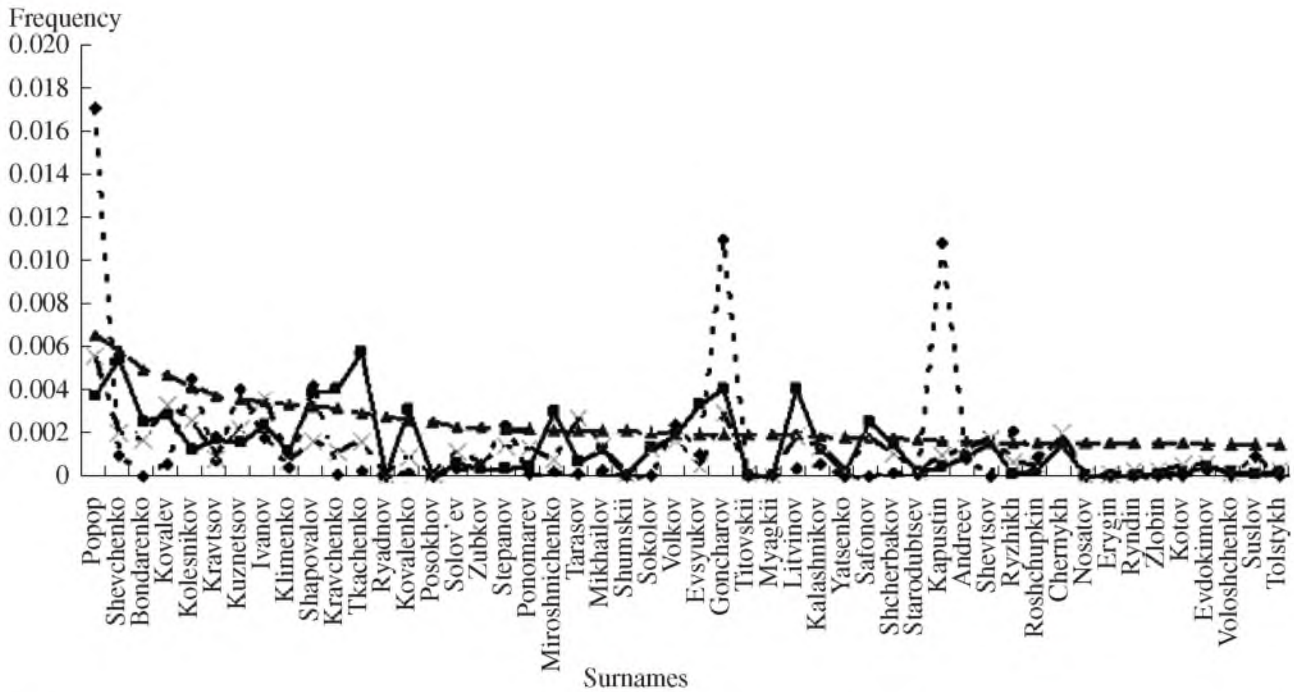

Fig. 6. Comparative analysis of the 50 most frequent surnames in group II of districts and the frequencies of these surnames in the other three groups. The 50 most frequent surnames in group II are listed in the order of decreasing frequency (triangles). The frequencies of the same surnames in the other three groups are shown: group I, crosses; group III, squares; group IV, diamonds.

two previous groups (Figs. 5, 6), the surname curve for group III plateaus immediately, whereas the curves for the other three groups are broken lines. The surnames Tkachenko, Mishenin, Litvinov, Cherkashin, and Vasilenko were more frequent in group III; and the surnames Goncharov and Popov, in group IV. Only one of the 50 surnames that were most frequent in group III (Volkov) had about the same frequency in all groups.

Note the specificity of the surname spectrum in group III. More than 50% of the 50 surnames most frequent in this group were Ukrainian. Only surnames ending in *-enko* (the most common Ukrainian surnames, often derived from Christian names [10]) were as many as 23 (47%), whereas there were as few as nine, three, and one such surnames among the 50 most frequent surnames in groups I, II, and IV. The high prevalence of Ukrainian surnames in districts of group III could be related to their population history. Not only Russians, but also Ukrainians actively populated the area of the modern Belgorod oblast in the 17th and 18th centuries. Ukrainian farmers (so-called cherkasses) fled from under the oppression of Polish landlords to the east in search for freedom. In the 1670s, Ukrainian populated lands in the basin of the Vorskla River near the mouth of the Graivoronka River, where they founded the Graivorony settlement (later, the town of Graivoron) [14]. Most of these migrants came from Right-Bank Ukraine. In the 18th century, the population of the Borisovskaya suzerainty (corresponding to the modern Borisovka and Graivoron raions) was mainly formed by migrants from Ukraine. In the first decades of the 18th century, count B.P. Sheremet'ev and, later, his wife and son bought much of the Borisovskaya suzerainty to populate these lands by Ukrainian immigrants (cherkasses). Count Sheremet'ev instructed his land steward to let only Ukrainians to settle in the suzerainty: "I suggest that you go on letting in my newly founded settlements cherkasses the free people that shall come from under the reign, except for Cossacks, […] and you should not let in Cossacks, and much less runaway soldiers and other servicemen and runaway Russian [serfs], and if you let them in, you will yourself bear responsibility for them" (cited from [15]). This was because, in that period, Ukrainian farmers still had the right to move to another land, whereas Russian farmers (serfs) were already deprived of this right [15].

Combined analysis of Figs. 5–7 shows the presence of not only surnames common for three groups (Volkov), but also the "groups' own" specific surnames, relatively rare in the other two groups. Each group had relatively many "its own" surnames. In group I, these were Nikulin, Medvedev, Alekseev, Chernov, Dmitriev, Besedin, Shekhovtsov, Proskurin, Bocharov, Potapov, Chuev, Bulgakov, Kulabukhov, Lazarev, Shlyakhov, and Ozerov (16 out 50 analyzed surnames or 32%) (Fig. 5). The surnames relatively specific for group II were Bondarenko, Kravtsov, Klimenko, Ryadnov, Posokhov, Zubkov, Shumskii, Titovskii, Myagkii, Yatsenko, Starodubtsev, Nosatov, Erygin, Ryndin, Zlobin, Kotov, Evdokimov, Voloshchenko, and Tolstykh

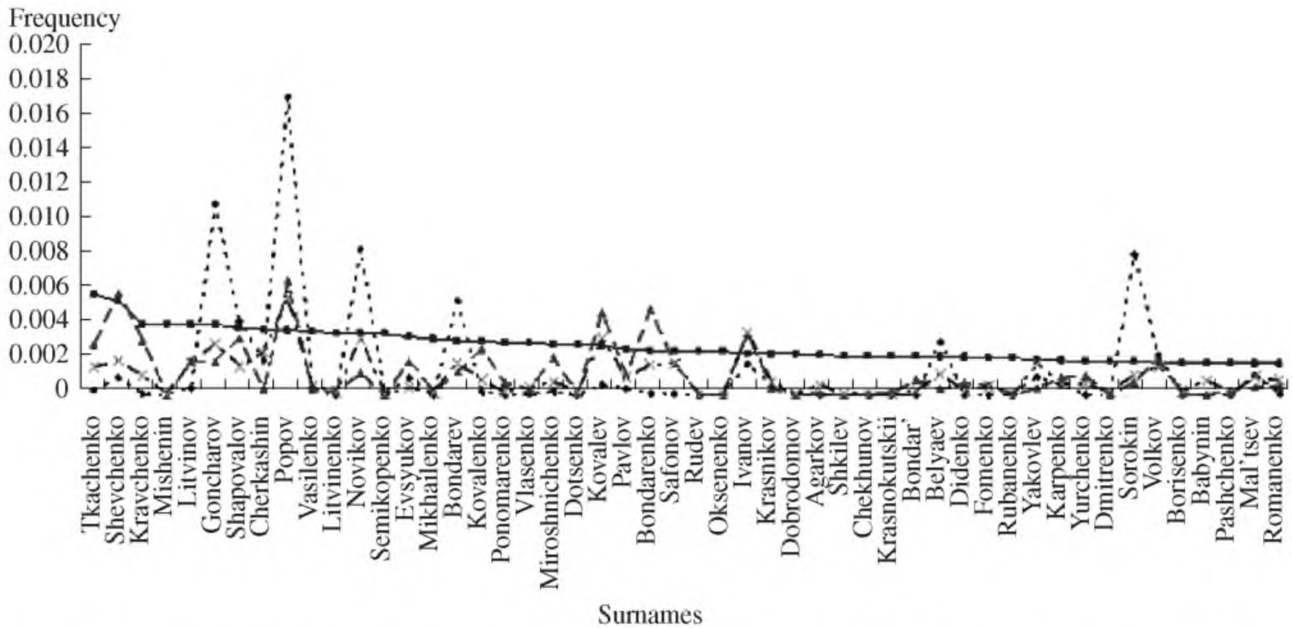

Fig. 7. Comparative analysis of the 50 most frequent surnames in group III of districts and the frequencies of these surnames in the other three groups. The 50 most frequent surnames in group III are listed in the order of decreasing frequency (squares). The frequencies of the same surnames in the other three groups are shown: group I, crosses; group II, triangles; group IV, diamonds.

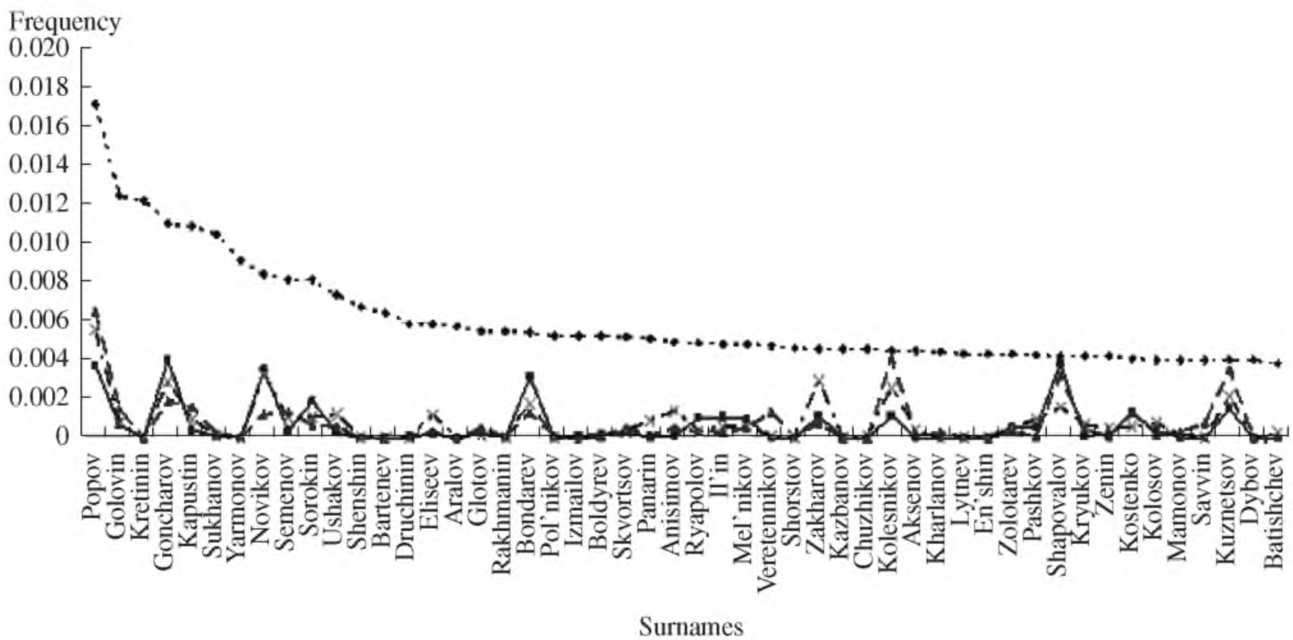

Fig. 8. Comparative analysis of the 50 most frequent surnames in group IV of districts and the frequencies of these surnames in the other three groups. The 50 most frequent surnames in group IV are listed in the order of decreasing frequency (diamonds). The frequencies of the same surnames in the other three groups are shown: group I, crosses; group II, triangles; group III, squares.

(19 out 50 analyzed surnames or 38%) (Fig. 6). Group III had even more "its own" surnames (26 out of 50 or 52%) (Fig. 7), mainly Ukrainian ones: Tkachenko, Mishenin, Litvinov, Vasilenko, Litvinenko, Semikopenko, Mikhailenko, Ponomarenko, Vlasenko, Dotsenko, Pavlov, Rudev, Oksenenko, Krasnikov, Dobrodomov, Agarkov, Shkilev, Chekhunov, Krasnokutskii, Bondar', Didenko, Fomenko, Rubanenko, Dmitrenko, Borisenko, and Pashchenko. Note that all frequent surnames in three groups of districts had approximately equal frequencies.

Despite the large number of specific surnames in each of the three groups (32–52%), group IV was the most different from others with respect to the frequencies and spectrum of 50 frequent surnames. Figure 8 shows 50 surnames that were the most frequent in group IV and the frequencies of the same surnames in the other three groups. In contrast to the surname frequency curves for groups I–III (Figs. 5–7), first, there are no distinct peaks of surname frequencies in the other three groups; second, the surname frequencies in group IV were substantially higher than the surname frequencies in the other three groups of districts.

The ten most frequent surnames in group IV (two eastern districts) were Popov, Golovin, Kretinin, Goncharov, Kapustin, Sukhanov, Yarmonov, Novikov, Semenov, and Sorokin. Many of the surnames that were common in Belgorod oblast in general, such as Shevchenko, Kovalev, Tkachenko, Bondarenko, Litvinov, and Kravchenko, were not in the list of the 50 most common surnames in group IV. Most of the 50 most frequent surnames were "local" surnames, which were entirely absent in other groups: Kretinin, Sukhanov, Yarmonov, Shenshin, Bartenev, Druchinin, Aralov, Rakhmanin, Pol'nikov, Shorstov, Kazbanov, Chuzhikov, Lytnev, En'shin, and Dybov. In general, 45 out 50 analyzed surnames (90%) were specific for group IV; i.e., they were considerably more frequent in the district populations of this group or were found only there. None of the 50 most frequent surnames of group IV was found with about the same frequency in all the four groups.

Note that this plot (Fig. 8) not only clearly demonstrates the specificity of the spectrum and frequencies of surnames of group IV, but also complements the pattern of the surname distribution in the first three groups. Figure 8 distinctly shows a significant correlation between the frequencies of surnames in groups I and II ($\rho = 0.59$, $p < 0.001$): the frequencies of most surnames in these groups increased and decreased in concert. The relationship between the frequencies of surnames in the first two groups and in group III was weaker: Spearman's coefficients (ρ) of correlation both between the surname frequencies in groups I and III and between those in groups II and III were 0.49 ($p < 0.001$). The correlations between the surname frequencies in the first two groups and group IV were even weaker; the respective correlation coefficients were 0.37 and 0.30 ($p < 0.001$). The correlation between the surname frequencies in groups III and IV was the weakest ($\rho = 0.20$, $p < 0.001$).

In summary, the main general tendencies were the distinct difference of group IV of district populations from the other groups in both the frequencies and spectrum of surnames and a peculiar surname spectrum (with a high proportion of Ukrainian surnames) in group III. Groups I and II were similar to each other in the surname frequency distribution. In an earlier study on "family name portraits" in groups of districts of Belgorod oblast with different degrees of subdivision [10], we found pronounced differences in the frequencies and spectrum of surnames in districts with a high subdivision level ($f_r^* > 0.00043$): Krasnaya Yaruga raion (the western cluster), Ivnya raion (the central cluster), Krasnoe raion (the eastern cluster), and Roven'ki raion (the southeastern cluster). Thus, the specificity of the frequencies and spectrum of surnames in the district populations of Belgorod oblast play a role in both the considerable subdivision of individual populations and the formation of an ordered system of genetic differentiation of districts in a large regional population.

To estimate the contribution of the surnames of each group to the formation of the total surname pool, we calculated the correlation coefficients between the frequencies of all surnames in each group and the "total regional" frequencies of all 50000 surnames. We used two estimations of the regional frequencies calculated, respectively, from the data on four groups of districts and on 22 districts (table). We found that the surname frequencies in all the four groups of districts were closely correlated with the frequencies in the total list of surnames calculated for 22 districts. The correlation coefficient expectedly decreased with the number of districts forming a group. The surname frequency distributions for the districts of the first and second groups (ten and six districts, respectively) were the most similar to that in the distribution of the mean surname frequencies calculated for 22 districts. The correlation coefficients between the surname frequencies in groups I and II and the mean regional frequencies were $r = 0.89$ and $r = 0.81$ ($p < 0.001$), respectively. This correlation coefficient for group III (four districts) was 0.70 ($p < 0.001$); for group IV (two districts), it was 0.58 ($p < 0.001$). The relationships between the surname frequencies in each of the four groups and their distribution averaged over the four groups of districts were somewhat different. The correlation coefficients of the surname frequencies in groups I and II with the mean regional frequencies were $r = 0.74$ and $r = 0.72$ ($p < 0.001$), respectively. The surname frequencies in group III exhibited the weakest correlation with the mean regional frequencies calculated from the data on four groups of districts ($r = 0.66$, $p < 0.001$), which may have been related to the high prevalence of Ukrainian surnames, which were considerably rare in the region in general. The highest correlation coefficient between the surname frequencies in group IV of districts and the mean regional frequencies ($r = 0.80$, $p < 0.001$) was an unexpected finding. This may be explained by the following. In group IV, the frequencies of most surnames were either the highest or the lowest among the four groups of districts, which ultimately resulted in a considerable shift of the mean regional frequencies of these surnames towards higher or lower values, respectively. This concerted variation of surname frequencies in group IV and in Belgorod oblast in general was reflected by the correlation coefficient, which became higher in this case. For example, the surname Medvedev (which was the 21st in the total regional list, with a frequency of 0.126×10^{-2}) was found in group IV with a frequency of 0.004×10^{-2}, which was 77, 25, and

22 times rarer than in groups I (0.309×10^{-2}), III (0.101×10^{-2}), and II (0.091×10^{-2}), respectively. Conversely, the surname Golovin (the 17th surname in the total regional list, with a frequency of 0.386×10^{-2}) was considerably more frequent in group IV (with a frequency of 1.242×10^{-2}) than in other groups, its frequency differing by a factor of 16 from those in groups I and III (0.078×10^{-2} and 0.087×10^{-2}, respectively) and by a factor of 8.5 from that in group II (0.146×10^{-2}).

Thus, districts of group IV (the eastern cluster) were characterized by very specific spectrum and frequencies of surnames. A considerable proportion of Ukrainians surnames were typical of the "family name portrait" of group III of districts (the western cluster). District populations of groups I (the central cluster) and II (the southeastern cluster)were similar to both each other and the total regional population with respect to the spectrum and frequencies of surnames.

To summarize the results of analysis of the "genetic landscape" of 22 district populations of the Central Chernozem region on the basis of anthroponymic data, we should note the following. First, district populations of Belgorod oblast form a definite, ordered system of four clusters corresponding to the actual geographic positions of the populations: the central, southeastern, western, and eastern clusters (comprising ten, six, four, and two districts, respectively). Second, the "family name portraits" of districts of the central and southeastern clusters are similar both to each other and to that of the whole region. The "family name portrait" of the western cluster is characterized by a high proportion of Ukrainian surnames. The eastern cluster has very specific surname spectrum and frequencies. Third, the frequencies and spectrum of surnames in districts of the Central Chernozem region and the geographic distances between them play and important role in the genetic differentiation of the large population of Belgorod oblast.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for the Humanities and the Russian Foundation for Basic Research.

REFERENCES

1. Deryabin, V.E., *Mnogomernye biometricheskie metody dlya antropologov* (Multivariate Biometrical Methods for Anthropologists), Available from VINITI, 2001, Moscow.
2. Pasekov, V.P., Genetic Distances, *Itogi Nauki Tekh., Ser. Obsh. Genet. Theor. Popul. Genet.*, 1983, vol. 8, pp. 3–75.
3. Paradeeva, G.M., Bol'shakova, L.P., Petrin, A.N., et al., Medical Genetic Study of the Population of Kostroma Oblast: III. Coefficient of Inbreeding and Its Dynamics in Rural Population and in Small Towns, *Genetica* (Moscow), 1986, vol. 22, no. 9, pp. 2355–2360.
4. El'chinova, G.I., Kadoshnikova, M.Yu., and Mamedova, R.A., Identification of Specific Population Genetic Structure by the Descriptive Method of "Genetic Landscape", *Russ. J. Genet.*, 1991, vol. 27, no. 11, pp. 327–334.
5. *Nasledstvennye bolezni v populyatsiyakh cheloveka* (Hereditary Diseases in Human Populations), Ginter, E.K., Ed., Moscow: Meditsina, 2002.
6. Khusnutdinova, E.K., *Molekulyarnaya etnogenetika narodov Volgo-Ural'skogo regiona* (Molecular Ethnogenetics of Populations of the Volga–Ural Region), Ufa, 1999.
7. Mamedova, R.A., The Effect of Genetic Drift on Territorial Distribution of the Genetic Pathology Load in Kirov Oblast, *Extended Abstract of Cand. Sci. (Med.) Dissertation*, Moscow, 1993, p. 23.
8. El'chinova, G.I., Application of the Methods of Population Genetic Analysis during the Study of Russia Populations with Different Genetic Demographic Structure, *Extended Abstract of Doctoral (Biol.) Dissertation*, Moscow: GU MGNTs Russ. Acad. Med. Sci., 2001, p. 48.
9. Churnosov, M.I., Sorokina, I.N., and Balanovska, E.V., The Gene Pool of the Belgorod Oblast Population: Dynamics of Endogamy Index in Region Populations, *Russ. J. Genet.*, 2008, vol. 44, no. 8, pp. 1117–1125.
10. Sorokina, I.N., Churnosov, M.I., and Balanovska, E.V., The Gene Pool of the Belgorod Oblast Population: II. "Family Name Portraits" in Groups of Districts with Different Degrees of Subdivision and the Role of Migration in Their Formation, *Russ. J. Genet.*, 2007, vol. 43, no. 8, pp. 1120–1128.
11. Sorokina, I.N., Churnosov, M.I., and Balanovska, E.V., The Gene Pool of the Belgorod Oblast Population: Dynamics of the Population Genetic Ratios during the Last 50 Years, *Russ. J. Genet.*, 2009, vol. 45, no. 4, pp. 555–563.
12. Nei, M., Analysis of Gene Diversity in Subdivided Populations, *Proc. Natl. Acad. Sci. USA*, 1973, vol. 70, pp. 3321–3323.
13. Nei, M., F-Statistics and Analysis of Gene Diversity in Subdivided Populations, *Ann. Hum. Genet.*, 1977, vol. 41, pp. 225–233.
14. Where does the Sward on the Coat of Arms of Prokhorovka Come From? *Komsomol'skaya Pravda (Belgorod)*, 28 November, 2003.
15. Borisovka. *Istoricheskie ocherki: Belgorod* (Historical Essays: Belgorod), Izd. dom V. Shapovalov, 2000.