

Population-Genetic Structure of the Residents of Central Russia

Evgenij N. Krikun¹, Kristina N. Sergeeva¹, Inna N. Sorokina², Irina V. Batlutskaya¹, Irina N. Polyakova¹, Anna V. Elikova¹

¹ Belgorod State University, 308015, Belgorod, Pobedy Street, 85, Russia

² Belgorod State University, 308015, Belgorod, Pobedy Street, 85, Russia; e-mail: sorokina@bsu.edu.ru

Abstract

In the territory of Central Russia, the presence of a certain, ordered system of grouping of elementary populations (raions) has been established, according to the data on the frequencies of surnames, using various methods of multivariate statistics (cluster analysis, multidimensional scaling, factor analysis). All the raions of the Belgorod oblast are combined into four clusters, corresponding to real geographical location of the populations: central cluster (12 raions), western (4 raions), south-eastern (3 raions), eastern (2 raions). Geographical distances between them play a significant role in the formation of genetic division between elementary populations.

1. INTRODUCTION

For a comprehensive analysis of the genetic structure of populations and the assessment of genetic relationships between them, the methods of multivariate statistics are widely used: cluster analysis, factor analysis, multidimensional scaling (2). Based on the surname frequencies, genetic distances between populations can be calculated. In population genetics, genetic distances are widely used for studying the relationships between elementary populations, according to the degree of their genetic similarity, and for establishing the nature of genetic relatedness between populations. That is, genetic distances allow to judge the degree of similarity and differences between populations, according to the frequency of genes (or their analogues - surnames) in terms of “genetic proximity” or “genetic distance” of populations. Based on the genetic distance matrices, cluster analysis is performed, and a dendrogram is developed. In order to obtain more illustrative picture of interrelationships between the populations, the appropriate dendrograms are used to make the “genetic landscape” of the area (3; 5). This method of description of “genetic landscape” can be effectively used to define the boundaries and size of the population, as a natural-historical unit, for the purpose of further detailed medical-genetic and genetic-epidemiological research of the population (4; 6).

This paper presents the results of using the surnames from telephone directories, in order to

estimate genetic distances and to carry out the taxonomic analysis of elemental populations of the Central Black Earth Region of the Russian Federation (10).

2. Materials and Methods

The estimation of genetic distances and taxonomic analysis were carried out at the level of elementary population (raion) (1) among the total population of the Belgorod oblast (20 raions), Pristenskii raion of the Kursk oblast, and Repyevsky raion of the Voronezh oblast. The material for research was the data on the distribution of surnames, obtained from the telephone directory, among more than 1 million people. Detailed analysis of the distribution of surnames, obtained from the telephone directory, in the above populations was presented by us earlier (7; 8).

Correspondence Author: Inna N. Sorokina

Email address: sorokina@bsu.edu.ru

Received: 04-07-2018

Revised: 21-9-2018

Accepted: 15-11-2018

Based on the frequency of surnames, the matrices of genetic distances were calculated, using the software DJ genetic (version 0.03 beta), developed by Yu.A. Seregin and E.V. Balanovskaya at the Research

Centre of Medical Genetics of the Russian Academy of Medical Sciences. The distances between elementary populations were calculated using the method M. Nei.

Based on the obtained matrix of genetic distances and the correlation matrix, the cluster analysis, multidimensional scaling and factor analysis were performed using the software Statistics (version 6). The cluster analysis was carried out by two methods: weighted average relationship and Ward's method. In this paper, we present a dendrogram, corresponding to the results, obtained by other methods of multivariate analysis — multidimensional scaling and factor analysis.

In accordance with the results of cluster analysis (dendrogram), the scheme of genetic landscape of the studied populations was developed. The equidistant lines were drawn through a certain unit of genetic distance (0.5), so that the number of levels in a universe does not exceed 10, since greater number of levels makes the scheme cumbersome and difficult for analysis.

In order to study the role of geographic distances in the formation of genetic structure of regional populations of Central Russia, we carried out a correlation analysis (the Spearman's rank correlation coefficient was calculated) between three types of matrices: the matrix of genetic distances, calculated on the basis of data, obtained from the telephone directory; the matrix of genetic distances, calculated on the basis of widespread surnames in the register of voters; and the matrix of geographic distances. Using the map of the Belgorod oblast (scale 1: 400000), geographical distances between the studied elementary populations were calculated.

3. Results

In accordance with the data of cluster analysis, the scheme of genetic landscape of all analysed populations was developed. Equidistant lines were drawn through 0.2 units of genetic distance. The number of levels in the universe should not exceed 10, because greater number of levels makes the scheme cumbersome and difficult for analysis. The spatial location of the results of cluster analysis showed the presence of a certain ordered system of grouping of raion populations of Central Russia.

According to the frequencies of surnames, obtained from the lists of telephone subscribers, the matrix of genetic distances was calculated. On its basis, using a cluster analysis, a dendrogram was developed, characterizing the genetic relationships between 22 studied elementary populations (raions). A similar procedure was conducted by us earlier, on the basis of surnames, obtained from the register of voters (9).

The analysis of dendrogram allows to identify four groups of clusters. The first group of clusters includes two subgroups, consolidating 9 raions. These are Gubkinsky, Starooskolsky, Novooskolsky, Chernyansky, Yakovlevsky, Ivnyansky, Prokhorovsky, Korochansky raions of the Belgorod oblast, and also Pristenskii raion of the Kursk oblast. They occupy the central part of the area.

The second group is constituted by Krasnoyarskiy, Rakityansky, Grayvoronsky and Borisovsky raions, located in the west of the oblast, and having common borders between themselves. This fully corresponds to the results of clustering by surnames, based on the data from the register of voters. The third group of clusters is formed by Alekseevsky, Veydelevsky, Rovensky, Valuysky, Volokonovsky, Krasnogvardeysky and Shebekinsky raions, which are located in the south and southeast of the oblast and are also characterized by a common territorial situation. Krasnensky raion of the Belgorod oblast is genetically closer to Repyevsky raion of the Voronezh oblast (they have a common border), both on the basis of clustering, according to the list of telephone subscribers, and according to the register of voters. This is evidenced by the independent fourth cluster, which they grouped themselves.

In accordance with the data of cluster analysis, the scheme of genetic landscape of all analysed populations was developed. Equidistant lines were drawn through 0.5 units of genetic distance. The spatial location of the results of cluster analysis showed the presence on the territory of Belgorod oblast of a certain system of elementary populations grouping.

Equidistant lines are constructed without intersection. Regional populations, located in the north and in the centre of Belgorod oblast, are consistently combined into a single group. In parallel with this, there is a consistent integration of other subsystems, located in the west, south-east and east of the region. Then, at the level 2.5, the central cluster joins with eastern, and the western cluster joins with south-eastern. At the final stage of the formation of genetic landscape, these groups are combined.

The correlation analysis of matrices of genetic distances, obtained from the list of telephone subscribers, and from the register of voters on the one hand, and geographical distances between raion populations, on the other, showed the presence of a significant positive correlation of average force ($r = 0.35$ and 0.45 , respectively, $p < 0.001$). Spearman's rank correlation coefficient between the matrices of genetic distances, calculated on the basis of surnames frequencies, according to the list of telephone subscribers, and to the register of voters, was 0.81 ($p < 0.0001$).

Using the method of multidimensional scaling, which is a kind of alternative to cluster analysis, the

graph of spatial relative position of the studied rural raion populations was obtained, on the basis of matrix of genetic distances, according to the data from telephone directory. The results of three-dimensional scaling were the most acceptable (after 48 iterations, the values of stress and alienation was $S_o = 0.10$ and $K_o = 0.12$, respectively, the Shepard curve is satisfactory).

It should be noted, that the grouping of areas in three-dimensional space is comported with the results of cluster analysis and equidistant schemes, where the areas of the Belgorod oblast are combined into 4 groups of clusters. Thus, according to the results of multidimensional scaling, the areas, included in the 1st cluster group, have a common location in three-dimensional space (Yakovlevsky, Shebekinsky, Prokhorovsky, Ivnyansky, Chernyansky, Korochansky, Novooskolsky, Volokonovsky, Valuysky, Krasnogvardeysky, Staroskolsky, Gubkinsky). Similarly, the areas, forming the 2nd and 3rd groups of clusters, are located together in space. These are Alekseevsky, Rovensky and Veydelevsky raions (the 2nd group of clusters) and Grayvoronsky,

Rakityansky, Borisovsky, Krasnoyaruhsky raions (the 3rd group of clusters). Two raions from the fourth group of clusters, Krasnensky and Repevsky, are also in close proximity to each other. Pristenskii raion is located somewhat apart from all other analysed areas.

When carrying out the factor analysis, using the method of principal factors, according to the scree-test, four significant main factors were identified (Table 1). The first main factor has negative loads on all raion populations, which are on average $-0.3 - -0.7$. It is most correlated with Starooskolsky (-0.735), Gubkinsky (-0.638), Yakovlevsky (-0.637), Valuysky (-0.630), Shebekinsky (-0.603) raions, which are included in the first cluster. The first main factor has the smallest correlations with Krasnensky (-0.331), Pristenskii (-0.332), Repevsky (-0.395) raions.

The second main factor is characterized by a bipolar nature. Moreover, it is most positively correlated with the areas, located in the central and north-eastern parts of the oblast - Pristenskii (0.393), Prokhorovsky (0.378), Ivnyansky (0.330), Gubkinsky (0.327) and Starooskolsky (0.271) raions, having common territorial borders.

Table (1) Results of factor analysis of raion populations of the Belgorod oblast, according to the lists of telephone subscribers

Raions	Main factors			
	1	2	3	4
Alekseevsky	-0.596	-0.249	-0.254	0.012
Borisovsky	-0.493	-0.327	0.108	0.029
Valuysky	-0.630	-0.090	-0.155	0.228
Weidel's	-0.452	-0.312	-0.115	0.340
Volokonovsky	-0.522	-0.089	-0.160	0.124
Graivoronsky	-0.415	-0.294	0.249	-0.095
Gubkinsky	-0.638	0.327	-0.151	-0.077
Ivnyanskiy	-0.456	0.329	0.192	0.0003
Korochansky	-0.503	0.157	0.244	0.161
Krasnenskiy	-0.331	0.204	-0.481	-0.316
Krasnogvardeysky	-0.523	-0.122	-0.125	0.020
Krasnoyaruhsky	-0.419	-0.348	0.302	-0.451
Novooskolsky	-0.574	-0.049	-0.079	0.014
Prokhorovsky	-0.387	0.378	0.389	0.177
Rakityansky	-0.462	-0.242	0.357	-0.359
Rovensky	-0.409	-0.331	0.008	0.269
Art. Oskolsky	-0.735	0.271	-0.178	-0.112
Chernyansky	-0.528	0.045	-0.175	0.007
Shebekinsky	-0.603	-0.066	0.024	0.042
Yakovlevsky	-0.637	0.222	0.192	0.041
Pristenskii (Kursk oblast)	-0.332	0.3926	0.221	0.1735
Repevsky (Voronezh oblast)	-0.395	0.1363	-0.124	-0.427
Dispersion	5.78	1.40	1.11	1.03
Contribution to total dispersion (%)	26.29	6.37	5.04	4.70

Krasnoyarskiy (0.348), Borisovskiy (-0.327), Grayvoronskiy (-0.294), Rakityanskiy (-0.242) raions, located in the west of the region, have maximum negative loads. The distinctive feature of the third main factor, which also has a bipolar nature, like the second one, is its maximum negative correlation with Krasnenskiy raion (-0.481).

It is important to emphasize that the grouping of raions in the space of the first three main factors is almost completely consistent with the results, obtained during the cluster analysis and multidimensional scaling, and also comports with the previously data, obtained from the register of voters (1).

4. Conclusion

Using various methods of multidimensional statistics (cluster analysis, multidimensional scaling, and factor analysis), based on the analysis of matrices of genetic distances and correlations, calculated according to the frequencies of surnames, received from the telephone directory and the register of voters, almost the same results were obtained. They reflect the features of organization of population-genetic structure of elementary populations (raions) of Central Russia.

REFERENCES

- [1] Churnosov, M.I., Sorokina, I.N., Balanovskaya, E.V., (2008). The gene pool of the Belgorod oblast population: changes in the endogamy indices of district populations with time. *Russian Journal of Genetics*. 44(8): 975-982.
- [2] Deryabin, V.E., (2001). *Multivariate biometric methods for anthropologists*. M.: VINITI, 105-265.
- [3] El'chinova, G.I., (2001). *The Use of Population Genetic Analysis for Studying Population of Russia with Different Genetic Demographic Structures*. Extended Abstract of Doctoral (Biol.) Dissertation. Moscow: Medical Genetic Research Center of the Russian Academy of Medical Sciences, p. 48.
- [4] Moskalenko, M.I., (2018). The involvement of genes of matrix metalloproteinase in the development of arterial hypertension and its complication (review). *Research Result. Medicine and Pharmacy*. 4(1):53-69.
- [5] *Hereditary Diseases in Human Populations*, (2002). Ginter, E.K., Ed., Moscow: Meditsina.
- [6] Rudyh, N.A., Sirotina, S.S., (2015). Genetic interrelations of Russian and Ukrainian populations of Belgorod region. *Research Result. Medicine and Pharmacy*, 1(3):72-79.
- [7] Sorokina, I.N., Balanovskaya, E.V., Churnosov, M.I., (2007a). The gene pool of the Belgorod oblast population: I. Differentiation of all district populations based on anthroponymic data. *Russian J. Genetics*. 43(6): 841-849.
- [8] Sorokina, I.N., Balanovskaya, E.V., Churnosov, M.I., (2007b). 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 migrations in their formation. *Russian J. Genetics*. 43(8): 1120-1128.
- [9] Sorokina, I.N., Churnosov, M.I., Balanovskaya, E.V., (2009). The gene pool of the Belgorod oblast population: description of the "genetic landscape" of 22 district populations. *Russian Journal of Genetics*. 45(5): 613-622.
- [10] Amelina, S.S., Vetrova, N.V., Amelina, M.A., Degtereva, E.V., Ponomareva, T.I., Elchinova, G.I., Zinchenko, R.A., Michailova, L.K., (2014). The load and diversity of hereditary diseases in four raions of Rostov oblast. *Russian Journal of Genetics*. 50 (1): 82-90.