

DISTRIBUTION OF POLYMORPHIC GENE VARIANTS ASSOCIATED WITH THE SEVERE COURSE OF COVID-19 IN THE SUBPOPULATION OF BASHKIRS

N.V. Ekomasova^{1,2*}, M.A. Dzhaubermezov^{1,2}, L.R. Gabidullina¹, Z.R. Sufianova¹, Y.O. Galimova¹, A.V. Kazantseva², A.Kh. Nurgalieva¹, D.S. Prokofieva¹, E.K. Khusnutdinova^{1,2}

¹ Federal State Educational Institution of Higher Education «Ufa University of Science and Technology», 32 Zaki Validi St., Ufa, 450076, Russia;

² Institute of Biochemistry and Genetics, Ufa Federal Research Center of the Russian Academy of Sciences, 71 Prospekt Oktyabrya, Ufa, 450054, Russia.

* Corresponding author: trofimova_nata_@mail.ru

Abstract. We have analyzed the distribution of alleles and genotypes of three polymorphic variants: rs11385942 of the *LZTFL1* gene, rs657152 of the *ABO* gene, and rs2109069 of the *DPP9* gene previously associated with COVID-19 in a sample of 80 presumably healthy individuals of the Permsky Bashkirs from the Permsky region of Russia, 48 representatives of the Burzyan region and 47 representatives of the Arkhangelsky district of the Republic of Bashkortostan of Russia. No statistically significant differences were shown between the Burzyan Bashkirs and the populations of Southern Siberia in the rs11385942 *LZTFL1* locus, in contrast to the subpopulations of the Permsky and Arkhangelsky Bashkirs. We also revealed statistically significant differences in the frequency of the minor allele, the rs11385942 *LZTFL1* and rs657152 loci of the *ABO* gene between Burzyan Bashkirs and a mixed sample of Komi and Udmurts. In turn, the subpopulations of the Permsky and Arkhangelsky Bashkirs in the rs657152 locus of the *ABO* gene were statistically significantly different from the Tatars.

Keywords: COVID-19, *LZTFL1*, *DPP9*, *ABO*, Permsky Bashkirs, Arkhangelsky Bashkirs, Burzyan Bashkirs.

List of Abbreviations

VUR – Volga-Ural region

GWAS – Genome wide association sequencing

DNA – Deoxyribonucleic acid

mtDNA – Mitochondrial deoxyribonucleic acid

HWE – Hardy–Weinberg equilibrium

EDTA – Ethylenediaminetetraacetic acid

Introduction

The study of how gene variants are distributed in the population, particularly those that affect the course and the development of infectious diseases, COVID-19 included, is an extremely relevant research field worldwide, predicting which regions host populations that may experience complications after a certain illness and a generally severe course of illness more frequently can ensure timely distribution of equipment for patients, the supply of correct amount of medicines and implementation of adequate preventive measures to minimize catastrophic consequences. In recent years, a large

number of scientific publications have focused on certain aspects of the development of COVID-19. Many of those works have been devoted to the study of polymorphic variants of genes associated with the severe course of the disease. One of the first major studies was the work of Ellenhaus *et al.*, where based on a sample of 1980 patients from Italy and Spain a genome-wide association study (GWAS) was conducted and an association of severe course of COVID-19, defined as respiratory failure, with a multigene cluster on chromosome 3 (3p21.31) was revealed (Ellenhaus *et al.*, 2020). The 3p21.31 locus is ~50 kb long and includes 6 genes (*SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCR1*); it was presumably obtained by humans from Neanderthals (Zeberg & Pääbo, 2020). Meta-analysis data allows us to isolate the *LZTFL1* gene (rs11385942), actively expressed in human lung cells (Ellenhaus *et al.*, 2020). For the first time, the frequency of the COVID-risk allele (GA) rs11385942 was shown to be ~1.5 times higher among ventilated hospitalized patients than among those who re-

ceived only supplemental oxygen (Ellenhaus *et al.*, 2020). An association of rs657152 of the 9q34.2 locus of the *ABO* gene with COVID-19 was also confirmed, and it was shown that the risky blood group is - blood type A, while 0 is protective (Ellenhaus *et al.*, 2020; Wu *et al.*, 2021; Zhao *et al.*, 2021). The significance of the *LZTFL1* gene in the pathogenesis of COVID-19 has been confirmed in a number of studies based on genome-wide analysis (GWAS) (Angulo-Aguado *et al.*, 2022; Cruz *et al.*, 2022; Downes *et al.*, 2021). When studying the population of China, no such association with the locus 3p21.31 and rs11385942 of the *LZTFL1* gene was found, which the authors attribute to the fact that the risk allele identified for European populations is not typical for populations of East Asia (Wu *et al.*, 2021).

Even though the role of the *ABO* gene has been questioned, and a number of studies demonstrating no association with the severe course of COVID-19, there is numerous evidence proving the significance of *ABO* gene (Dzik *et al.*, 2020; Latz *et al.*, 2020), in particular in the study of patients and control groups from Japan, the association of severe COVID-19 with AB blood group was revealed (Namkoong *et al.*, 2022).

Another variant involved in the pathogenesis of COVID-19 is rs2109069 of the dipeptidyl peptidase 9 (*DPP9*) gene (Pairo-Castineira *et al.*, 2021). This locus has previously been shown to be associated with idiopathic pulmonary fibrosis (Wu *et al.*, 2021). *DPP9* encodes serine protease that performs a variety of intracellular functions, including break down of the key antiviral signaling mediator CXCL10 (Zhang *et al.*, 2015), and plays a key role in the activation of inflammation processes (Griswold *et al.*, 2019).

Despite the fact that the study of the distribution of polymorphic variants of genes associated with the severe course of COVID-19 is being carried out worldwide, data on such distribution in populations and subpopulations of Russia are very limited. To date, a major article is published by Balanovsky *et al.* (Balanovsky *et al.*, 2021) that includes data on the distribution of alleles of the rs11385942 variants of the

LZTFL1 gene and rs657152 of the *ABO* gene in Russian populations. However, in the paper, many ethnic groups were mixed together as part of a general sample, for example, the Chuvash and Mari or the Udmurts and Komi, while the actual subethnic groups were not studied at all. The study considers a general sample of Bashkirs, however, territorial subpopulations that have significant differences in the genetic structure (Trofimova, 2015), were not considered. The Bashkirs are a Turkic-speaking ethnic group living mainly in the South Urals. According to the 2020 census, its population is approximately 1 million people 572 thousand people (All-Russian population census, 2020). The ethnonym «kort» has so far raised controversy in scientific circles. So, V.N. Tatishchev believes that it means «the main wolf» - «Bashkurt», in connection with the wolf that brought missionaries preaching Islam to the Ural Mountains to enlighten the ancestors of the Bashkirs who at the time practiced paganism (Tatishchev, 1962). Kuzeev supports the Turkic origin of the ethnonym and also associates the self-name of the Bashkirs with five wolves («bish» – «kurt» – wolf; or «boz» – gray, «kurt» – wolf), taking into account the special reverence of the Bashkirs, and in general the attitude of all the Turk peoples, to the wolf. So, in the southeastern regions of Bashkiria, a legend is widespread that the Bashkirs came to the Urals from the lower reaches of the Syr Darya, from Altai, following the wolf that showed them the way (Kuzeev, 1974). Most of the Bashkirs live within the administrative territories of the Republic – Bashkortostan. Our study involved 2 samples from the territory of Bashkortostan, namely the Bashkirs from the Burzyansky district and the Bashkirs from the Arkhangelsky district. Previously, significant genetic differences were identified in uniparental markers, mtDNA, it was found that in the subpopulation of the Burzyan Bashkirs, the East Eurasian component is 44.2%, and in the subpopulation of the Arkhangelsky Bashkirs, 25.1% (Trofimova, 2015). The location of this subpopulation is also important, if the Burzyan district belongs to the southeastern part of the republic, the Arkhangelsky is centrally located. Another

extremely interesting subpopulation of the Bashkirs is the Permsky Bashkirs, who also occupy an important place in their gene pool of East Eurasian mtDNA lines – 52.2% (Trofimova, 2015). The endonym of the Permsky Bashkirs of Russia corresponds to the ethnonym of one of the most ancient Bashkir tribes - «Gaynitsy». The language of the Permsky Bashkirs in terms of vocabulary and phonetic features belongs to the Gayn dialect of the northwestern Bashkir language (Mirzhanova, 2006). Since ancient times, agriculture has been the basis of the economy of this group of Bashkirs, as well as cattle breeding, hunting, fishing and gathering (Yusupov *et al.*, 2009). Anthropologically there are 2 types prevalent among the Bashkir-Gaynians: Pontic (dark-pigmented of southern origin), more typical for men; and Ural (with variants), more common for women (Yusupov, 1987; Yusupov, 1991; Yusupov, 2002; Yusupov, 2006; Yusupov *et al.*, 2009). According to historical legends, the ancestors of the Gaynians moved to the basin of the Tulva river in the pre-Mongolian period, from the «Minzelin side», and were «natives of the Bulgars» and their clan begins «from the generation of Tarkhans» (Nebolsin, 1852).

In our work, we present a study of the Permsky, Burzyan and Arkhangelsky Bashkirs subpopulations, which will significantly complement the available data on such a fascinating ethnic group, from the point of view of genetics and anthropology, as the Bashkirs.

Materials and Methods

The study included 80 presumably healthy individuals of the Permsky subpopulation of Bashkirs from the Permsky region of the Russian Federation, 48 representatives of the Burzyan region and 47 representatives of the Arkhangelsky district of the Republic of Bashkortostan. Sampling was carried out in accordance with the ethical standards of the Bioethics Committee, developed by the WMA Declaration of Helsinki – «Ethical Principles for the Conduct of Medical Research Involving Human Subjects». All subjects filled out a questionnaire taking into account their ethnicity up to three generations, year of birth. All respond-

ents signed an informed voluntary consent to participate in the study. The work was approved by the Local Ethics Committee of the Institute of Biochemistry and Genetics of the UFRC RAS (protocol No. 19 of November 25, 2021).

The DNA was extracted from peripheral blood samples using phenol-chloroform (Mathew, 1984). Vacutainer® tubes were used to collect, transport, and store the blood samples using 0.5 M EDTA solution as a preservative. After drawing the sample each tube was shaken and stored at 4 °C. Genotyping was carried out by determining single nucleotide polymorphisms using the KASP (Kompetitive Allele Specific PCR) method. The KASP genotyping method is based on competitive allele-specific PCR and makes it possible to determine both single nucleotide and insertion-deletion polymorphisms in both alleles. A mixture of SNP-specific primers and a 2-fold reaction mixture universal for genotyping (mastermix) were added to the DNA sample, then a polymerase chain reaction was performed, followed by endpoint fluorescence reading on a BioRad CFX96 Touch™ Real-Time PCR Detection Systems instrument. For RFLP genotyping, specific primers were selected for rs11385942 of the *LZTFL1* gene (F-5'-AAGCACAGTCACAGCACATCAGAT-3', R 5'-AGCACACCTTCTCAGAGTTTCT-3'). The incidence of allele variants in given population were calculated based on observed genotype frequencies. The correspondence of the genotype frequencies to the Hardy–Weinberg equilibrium was assessed using Pearson's χ^2 test (at $p > 0.05$). The significance of differences in allele frequencies in the sample was calculated by the χ^2 test using the Yates correction for continuity.

Results

In our work, we studied the distribution of alleles and genotypes of three polymorphic variants rs11385942 of the *LZTFL1* gene, rs657152 of the *ABO* gene, and rs2109069 of the *DPP9* gene previously associated with severe COVID-19. For all studied loci, the distribution of genotype frequencies corresponded to the Hardy–Weinberg equilibrium (Tables 1, 2, 3).

Table 1

Distribution of the rs11385942 genotypes of the *LZTFL1* Permsky, Burzyan and Arkhangelsky Bashkirs subpopulations of the Russian Federation

Population	N	G/G			G/GA			GA/GA			Minor allele frequency(95% CI)	χ^2	Deviations from HWE, p
		Observed (N)	Expected (N)	%	Observed (N)	Expected (N)	%	Observed (N)	Expected (N)	%			
Permsky Bashkirs	80	63	63.9	78.75	17	15.2	21.25	0	0.9	0	10.63 (6.31-16.47)	1.131	0.29
Burzyan Bashkirs	48	41	41.3	85.42	7	6.5	14.58	0	0.3	0	7.29 (2.98-14.45)	0.297	0.58
Arkhangelsky Bashkirs	47	37	36.6	78.72	9	9.7	19.15	1	0.6	2,13	11.70 (5.99-19.97)	0.253	0.61

Table 2

Distribution of the rs657152 genotypes of the *ABO* Permsky, Burzyan and Arkhangelsky Bashkirs subpopulations of the Russian Federation

Population	N	C/C			C/A			A/A			Minor allele frequency (95% CI)	χ^2	Deviations from HWE, p
		Observed (N)	Expected (N)	%	Observed (N)	Expected (N)	%	Observed (N)	Expected (N)	%			
Permsky Bashkirs	80	26	26.5	32.50	40	39.1	50	14	14.5	17.50	42.50 (34.73-50.55)	0.042	0.84
Burzyan Bashkirs	48	11	10.5	22.9	23	23.9	47.9	14	13.5	29.2	53.13 (42.66-63.39)	0.069	0.79
Arkhangelsky Bashkirs	47	17	15.5	36.17	20	23	42.55	10	8.5	21.28	42.55 (32.41-53.18)	0.790	0.37

Table 3

Distribution of the rs2109069 genotypes of the *DPP9* Permsky, Burzyan and Arkhangelsky Bashkirs subpopulations of the Russian Federation

Population	N	G/G			G/A			A/A			Minor allele frequency (95% CI)	χ^2	Deviations from HWE, p
		Observed (N)	Expected (N)	%	Observed (N)	Expected (N)	%	Observed (N)	Expected (N)	%			
Permsky Bashkirs	80	45	4.1	56.25	34	27.9	42.50	1	48.1	1.25	22.50 (16.28-29.76)	3.824	0.05
Burzyan Bashkirs	48	35	34.2	72.92	11	12.7	22.92	2	1.2	4.17	15.63 (9.02-24.46)	0.822	0.36
Arkhangelsky Bashkirs	47	29	28.3	61.70	15	16.3	31.91	3	2.3	6.38	22.34 (14.39-32.10)	0.303	0.58

We carried out the pairwise comparisons of the allele frequencies of all studied variants of the Permsky and Arkhangelsky Bashkirs subpopulations with data on the distribution of alleles in some other world populations, previously published in the academic literature, as well as from The1000 Genomes project (Supplementary tables 1, 2) (The 1000 Genomes Project Consortium, 2012).

Based on the distribution of alleles of the rs11385942 variant of the *LZTFL1* gene, which is associated with the severe course of COVID-19 and shows the strongest relationship based on a number of studies (Angulo-Aguado *et al.*, 2022; Cruz *et al.*, 2022; Downes *et al.*, 2021; Ellenhaus *et al.*, 2020; Wu *et al.*, 2021; Zhao *et al.*, 2021) we found statistically significant differences in the distribution of alleles with the populations of Altaians, Siberian Tatars, Mongols, Japanese, mixed samples of Africans; Native Americans; southern India; Tuvans and Tofalars; Buryats, Khaminigans and Yakuts; Chukchi, Koryaks and Itelmens; Nanais, Uliches, Nivkhs and Evens; Khanty, Mansi and Nenets; Uzbeks, Turkmens and Kirghiz. It should be noted that we recorded differences in the distribution of alleles only with the populations of East Asia, Africa, and America. Differences with the populations of the Volga-Ural region and the Caucasus were not identified, and no statistically significant differences were found with the general sample of the Bashkirs, previously published in the article by Balanovsky *et al.* (Balanovsky *et al.*, 2021). It should be noted that the picture changes when a similar analysis is carried out for the Burzyan Bashkirs, who traditionally live in the eastern range of the population. Unlike the Permsky and Arkhangelsky Bashkirs, there are no statistically significant differences with the populations of Altaians, Tuvans, Tofalars, Buryats, Yakuts, and, on the contrary, there is a statistical difference with the population of Ukrainians (Table 1, Supplementary table 1).

In the study of the rs657152 locus of the *ABO* gene, the differences in allele frequencies between the studied subpopulation of Permsky and Arkhangelsky Bashkirs and the populations of the world were found to be much less signif-

icant, we recorded statistically significant differences in the distribution of alleles of the Permsky and Arkhangelsky Bashkirs with the populations of Tatars, Mongols, a mixed sample of the Western Caucasus, populations of Anatolia and the Levant and Native Americans (Table 2, Supplementary table 2). In turn, the Burzyan Bashkirs showed a much more complex picture with statistically significant differences from the Finno-Ugric populations of Komi, Udmurts, Karelians and Veps, Khanty and Mansi, Samoyed Nenets, Turkic and Mongolian populations of Yakuts, Tuvans, Tofalars, Buryats, Khamingans, mixed Caucasian sample, as well as some European populations.

The study of the distribution of alleles of the rs657152 variant of the *DPP9* gene revealed statistically significant differences between the subpopulation of Permsky Bashkirs and the populations of South America (Peruvians), East Asia (Japanese, Han and Dai Chinese), and Europeans (Finns, British, and Spaniards) (Table 3, Supplementary table 3)

Discussion

Our data on the distribution of alleles and genotypes showed that statistically significant differences in the distribution of the minor allele rs11385942 of the *LZTFL1* gene in the subpopulation of Permsky and Arkhangelsky Bashkirs are present with the populations of Asia, Africa and America (Supplementary table 1). At the same time, there is no difference between the Burzyan Bashkirs and the populations of Siberia. Of particular interest in such a difference in the distribution of the GA risk allele in various subpopulations of the Bashkirs can be observed both in connection with the wide area of residence and in connection with the different genetic background of different subethnic groups of the Bashkirs. Of particular interest are the differences in the distribution of the risk allele GA with Asian populations, since, according to one hypothesis, the origin of the Bashkirs is from South Siberia (Yanguzin *et al.*, 2007) moreover after bioinformatic processing of genome-wide analysis data, using the ADMIXTURE method, it was shown that the Bashkirs typically have the largest Siberian and

East Asian component in their genetic composition among all the populations of the VUR (Yunusbayev *et al.*, 2015). This suggests that the distribution of alleles of the rs11385942 variant of the *LZTFL1* gene is consistent with data from European populations, for which a strong association of this variant with the severe course of COVID-19 was shown.

When studying the rs657152 variant of the *ABO* gene, we showed that the strongest differences in the distribution of alleles were recorded between all studied subpopulations of the Bashkirs and Native Americans, in whose gene pool one of the alleles was completely absent. Also, interestingly, statistically significant differences are reported between the subpopulations of Permsky and Arkhangelsky Bashkirs and the population of Tatars ($p = 0.005$) (Supplementary table 2), belonging to the Kipchak group of Turkic languages common with the Bashkirs and living in close proximity. However, the easternmost subpopulation of the Burzyan Bashkirs does not show such a difference ($p = 0.587$). We also found statistically significant differences the subpopulations of Permsky and Arkhangelsky Bashkirs with the populations of the Western Caucasus, Mongolia, and the Middle East, but none were found with the populations of Africa, Europe, and East Asia (with the exception of the Mongols). In turn, the Burzyan Bashkirs showed a much more complex picture with statistically significant differences from various Finno-Ugric, Turkic and Caucasian populations.

When studying the variant rs2109069 of the *DPP9* gene, we found that the risk allele A occurs with a frequency of 22.2% and the risk genotype AA with a frequency of 1.3%. With a slightly lower frequency of the genotype (22.3%), but a higher frequency of the A allele (6.38%) in the subpopulation of the Arkhangelsky Bashkirs and with a frequency of the AA genotype of 15.6% and the A allele of 4.17% in the subpopulation of the Burzyan Bashkirs. It should be noted that the highest rates are observed in European populations, where the frequency of allele A is more than 30%, particularly in the Spanish population (32.7%), a mixed sample from Britain and Scotland

(35.2%) and the population of Finns (36.4%). We found that the all studied subpopulations in this paper in terms of the distribution of alleles is statistically significantly different from the listed European populations. We also demonstrated differences in the distribution of alleles between the studied subpopulation of Permsky and Arkhangelsky Bashkirs and the populations of East Asia: Chinese Dai in Xishuandbanna, Han Chinese in Beijing, and Japanese in Tokyo, which also confirms the differences between the populations of the VUR and East Asia. No statistically significant differences with the indigenous populations of Africa were found.

Thus, it was found that the lowest frequency of risk alleles of the rs11385942 *LZTFL1* and rs2109069 *DPP9* variants was found in the subpopulation of Burzyan Bashkirs, 7.3% and 15.6%, respectively, while the highest frequency was observed for the rs657152 locus of the *ABO* gene in this subpopulation – 53.1%. Statistically significant differences between the studied subpopulations were not identified. However, statistically significant differences were shown with other populations included in the study. Thus, it was shown that there were no statistically significant differences between the Burzyan Bashkirs and the populations of Southern Siberia at the rs11385942 *LZTFL1* locus, in contrast to the subpopulations of the Permsky and Arkhangelsky Bashkirs.

The significance of polymorphic variants in the pathogenesis of COVID-19, the distribution of which was studied in this work, was confirmed in a number of genome-wide studies (GWAS) and meta-analyses (Dzik *et al.*, 2020, Ellenhaus *et al.*, 2020, Downes *et al.*, 2021, Pairo-Castineira *et al.*, 2021, Wu *et al.*, 2021; Zhao *et al.*, 2021; Latz *et al.*, 2020, Angulo-Aguado *et al.*, 2022; Cruz *et al.*, 2022; Namkoong *et al.*, 2022). The study of the distribution of their risk alleles in different populations can contribute to early diagnosis and more accurate prediction of the risks associated with the severe course of the disease. Also, it can be assumed that such a distribution of alleles affects not only the risk of severe COVID-19, but also the effectiveness of treatment and, consequently, the survival of patients.

Conflict of interest: the authors declare no conflict of interest.

Acknowledgments

This work has been supported by the grants the Russian Science Foundation, RSF 21-74-00104, in the part of genotyping. and

statistical analysis; was funded by Ministry of Science and Higher Education of Russian Federation (№ 075-03-2021-193/5) and Ministry of Education and Science of the Republic of Bashkortostan (agreement no. 1, December 28, 2021) in the part of collection of biological materials.

References

- ANGULO-AGUADO M., CORREDOR-ORLANDELLI D., ... & ORTEGA-RECALDE O. (2022): Association Between the LZTFL1 rs11385942 Polymorphism and COVID-19 Severity in Colombian Population. *Frontiers in Medicine* **9**, 910098.
- BALANOVSKY O.P., PETRUSHENKO V.V., MIRZAEV K.B., ... & SYCHEV D. (2021): Variation of Genomic Sites Associated with Severe Covid-19 Across Populations: Global and National Patterns. *Pharmacogenomics and Personalized Medicine* **14**, 1391–1402.
- CRUZ R., DIZ-DE ALMEIDA S., ... & SCOURGE COHORT GROUP; HOSTAGE COHORT GROUP; GRACE COHORT GROUP, GUILLEN-NAVARRO E., AYUSO C., GONZÁLEZ-NEIRA A., RIANCHO J.A., ROJAS-MARTINEZ A., FLORES C., LAPUNZINA P. & CARRACEDO A. (2022): Novel genes and sex differences in COVID-19 severity. *Human Molecular Genetics* **31**(22), 3789–3806.
- DOWNES D. J., CROSS A.R., HUA P., ... & HUGHES J.R. (2021): Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. *Nature Genetics* **53**(11), 1606–1615.
- DZIK S., ELIASON K., MORRIS E.B., KAUFMAN R.M. & NORTH C.M. (2020): COVID-19 and ABO blood groups. *Transfusion* **60**(8), 1883–1884.
- ELLINGHAUS D., DEGENHARDT F., BUJANDA L., ... & KARLSEN T.H. (2020): Genomewide association study of severe Covid-19 with respiratory failure. *The New England Journal of Medicine* **383**(16), 1522–1534.
- FINGERLIN T.E., MURPHY E., ZHANG W., PELJTO A.L., ... & SCHWARTZ D.A. (2013): Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nature Genetics* **45**(6), 613–620.
- GRISWOLD A.R., BALL D.P., BHATTACHARJEE A., CHUI A.J., RAO S.D., TAABAZUING C.Y. & BACHOVCHIN D.A. (2019): DPP9's Enzymatic Activity and Not Its Binding to CARD8 Inhibits Inflammasome Activation. *ACS Chemical Biology* **14**(11), 2424–2429.
- KUZEEV R.G. (1974): *The origin of the Bashkir people*. Moscow: Science, 570 pp. (In Russ.)
- LATZ C.A., DECARLO C., ... & DUA A. (2020): Blood type and outcomes in patients with COVID-19. *Annals of Hematology* **99**(9), 2113–2118.
- MATHEW C.G. (1984): The isolation of high molecular weight eukaryotic DNA. *Methods in Molecular Biology* **2**, 31–34.
- MIRZHANOVA S.F. (2006): *Northwestern dialect of the Bashkir language*. Ufa: Kitap, 210-288 p. (In Russ.)
- NAMKOONG H., BIOBANK JAPAN PROJECT, OMAE Y., NANNYA Y., UENO T., KATAYAMA K., AI M., FUKUI Y., KUMANOGOH A., SATO T., HASEGAWA N., TOKUNAGA K., ISHII M., KOIKE R., KITAGAWA Y., KIMURA A., IMOTO S., MIYANO S., OGAWA S., KANAI T., FUKUNAGA K. & OKADA Y. (2022): DOCK2 is involved in the host genetics and biology of severe COVID-19. *Nature* **609**(7928), 754–60.
- NEBOLSIN P. (1852): Bashkirs of the 1st canton (Osinsky district). *Bulletin of the Russian Geographical Society*. Department of Ethnography, St. Petersburg, 18. (In Russ.)
- PAIRO-CASTINEIRA E., CLOHISEY S., KLARIC L., ... & BAILLIE J.K. (2021): Genetic mechanisms of critical illness in COVID-19. *Nature* **591**(7848), 92–98.
- THE 1000 GENOMES PROJECT CONSORTIUM (2012): An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**(7422), 56–65.
- TATISHCHEV V.N. (1962) *Russian history*. Moscow: L. Publishing House of the Academy of Sciences of the USSR V. 1, 500 pp. (In Russ.)
- TROFIMOVA N.V. (2015): *Variability of mitochondrial DNA and Y-chromosome in the populations of the Volga-Ural region*. Dissertation, Ufa, 192 pp. (In Russ.)

- WU P., DING L., LI X., LIU S., ... & WANG C. (2021): Trans-ethnic genome-wide association study of severe COVID-19. *Communications Biology* **4**(1), 1034.
- YANGUZIN R.Z., KHISAMITDINOVA F.G. (2007): *Indigenous peoples of Russia. Bashkirs*. Ufa: Kitap, 352 pp. (In Russ.)
- YUNUSBAYEV B., METSPALU M., METSPALU E., VALEEV A., LITVINOV S., VALIEV R., AKHMETOVA V., BALANOVSKA E., BALANOVSKY O., TURDIKULOVA S., ... & VILLEMS R. (2015): The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genetics* **11**(4), e1005068.
- YUSUPOV R.M. (1987): On the history of the Gainin Bashkirs. *Dawn (Barda)* 127, 128. (In Russ.)
- YUSUPOV R.M. (1991): *Historical anthropology of the Southern Urals and the formation of the racial type of the Bashkirs*, Ufa (In Russ.)
- YUSUPOV R.M. (2002): Anthropological composition of the Bashkirs and its formation. *Bashkirs. Ethnic history and traditional culture*, Ufa, 11–20 (In Russ.)
- YUSUPOV R.M. (2006): Ethnology of the Bashkirs at the turn of the millennium. *Problems of ethnogenesis and ethnic history of the Bashkir people*, Ufa, 95–101 (In Russ.)
- YUSUPOV R.M., KHUSNUTDINOVA E.K. ... & RAKHMANGULOV A.A. (2009): *Anthropology and population genetics of the Permian Bashkirs*. Monograph, Ufa, 188 pp. (In Russ.)
- ZEBERG H. & PÄÄBO S. (2020): The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* **587**(7835), 610–612.
- ZHANG H., MAQSUDI S., RAINCZUK A., DUFFIELD N., LAWRENCE J., KEANE F.M., JUSTASCHUCH D., GEISS-FRIEDLANDER R., GORRELL M.D. & STEPHENS A.N. (2015): Identification of novel dipeptidyl peptidase 9 substrates by two-dimensional differential in-gel electrophoresis. *FEBS Journal* **282**(19), 3737–3757.
- ZHAO J., YANG Y., HUANG H., ... & WANG P.G. (2021): Relationship between the ABO Blood Group and the COVID-19 Susceptibility. *Clinical Infectious Diseases* **73**(2), 328–331.

Supplementary table 1

Frequencies of the minor allele rs11385942 *LZTFLI* gene in the studied samples of Permsky, Burzyan and Arkhangelsky Bashkirs peoples, as well as in some subpopulation of the world

Population	N	Minor allele frequency%	Permsky Bashkirs	Burzyan Bashkirs	Arkhangelsky Bashkirs
Permsky Bashkirs	80	10.5		0.506	0.954
Burzyan Bashkirs	48	7.3	0.506		0.429
Arkhangelsky Bashkirs	47	11.7	0.954	0.429	
Russians, northernmost ¹	140	6	0.125		
Ukrainians ¹	158	16	0.137	0.044	0.373
Komi and Udmurts ¹	168	15	0.169	0.077	0.540
Chuvash and Mari ¹	106	10	0.957	0.599	0.786
Mordovians ¹	80	5	0.096	0.630	0.087
Tatars ¹	104	8	0.295	0.098	0.532
Bashkirs ¹	86	12	0.779	0.292	0.940
Karelians and Veps ¹	118	14	0.405	0.130	0.711
Altaians ¹	154	4	0.008	0.273	0.009
Tuvinians and Tofalars ¹	110	4	0.022	0.360	0.023
Buryats, Khamnigans and Yakuts ¹	114	4	0.017	0.323	0.018
Chukchi, Koryaks and Itelmens ¹	134	1	0.00002	0.005	0.00002
Far East (Nanais, Ulchi, Nivkh, Evens) ¹	168	2	0.00009	0.027	0.0001
Khanty, Mansi and Nenets ¹	106	2	0.0007	0.042	0.0007
Siberian Tatars ¹	136	2	0.0001	0.024	0.0002
West Caucasus ¹	174	11	0.957	0.394	0.977
East Caucasus (Dagestan) ¹	158	11	0.995	0.378	0.986
Central Caucasus	128	5	0.053	0.589	0.053
Transcaucasia (South Caucasus) ¹	154	14	0.380	0.120	0.697
Khalkha Mongols ¹	98	1	0.0002	0.011	0.0001
Uzbeks, Turkmens and Kyrgyz ¹	160	4	0.009	0.305	0.011
Tajiks, Pamiris and Yaghnobis ¹	144	15	0.255	0.080	0.542
South Indians ¹	96	22	0.008	0.003	0.055
Pakistanis ¹	336	15	0.151	0.059	0.484
Italians ¹	88	13	0.602	0.211	0.897
Finns ²	99	10.1	0.990	0.571	0.832
North Europeans ¹	66	9	0.811	0.808	0.677
Central Europeans ¹	90	11	0.975	0.421	0.957
Anatolian and Levant populations ¹	114	11	0.952	0.419	0.997
Ethiopians ¹	38	8	0.670	0.887	0.571
Africa ²	661	5.3	0.011	0.548	0.018
Native Americans from South America ¹	58	0	0.0008	0.008	0.0004
Japanese ¹	56	0	0.001	0.010	0,0005

Note: Bold indicates statistically significant differences ($P < 0.05$). ¹ – (Balanovsky *et al.*, 2021); ² – (The 1000 Genomes Project Consortium, 2012)

Frequencies of the minor allele rs657152 AB0 gene in the studied samples of Permsky, Burzyan and Arkhangelsky Bashkirs peoples, as well as in some subpopulation of the world

Population	N	Minor allele frequency%	Permsky Bashkirs	Burzyan Bashkirs	Arkhangelsky Bashkirs
Permsky Bashkirs	80	42.4		0.128	0.902
Burzyan Bashkirs	48	53.1	0.128		0.189
Arkhangelsky Bashkirs	47	42.4	0.902	0.189	
Russians. northernmost ¹	140	42	0.942	0.062	0.959
Ukrainians ¹	158	51	0.081	0.797	0.189
Komi and Udmurts ¹	168	40	0.579	0.028	0.728
Chuvash and Mari ¹	106	46	0.474	0.318	0.637
Mordovians ¹	80	38	0.425	0.027	0.573
Tatars ¹	104	57	0.005	0.587	0.025
Bashkirs ¹	86	42	0.906	0.099	0.984
Karelians and Veps ¹	118	36	0.193	0.006	0.327
Altaians ¹	154	48	0.253	0.452	0.350
Tuvinians and Tofalars ¹	110	38	0.396	0.019	0.468
Buryats, Khamnigans and Yakuts ¹	114	42	0.938	0.089	0.960
Chukchi, Koryaks and Itelmens ¹	134	42	0.946	0.083	0.956
Far East (Nanais, Ulchi, Nivkh, Evens) ¹	168	38	0.348	0.012	0.507
Khanty, Mansi and Nenets ¹	106	38	0.403	0.020	0.555
Siberian Tatars ¹	136	41	0.788	0.043	0.911
West Caucasus ¹	174	27	0.0005	0.000003	0.005
East Caucasus (Dagestan) ¹	158	52	0.053	0.925	0.141
Central Caucasus	128	42	0.950	0.086	0.952
Transcaucasia (South Caucasus) ¹	154	47	0.345	0.358	0.514
Khalkha Mongols ¹	98	32	0.044	0.0009	0.109
Uzbeks, Turkmen and Kyrgyz ¹	160	51	0.081	0.795	0.189
Tajiks, Pamiris and Yaghnobis ¹	144	46	0.496	0.262	0.663
South Indians ¹	96	40	0.649	0.049	0.789
Pakistanis ¹	336	45	0.577	0.163	0.745
Italians ¹	88	35	0.172	0.006	0.293
Finns ²	5287	47.8	0.181	0.351	0.360
North Europeans ¹	66	39	0.504	0.042	0.650
Central Europeans ¹	90	42	0.959	0.109	0.939
Anatolian and Levant populations ¹	114	54	0.026	0.989	0.082
Ethiopians ¹	38	50	0.279	0.800	0.416
Africa ²	661	51.5	0.743	0.079	0.883
Native Americans from South America ¹	58	0	0	0	0
Japanese ¹	56	32	0.084	0.003	0.162

Note: Bold indicates statistically significant differences ($P < 0.05$). ¹ – (Balanovsky *et al.*, 2021); ² – (The 1000 Genomes Project Consortium. 2012)

Frequencies of the minor allele rs2109069 DPP9 gene in the studied samples of Permsky, Burzyan and Arkhangelsky Bashkirs peoples, as well as in some subpopulation of the world

Population	N	Minor allele frequency%	Permsky Bashkirs	Burzyan Bashkirs	Arkhangelsky Bashkirs
Permsky Bashkirs	80	22.5		0.241	0.899
Burzyan Bashkirs	48	15.6	0.241		0.319
Arkhangelsky Bashkirs	47	22.3	0.899	0.319	
African Caribbean in Barbados ²	96	16.7	0.213	0.955	0.318
Esan in Nigeria ²	99	20.2	0.690	0.433	0.790
Yoruba in Ibadan, Nigeria ²	108	16.2	0.159	0.969	0.258
Luhya in Webuye, Kenya ²	99	23.7	0.881	0.148	0.908
Mende in Sierra Leone ²	85	17.6	0.335	0.801	0.446
Gambian in Western Division, The Gambia ²	113	22.1	0.971	0.239	0.916
Colombian in Medellin, Colombia ²	94	25.5	0.594	0.080	0.659
Peruvian in Lima, Peru ²	85	12.9	0.033	0.672	0.070
Puerto Rican in Puerto Rico ²	104	24	0.825	0.130	0.860
Chinese Dai in Xishuandbanna, China ²	93	10.8	0.005	0.324	0.016
Han Chinese in Beijing, China ²	103	13.6	0.037	0.769	0.083
Southern Han Chinese, China ²	105	18.1	0.358	0.714	0.479
Japanese in Tokyo, Japan ²	104	12.5	0.016	0.575	0.044
Kinh in Ho Chi Minh City, Vietnam ²	99	14.6	0.075	0.963	0.143
Finnish in Finland ²	99	36.4	0.004	0.0004	0.023
British in England and Scotland ²	91	35.2	0.010	0.001	0.040
Iberian populations in Spain ²	107	32.7	0.030	0.003	0.089
Toscani in Italy ²	107	26.6	0.427	0.048	0.512
Bengali in Bangladesh ²	86	23.3	0.974	0.158	0.986
Punjabi in Lahore, Pakistan ²	96	18.8	0.462	0.623	0.578

Note: Bold indicates statistically significant differences ($P < 0.05$). ² – (The 1000 Genomes Project Consortium, 2012).