2021 • 25 • 8

Print ISSN 2500-0462
Online ISSN 2500-3259

Научный рецензируемый журнал

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.
Периодичность 8 выпусков в год
DOI 10.18699/VJ21.091

Учредители

Сибирское отделение Российской академии наук

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук»

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

2021 • 25 • 8

Print ISSN 2500-0462
Online ISSN 2500-3259

Scientific Peer Reviewed Journal

VAVILOV JOURNAL OF GENETICS AND BREEDING

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

Founded in 1997
Published 8 times annually
DOI 10.18699/VJ21.091

Founders

Siberian Branch of the Russian Academy of Sciences

Federal Research Center Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences

The Vavilov Society of Geneticists and Breeders

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

СОДЕРЖАНИЕ • 2021 • 25 • 8

Генетика и селекция растений

Генетика животных

Генетика человека

Медицинская генетика

Генетические коллекции

Актуальные технологии

VAVILOV JOURNAL OF GENETICS AND BREEDING

CONTENTS • 2021 • 25 • 8

Вавиловский журнал генетики и селекции. 2021;25(8):797-804
DOI 10.18699/VJ21.092

Original Russian text www.bionet.nsc.ru/vogis/

Genetic polymorphism of local Abkhazian grape cultivars

E.T. Ilnitskaya[1] ✉, M.V. Makarkina[1], I.V. Stepanov[1], I.I. Suprun[1], S.V. Tokmakov[1], V.Ch. Aiba[2], M.A. Avidzba[2], V.K. Kotlyar[1]

[1] North-Caucasian Federal Scientific Center of Horticulture, Viticulture, Winemaking, Krasnodar, Russia
[2] Institute of Agriculture of the Academy of Sciences of Abkhazia, Sukhum, Abkhazia
✉ Ilnitskaya79@mail.ru

Abstract. Local grape cultivars from different countries of the world are an important part of the gene pool of this culture. Of particular interest are the genotypes of the most ancient regions of viticulture. The territories of the subtropical zone of Georgia and the central part of Abkhazia belong to one of the centers of origin of the cultural grapevine. The purpose of the work was to genotype native Abkhazian grape cultivars, to study their genetic diversity based on DNA profiling data and to compare them with the genotypes of local varieties of other viticultural regions. Samples of plants were taken on the territory of the Republic of Abkhazia in private farmsteads and in the collection of the agricultural firm "Vina i Vody Abkhazii" ("Wines and Waters of Abkhazia"). The genotyping of the Abkhazian cultivars Avasirhva, Agbizh, Azhapsh, Azhizhkvakva, Azhikvaca, Atvizh, Atyrkuazh, Achkykazh, Kachich was carried out using 14 DNA markers, 9 of which are standard microsatellite markers recommended for the identification of grape varieties. To improve our knowledge about the sizes of the identified alleles, we used the DNA of grape cultivars with a known allelic composition at the analyzed loci. Statistical analysis of the data showed that the observed heterozygosity for the analyzed loci exceeded expected values, which indicates a genetic polymorphism of the studied sample of varieties. Evaluation of genetic similarity within the analyzed group based on the results of genotyping at 14 loci showed that the cultivars Kachich and Azhapsh differed from the other Abkhazian varieties. The obtained DNA profiles of the Abkhazian cultivars were checked for compliance with DNA-fingerprints of grape varieties in the *Vitis* International Variety Catalogue. The Georgian varieties Azhizhkvakva and Tsitska turned out to be synonyms according to DNA profiles, two varieties from the Database (Italian Albana bianca and Georgian Ojaleshi) have differences in DNA-fingerprints from the varieties Atyrkuazh and Azhikvatsa only in one allele, respectively. When comparing the identified Abkhazian grape genotypes, their difference from the sample of Dagestan, Don, Greek, Turkish, Italian, Spanish, and French varieties and genetic similarity with the genotypes of Georgian grapes were shown.
Key words: *Vitis vinifera* L.; local grape varieties; genetic diversity; SSR-loci.

For citation: Ilnitskaya E.T., Makarkina M.V., Stepanov I.V., Suprun I.I., Tokmakov S.V., Aiba V.Ch., Avidzba M.A., Kotlyar V.K. Genetic polymorphism of local Abkhazian grape cultivars. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8):797-804. DOI 10.18699/VJ21.092

Генетический полиморфизм аборигенных абхазских сортов винограда

Е.Т. Ильницкая[1] ✉, М.В. Макаркина[1], И.В. Степанов[1], И.И. Супрун[1], С.В. Токмаков[1], В.Ш. Айба[2], М.А. Авидзба[2], В.К. Котляр[1]

[1] Северо-Кавказский федеральный научный центр садоводства, виноградарства, виноделия, Краснодар, Россия
[2] Институт сельского хозяйства Академии наук Абхазии, Сухум, Абхазия
✉ Ilnitskaya79@mail.ru

Аннотация. Аборигенные сорта винограда различных стран мира – важная часть генофонда культуры. Особый интерес вызывают генотипы наиболее древних регионов виноградарства. Территории субтропической зоны Грузии и центральной части Абхазии относят к одному из центров возникновения культурной виноградной лозы. Целью работы было генотипирование аборигенных абхазских сортов винограда, изучение их генетического разнообразия на основе данных ДНК-профилирования и сравнение с генотипами местных сортов других регионов виноградарства. Образцы растений были отобраны на территории Республики Абхазия в частных подворьях и в коллекции агрофирмы «Вина и воды Абхазии». Генотипирование абхазских сортов Авасирхва, Агбиж, Ажапшь, Ажижкваква, Ажиквака, Атвижь, Атыркуажь, Ачкыкажь, Качич выполнено с помощью 14 ДНК-маркеров, девять из которых являются стандартными микросателлитными маркерами, рекомендованными для паспортизации сортов винограда. Для уточнения размеров идентифицированных аллелей в работе использовали ДНК сортов винограда с известным аллельным составом по анализируемым локусам. Статистический анализ данных показал, что фактическая гетерозиготность по анализируемым локусам превысила ожидаемую, что говорит о генетическом полиморфизме исследуемой выборки сортов. Оценка генетического сходства внутри анализируемой группы по результатам генотипирования по 14 локусам показала отличие сортов Качич и Ажапшь от остальных абхазских сортов. Полученные ДНК-профили абхазских сортов были проверены на предмет соответствия

ДНК-паспортам сортов винограда, представленных в международной базе данных. Грузинские сорта Ажижкваква и Цицка оказались синонимами по паспортам ДНК; у двух сортов из базы данных (итальянский Albana bianca и грузинский Оджалеши) обнаружены отличия в ДНК-паспортах от сортов Атыркуажь и Ажикваца соответственно только по одному аллелю. При сравнении идентифицированных абхазских генотипов винограда показано их отличие от выборки дагестанских, донских, греческих, турецких, итальянских, испанских и французских сортов и генетическое сходство с генотипами винограда Грузии.
Ключевые слова: *Vitis vinifera* L.; местные сорта винограда; генетическое разнообразие; SSR-локусы.

Introduction

The grapevines *Vitis vinifera* L. are cultivated by humans for already about 5000 years and it is the most economically important fruit crop in the world now. Local varieties from different regions of the world are an important part of the crop's gene pool. Of particular interest are the genotypes of the most ancient viticulture regions. Western Transcaucasia and, especially, the subtropical zone of Georgia and the central part of Abkhazia are recognized as one of the centers of the origin of cultural grapevine. Many ancient local varieties as well as wild grapevines are found in these regions.

Currently, molecular genetic methods are widely used to assess the diversity of the gene pool of cultivated plants, including grape. Microsatellite (SSR) markers are most often used for these purposes. The standard set of 9 microsatellite loci has been developed for DNA fingerprinting of grape genotypes (Bowers et al., 1999; This et al., 2004; VIVC, 2021). The use of DNA markers in order to identify grape varieties and study their polymorphism made it possible to clarify many questions regarding homonym and synonym varieties, as well as to determine the most genetically similar and distant genotypes (Crespan, Milani, 2001; Fossati et al., 2001; Vokurka et al., 2003; Santiago et al., 2005; Moreno-Sanz et al., 2008; Cipriani et al., 2010; Goryslavets et al., 2015; Raimondi et al., 2015; Mandić et al., 2019; Papapetrou et al., 2020; Pastore et al., 2020). For example, using a genetic analysis of 35 autochthonous varieties of Bosnia and Herzegovina at 9 standard microsatellite loci, several synonyms and homonyms were identified. Comparison of genotypes from Bosnia and Herzegovina with Croatian grape varieties also revealed synonyms and homonyms among the two groups (Mandić et al., 2019).

PCR analysis of Crimean local varieties revealed that the Shabash, Manzhil al and Shabash krupnoyagodnyi varieties have identical DNA profiles. The variety Shabash krupnoyagodnyi is a clone of Shabash, and variety Manzhil al is a synonym of Shabash (Goryslavets et al., 2015).

178 grape varieties ranging from widely cultivated to nearly extinct, harvested in Emilia-Romagna (Northern Italy), were analyzed by 10 microsatellite markers (Pastore et al., 2020). The data obtained showed that, in this region, there are varieties cultivated in other regions of Italy and in other countries, but under different local names; however, unique genotypes were also identified. Of the 122 unique genotypes identified, 62 are not described in the literature, except for mentions in historical documents. They probably belong to the local gene pool and, possibly, are the autochthons of this region.

A collection of 1005 grape samples was genotyped at 34 microsatellite loci (SSR) in order to analyze genetic diversity and origins study (Cipriani et al., 2010). The comparison of molecular profiles revealed 200 groups of synonymies. The list was corrected taking into account full synonyms, which reduced the database to 745 unique genotypes.

Viticulture and winemaking are of particular importance for the people of Abkhazia. The antiquity of viticulture and winemaking of this region is evidenced by many archaeological and paleobotanical finds (Chamagua, 1968). The local type of grape culture even exists here: maglari – the cultivation of vines on trees. The autochthonous varieties of Abkhazia are characterized by a late ripening period and the ability to preserve the crop for a long time on the bushes. The fame of Abkhazian wines is mainly associated with the varieties Auasyrhua (Avasirhva), Kachich (Kachichi), Amlahu.

In the history of the viticulture of Abkhazia, there were periods due to political and economic reasons, when this industry either developed intensively, or fell into decline. The greatest damage was done when phylloxera appeared in the region: the plantings of local varieties were actively destroyed; significant areas were planted with a resistant American variety Isabella. Many varieties have been lost and are preserved in private farmsteads as single bushes. However, interest to local varieties is growing, and molecular genetic studies can be used for DNA profiling (fingerprinting) of local varieties, assessment of genotype polymorphism, and clarification of the origin of unknown cultivars.

The purpose of the study was the genotyping of Abkhazian grape varieties and assessment of genetic diversity of the studied sample based on the polymorphism of microsatellite loci.

Materials and methods

In this study, plants corresponding to the ampelographic descriptions of Abkhazian varieties given in the "Ampelography of the USSR…" were taken (Frolov-Bagrev, 1953, 1954; Negrul et al., 1963, 1970). Some samples of the varieties included in the research were collected from several geographical points of Abkhazia (Kachich, Avasirhva, Agbizh), others were taken from the collection of the agricultural company "Vina i Vody Abkhazii" (Atyrkuazh, Azhikvaca, Atvizh, Azhizhkvakva, Achkykazh, Azhapsh) (Table 1).

DNA was isolated from the crown leaves from annual shoots of three to five typical variety plants by the method using the CTAB buffer (Rogers, Bendich, 1985). A standard set of 9 microsatellite (SSR) markers recommended for identification of grapevine genotypes was used for DNA fingerprinting (Bowers et al., 1999; This et al., 2004; VIVC, 2021). For a more complete assessment of the studied sample's polymorphism, 5 SSR markers were additionally included in the work (UDV737, GF09-46, ScORGF15-02, GF15-42, CenGen6) (Di Gaspero et al., 2012; Schwander et al., 2012; van Heerden et al., 2014; Zendler et al., 2017). DNA fingerprinting of the Azhizhkvakva, Kachich and Avasirhva varieties for 9 standard

Table 1. List of grape varieties used in the study

No.	Grape varieties	Country of origin	No.	Grape varieties	Country of origin
1	Narince	Turkey	34	Sibirkovyi	Don (Russia)
2	Vasiliko		35	Tsimlyanskii chernyi	
3	Bogazkere		36	Varyushkin	
4	Yapincak		37	Krasnostop zolotovskii	
5	Papaskarasi		38	Bruskovatenkii	
6	Hasandede beyazi		39	Shilokhvostyi	
7	Horoz Karasi		40	Cabernet Sauvignon	France
8	Achladi	Greece	41	Chardonnay Blanc	
9	Kotsifali		42	Merlot	
10	Mandilaria		43	Sauvignon Blanc	
11	Akiki		44	Aligote	
12	Daphnia		45	Pinot noir	
13	Fokiano		46	Montepulciano	Italy
14	Kakotrygis		47	Sangiovese	
15	Mtsvane kakhuri	Georgia	48	Nebbiolo	
16	Rkatsiteli		49	Lacrima	
17	Tavkveri		50	Vermentino	
18	Tsolikouri		51	Garganega	
19	Alexandroouli		52	Tempranillo	Spain
20	Chinuri		53	Parellada	
21	Ojaleshi		54	Shiava grossa	
22	Saperavi		55	Garnacha	
23	Meskhuri mtsvane		56	Albillo Real	
24	Alyi Terskii	Dagestan (Russia)	57	Zalema	
25	Asyl Kara		58	Kachich	Abkhazia
26	Rish baba		59	Atyrkuazh	
27	Sarakh		60	Avasirhva	
28	Shavrany		61	Atvizh	
29	Bayat Kapy		62	Azhizhkvakva	
30	Agadai		63	Achkykazh	
31	Gok ala		64	Azhapsh	
32	Tavlinskii pozdnii		65	Agbizh	
33	Kukanovskii	Don (Russia)	66	Azhikvaca	

microsatellite loci was performed by us earlier (Ilnitskaya et al., 2019–2021).

PCR was carried out in a 20 μL reaction mixture containing 50 ng of genomic DNA, 1.5 units of Taq polymerase, 1× buffer for Taq polymerase with ammonium sulfate and magnesium, 2 mM $MgCl_2$, 0.2 mM each dNTP (deoxynucleotide triphosphates) (SibEnzyme-M, Moscow) and 200 μM of each primer (Syntol, Moscow) using a BioRad device (USA), following these protocols: initial denaturation – 10 s at +95 °C; then 34 synthesis cycles: denaturation – 10 s

at +95 °C, primer annealing – 30 s at +55 °C for markers VVS2, VVMD5, VVMD7, VVMD27, UDV737, CenGen6, at +58 °C – VrZAG62, VrZAG79, ScORGF15-02, GF15-42, and at +60 °C – VVMD25, VVMD28, VVMD32, GF09-46, elongation – 30 s at +72 °C; the final cycle (final elongation) – 3 min at +72 °C.

Separation and analysis of the PCR products lengths was carried out by capillary electrophoresis using ABI Prism 3130 genetic analyzer. The amplified fragments were aligned relative to the control (reference) genotypes that we included in the work, with a known allelic composition for the analyzed loci: Pinot noir (for markers VVS2, VVMD5, VVMD7, VVMD25, VVMD27, VVMD28, VVMD32, VrZAG62, VrZAG79), Saperavi severnyy (GF09-46), Seyve Villard 12-375 (UDV737), Regent (ScORGF15-02, GF15-42, CenGen6).

Statistical processing of data on loci polymorphism in the studied sample was carried out using the program GenAlEx 6.5 (Peakall, Smouse, 2012). Genetic relationships were assessed with the PAST 2.17c program using the pairwise within-group unweighted mean (UPGMA) and principal coordinate methods (PCoA) (Hammer et al., 2001).

To study the genetic similarity of the autochthonous Abkhazian varieties with the grapevine's local gene pool from other viticulture zones, we included a sample of varieties that are classified as aboriginal genotypes of Georgia, Greece, Dagestan, Don (Rostov region, Russian Federation), Spain, Italy, France, Turkey (see Table 1). These regions of viticulture also have an ancient history of *V. vinifera* L. cultivation, geographical proximity or historical ties with Abkhazia. The DNA profiles of local varieties genotypes for nine SSR loci, standard for genotyping *V. vinifera*, were taken by us from the international database *Vitis* International Variety Catalog (VIVC). Bayesian analysis was carried out in the Structure 2.3.4 program using 66 genotypes (see Table 1), the optimal number of clusters was 3, established using the Evanno method, calculation was carried out in the online program Structure Harvester (Evanno et al., 2005; Earl, vonHoldt, 2012).

Results and discussion

The results of genotyping nine Abkhazian grape varieties (Kachich, Atyrkuazh, Avasirhva, Atvizh, Azhizhkvakva, Achkykazh, Azhapsh, Agbizh, Azhikvaca) by 14 microsatellite loci are presented in Table 2. The identified profiles of each variety for nine SSR loci (VVS2, VVMD5, VVMD7, VVMD25, VVMD27, VVMD28, VVMD32, VrZAG62, VrZAG79) were checked in the international database of grape varieties DNA fingerprints of *Vitis* International Variety Catalogue (VIVC, 2021). It was revealed that the allelic composition of the DNA profile of the Azhizhkvakva grape variety by nine microsatellite loci fully corresponds to the DNA profile of the Georgian aboriginal Tsitska grape variety, presented in the Database (Ilnitskaya et al., 2021). These varieties are similar in phenotypic characteristics and may be synonymous varieties or clone variations. The DNA profile of the Azhikvaca grape variety also showed a close similarity with the DNA

Table 2. DNA profiles of Abkhazian grape varieties

Variety	Allele size, base pairs													
	VVS2	VVMD5	VVMD7	VVMD25	VVMD27	VVMD28	VVMD32	VrZAG62	VrZAG79	UDV737	GF09-46	ScORGF15-02	GF15-42	CenGen6
Kachich	153 155	234 240	239 249	239 267	186 193	234 248	262 272	194 196	237 255	291 291	395 425	240 240	179 185	271 285
Atyrkuazh	133 143	228 234	247 249	241 255	180 190	234 236	250 272	194 200	237 251	295 295	425 425	195 240	177 177	263 277
Avasirhva	141 145	234 242	239 249	239 249	184 190	234 248	248 262	200 204	251 257	285 285	395 413	240 240	193 199	273 299
Atvizh	137 145	228 234	239 239	239 249	184 193	236 236	258 262	188 188	251 255	285 295	413 425	240 240	193 199	263 301
Azhizhkvakva	143 145	228 236	239 253	239 255	186 188	236 258	262 272	194 196	251 251	285 289	423 425	240 240	193 197	273 277
Achkykazh	143 145	228 228	239 249	239 255	186 193	236 251	262 262	200 204	251 255	285 295	395 413	240 240	179 185	277 303
Azhapsh	123 155	234 234	239 249	240 267	184 193	226 248	262 272	196 204	247 255	285 291	395 423	240 240	179 179	275 289
Agbizh	143 153	234 240	233 239	239 267	180 190	234 236	248 262	188 204	237 239	285 285	395 425	195 240	193 193	263 301
Azhikvaca	141 153	234 242	241 247	239 255	180 184	228 236	262 262	194 208	237 251	285 289	395 395	240 240	177 179	277 283

profile of the Georgian aboriginal Ojaleshi grape variety: the difference was revealed only in one allele at VrZAG79 locus (VIVC, 2021).

Since Georgia is the closest neighbor geographically and also a country with an ancient viticulture culture, the genetic similarity of native varieties of these regions is quite expected. However, we also revealed the similarity of the Atyrkuazh variety with a local Italian grape variety Albana bianca: the DNA profiles are identical except for one allele at the VrZAG79 locus (VIVC, 2021). Albana bianca is a relatively common variety and is found under other names in different countries (VIVC, 2021). The study of the ampelographic characteristics of ancient Italian variety Albana bianca, presented in the literature, shows a certain phenotypic similarity with variety Atyrkuazh. Due to the lack of DNA samples of the Albana bianca variety, it was not possible to compare genotypes for a larger number of SSR loci and clarify their level of genetic similarity.

In general, the genotypes of local Abkhazian varieties showed a fairly high polymorphism. The average value of observed heterozygosity (Ho = 0.810) exceeded the expected value (He = 0.712) in the studied sample of nine Abkhazian grape varieties at 14 microsatellite loci (Table 3). The least polymorphic locus was ScORGF15-02: only 2 alleles were detected. 11 alleles were identified on the most polymorphic locus CenGen6. It is worth noting that in the DNA profile of the Azhapsh grape variety at VVMD25 locus a very rare allele of 240 base pairs (bp) was identified, which was previously described in only one of the varieties presented in VIVC Database.

To assess the genetic similarity of the studied Abkhazian varieties, a cluster analysis was carried out based on the data of SSR-genotyping. The analyzed varieties were divided into two clusters, one of which contains seven of the nine studied varieties and within which Agbizh, Atyrkuazh, Azhikvaca were grouped into a separate subcluster, and Avasirhva, Atvizh, Achkykazh, Azhizhkvakva were grouped into another subcluster (Fig. 1). It is important to note that the genotypes of varieties Azhapsh and Kachich were allocated into one cluster. Localization of nine Abkhazian varieties in the PCA showed that the Azhapsh and Kachich varieties are grouped together and are located farther from the other seven varieties (Fig. 2).

For a broader understanding of the genetic structure of the population of grape varieties in Abkhazia and the relationship with the world gene pool, they were compared with the genotypes of aboriginal varieties of other regions of viticulture (Georgia, Greece, Dagestan, Don (Rostov region of Russian Federation), Spain, Italy, France, Turkey). The DNA profiles of varieties on nine standard SSR loci were taken from the international VIVC Database. Bayesian analysis showed the greatest degree of similarity between Abkhazian varieties and Georgian grape varieties (Fig. 3). Moreover, among the group of Georgian varieties, three genotypes showed similarities with other groups of varieties (Tavkveri, Saperavi, Meskhuri mtsvane), while the group of Abkhazian varieties is more homogeneous (see Fig. 3, *a*).

Also, the group of French varieties taken in the study stands out for its uniformity and difference from others. In the group of Italian varieties, the genotype of the Nebbiolo variety is closest to the Abkhazian ones, it has many synonyms and is characterized by a late maturation period. Among the Greek varieties, we can note the Mandilaria variety genotype, which differs from the others in this group and is similar in structure to the varieties of Abkhazia. This may indicate the genetic

Table 3. Characteristics of microsatellite loci in the studied sample of Abkhazian grape varieties

Locus	Na	Ne	Ho	He
VVS2	8.000	6.231	1.000	0.840
VVMD5	5.000	3.306	0.778	0.698
VVMD7	6.000	3.375	0.889	0.704
VVMD25	6.000	4.050	1.000	0.753
VVMD27	6.000	5.400	1.000	0.815
VVMD28	7.000	4.154	0.889	0.759
VVMD32	5.000	2.656	0.778	0.623
VrZAG62	6.000	5.400	0.889	0.815
VrZAG79	6.000	3.857	0.889	0.741
UDV737	4.000	2.945	0.556	0.660
GF09-46	4.000	3.306	0.778	0.698
ScORGF15-02	2.000	1.246	0.222	0.198
GF15-42	6.000	4.765	0.667	0.790
CenGen6	11.000	8.100	1.000	0.877
Mean	5.857	4.199	0.810	0.712

Note. Na – total number of alleles per locus; Ne – number of effective alleles; Ho – observed heterozygosity; He – expected heterozygosity.

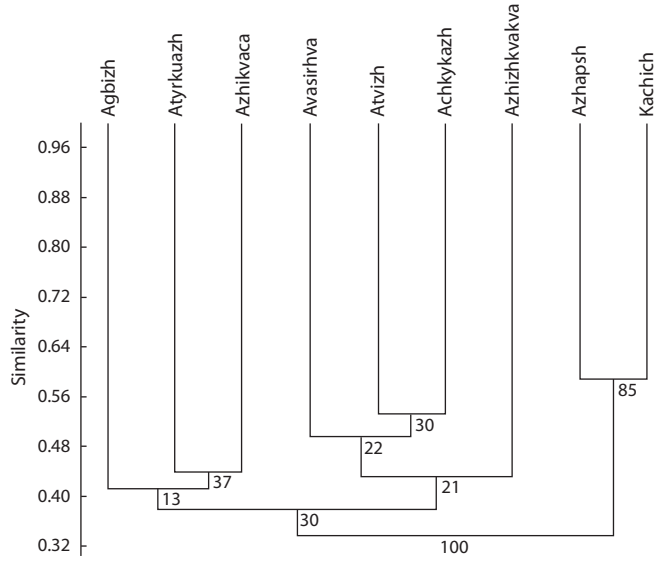


Fig. 1. Dendrogram of genetic similarity of Abkhazian local grape varieties according to DNA profiling data.

Branching nodes show bootstrap percentage values calculated on the basis of 50,000 random samples.

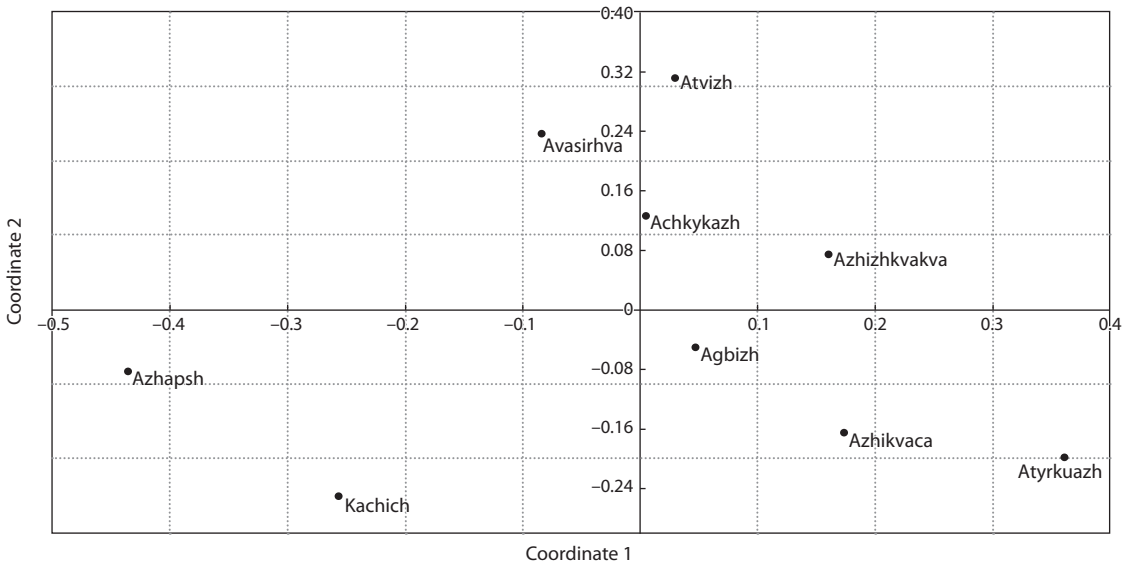


Fig. 2. Distribution of Abkhazian grape varieties according to PCA analysis of DNA profiling data.

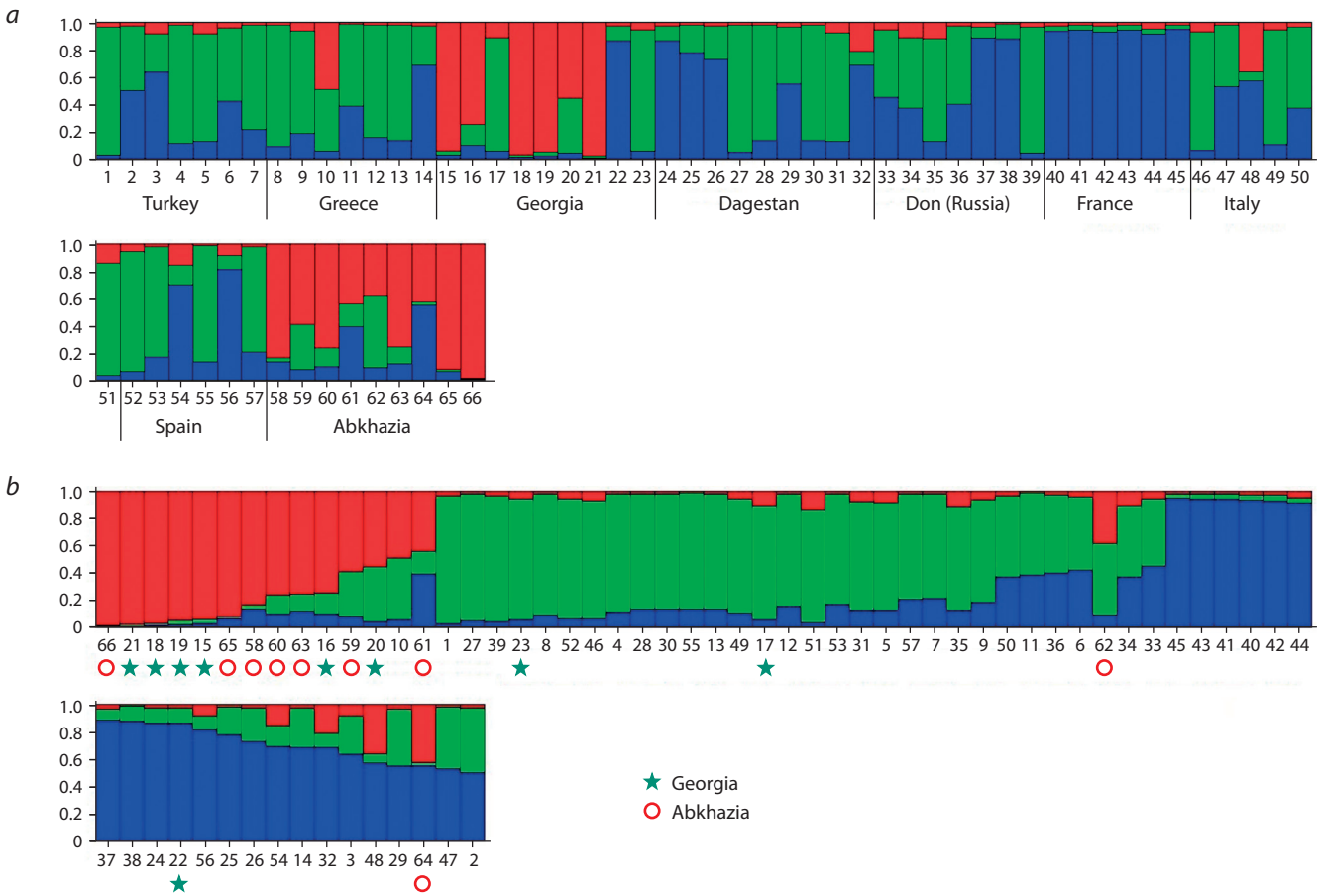


Fig. 3. Bar plot of the results from Bayesian analysis on 66 grape varieties, genotyping with nine SSR markers: *a*, grouping of varieties by origin; *b,* clustering by genetic similarity. The vertical axis denotes the probability of assigning each genotype to the putative clusters, indicated by different colors.

1 – Narince, 2 – Vasiliko, 3 – Bogazkere, 4 – Yapincak, 5 – Papaskarasi, 6 – Hasandede beyazi, 7 – Horoz Karasi, 8 – Achladi, 9 – Kotsifali, 10 – Mandilaria, 11 – Akiki, 12 – Daphnia, 13 – Fokiano, 14 – Kakotrygis, 15 – Mtsvane kakhuri, 16 – Rkatsiteli, 17 – Tavkveri, 18 – Tsolikouri, 19 – Alexandroouli, 20 – Chinuri, 21 – Ojaleshi, 22 – Saperavi, 23 – Meskhuri mtsvane, 24 – Alyi Terskii, 25 – Asyl Kara, 26 – Rish baba, 27 – Sarakh, 28 – Shavrany, 29 – Bayat Kapy, 30 – Agadai, 31 – Gok ala, 32 – Tavlinskii pozdnii, 33 – Kukanovskii, 34 – Sibirkovyi, 35 – Tsimlyanskii Chernyi, 36 – Varyushkin, 37 – Krasnostop zolotovskii, 38 – Bruskovatenkii, 39 – Shilokhvostyi, 40 – Cabernet Sauvignon, 41 – Chardonnay Blanc, 42 – Merlot, 43 – Sauvignon Blanc, 44 – Aligote, 45 – Pinot noir, 46 – Montepulciano, 47 – Sangiovese, 48 – Nebbiolo, 49 – Lacrima, 50 – Vermentino, 51 – Garganega, 52 – Tempranillo, 53 – Parellada, 54 – Shiava Grossa, 55 – Garnacha, 56 – Albillo Real, 57 – Zalema, 58 – Kachich, 59 – Atyrkuazh, 60 – Avasirhva, 61 – Atvizh, 62 – Azhizhkvakva, 63 – Achkykazh, 64 – Azhapsh, 65 – Agbizh, 66 – Azhikvaca.

relationships of the grape's gene pool in Greece and Abkhazia. It is known that in the times of Ancient Greece, the territory of Abkhazia was ruled by Greece for a period of time. During this period, perhaps, the exchange of the gene pool of grapes could also have taken place.

Figure 3, *b* shows that most varieties of Georgia and Abkhazia form a single group and differ from other genotypes. Two varieties, Azhizhkvakva and Azhapsh, are outside this group, but it is also noticeable that the hypothetical population marked in red, which prevails in the group of Georgian and Abkhazian varieties, makes a significant contribution to the structures of these genotypes.

Conclusion

Genotyping of local Abkhazian grape varieties Avasirhva, Agbizh, Azhapsh, Azhizhkvakva, Azhikvaca, Atvizh, Atyrkuazh, Achkykazh, Kachich was carried out using 14 DNA markers, including 9 generally accepted for DNA identification of grape varieties. The comparison of the identified DNA profiles on microsatellite loci VVS2, VVMD5, VVMD7, VVMD25, VVMD27, VVMD28, VVMD32, VrZAG62, VrZAG79 in the International Database shows the coincidence of the allelic compositions of the Azhizhkvakva cultivar with the Georgian local variety Tsitska, the Azhikvaca cultivar differs by one allele from the Georgian variety Ojaleshi. Also, the Atyrkuazh variety differs in one allele out of nine analyzed from the Italian grape variety Albana bianca.

The assessment of genetic structure of the population of grape varieties of Abkhazia and its relationship with the local gene pool of other regions of viticulture showed the similarity of Abkhazian varieties with Georgian and a difference from other groups of varieties of neighboring regions (Dagestan, Don, Turkey) and more remote regions of ancient viticulture (Greece, Italy, Spain and France). According to the results of the study, it is possible to assume that local varieties from the populations of the wild gene pool of grapevines are of autochthonous origin.

References

Bowers J., Dangl G., Meredith C.P. Development and characterization of additional microsatellite DNA markers for grape. *Am. J. Enol. Vitic.* 1999;50:243-246.

Chamagua E.I. Grapes of Abkhazia. Sukhumi: Alashara Publ., 1968. (in Russian)

Cipriani G., Spadotto A., Jurman I., Di Gaspero G., Crespan M., Meneghetti S., Frare E., Vignani R., Cresti M., Morgante M., Pezzotti M., Pe E., Policriti A., Testolin R. The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *Theor. Appl. Genet.* 2010; 121:1569-1585. DOI 10.1007/s00122-010-1411-9.

Crespan M., Milani N. The Muscats: a molecular analysis of synonyms, homonyms and genetic relationships within a large family of grapevine cultivars. *Vitis.* 2001;40(1):23-30. DOI 10.5073/vitis.2001.40.23-30.

Di Gaspero G., Copetti D., Coleman C., Castellarin S.D., Eibach R., Kozma P., Lacombe T., Gambetta G., Zvyagin A., Cindrić P., Kovács L., Morgante M., Testolin R. Selective sweep at the *Rpv3* locus during grapevine breeding for downy mildew resistance. *Theor. Appl. Genet.* 2012;124:227-286. DOI 10.1007/s00122-011-1703-8.

Earl D.A., vonHoldt B.M. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 2012;4(2):359-361. DOI 10.1007/s12686-011-9548-7.

Evanno G., Regnaut S., Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol*. 2005;14(8):2611-2620. DOI 10.1111/j.1365-294x.2005.02553.x.

Fossati T., Labra M., Castiglione S., Failla O., Scienza A., Sala F. The use of AFLP and SSR molecular markers to decipher homonyms and synonyms in grapevine cultivars: the case of the varietal group known as "Schiave". *Theor. Appl. Genet.* 2001;102:200-205. DOI 10.1007/s001220051636.

Frolov-Bagrev A.M. (Ed.) Ampelography of the USSR. Standard and promising grape varieties. Vol. 2. Moscow: Pishchepromizdat Publ., 1953. (in Russian)

Frolov-Bagrev A.M. (Ed.) Ampelography of the USSR. Standard and promising grape varieties. Vol. 3. Moscow: Pishchepromizdat Publ., 1954. (in Russian)

Goryslavets S.M., Memetova E.Sh., Risovannaya V.I. DNA profiling of grape varieties Mangil al, Shabash and Shabash krupnoyagodnyi and clarification of their genetic relationships using microsatellite loci. *Magarach. Vinogradarstvo i Vinodelie = Magarach. Viticulture and Winemaking.* 2015;3:17-18. (in Russian)

Hammer O., Harper D.A.T., Ryan P.D. PAST: PAleontological STatistics software package for education and data analysis. *Palaeontol. Electronica*. 2001;4(1):1-9.

Ilnitskaya E.T., Ayba V.Sh., Makarkina M.V., Tokmakov S.V., Avidzba M.A. DNA-fingerprinting of local Abkhazian grape variety Kachich. *Sadovodstvo i Vinogradarstvo = Horticulture and Viticulture*. 2019;6:9-13. DOI 10.31676/0235-2591-2019-6-9-13. (in Russian)

Ilnitskaya E.T., Makarkina M.V., Tokmakov S.V., Krasilnikov A.A., Aiba V.Ch., Avidzba M.A. Genotyping of grape plants of "Kachich" variety from different places of growing. *Plodovodstvo i Vinogradarstvo Yuga Rossii = Fruit Growing and Viticulture of South Russia.* 2020;61(1):33-43. DOI 10.30679/2219-5335-2020-1-61-33-43. (in Russian)

Ilnitskaya E.T., Makarkina M.V., Tokmakov S.V., Aiba V.Ch. Identification of the genotype of local Abkhazian grape cultivar Azhshkuakua (Azhizhkvakva). *Plodovodstvo i Vinogradarstvo Yuga Rossii = Fruit Growing and Viticulture of South Russia.* 2021;67(1): 91-101. DOI 10.30679/2219-5335-2021-1-67-91-101. (in Russian)

Mandić A., Žulj Mihaljević M., Leko M., Primorac J., Beljo J. Synonyms and homonyms in Herzegovinian and Dalmatian grapevine cultivars. *Acta Hortic.* 2019;1248:15-20. DOI 10.17660/ActaHortic.2019.1248.3.

Moreno-Sanz P., Suárez B., Loureiro M.D. Identification of synonyms and homonyms in grapevine cultivars (*Vitis vinifera* L.) from Asturias (Spain). *J. Hortic. Sci. Biotechnol.* 2008;83(6):683-688. DOI 10.1080/14620316.2008.11512444.

Negrul A.M. et al. (Eds.) Ampelography of the USSR. Less common grape varieties. Vol. 1. Moscow: Pishchepromizdat Publ., 1963. (in Russian)

Negrul A.M. et al. (Eds.) Ampelography of the USSR. Reference volume. Moscow: Pishchevaya Promyshlennost' Publ., 1970. (in Russian)

Papapetrou M., Loukovitis D., Papadopoulos O., Kazlari Z., Peristeraki A., Arsenova S., Bardarova D., Doncheva D., Theocharis S., Karagiannidis C., Koundouras S., Giannakoula A., Aggelopoulos S., Chatziplis D. Genetic diversity of local Greek and Bulgarian grapevine (*Vitis vinifera* L.) varieties. *Diversity*. 2020;12(7):273. DOI 10.3390/d12070273.

Pastore C., Fontana M., Raimondi S., Ruffa P., Filippetti I., Schneider A. Genetic characterization of grapevine varieties from Emilia-Romagna (northern Italy) discloses unexplored genetic resources. *Am. J. Enol. Vitic*. 2020;71(4):334-343. DOI 10.5344/ajev.2020.19076.

Peakall R., Smouse P.E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*. 2012;28:2537-2539. DOI 10.1093/bioinformatics/bts460.

Raimondi S., Ruffa P., De Lorenzis G., Imazio S., Fiori S., Failla O., Schneider A. Detection of grapevine synonyms in Lombardy and Piedmont regions (northern Italy). *Vitis.* 2015;54:31-36. DOI 10.5073/vitis.2015.54.special-issue.31-36.

Rogers S.O., Bendich A.J. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol. Biol*. 1985;19(1):69-76. DOI 10.1007/BF00020088.

Santiago J.-L., Boso S., Vilanova M., Martínez M.-C. Characterisation of cv. Albarin Blanco (*Vitis vinifera* L.). Synonyms, homonyms and errors of identification associated with this cultivar. *J. Int. Sci. Vigne Vin*. 2005;39(2):57-65. DOI 10.20870/oeno-one.2005.39.2.901.

Schwander F., Eibach R., Fechter I., Hausmann L., Zyprian E., Töpfer R. *Rpv10*: a new locus from the Asian *Vitis* gene pool for pyramiding downy mildew resistance loci in grapevine. *Theor. App. Genet.* 2012;124(1):163-176. DOI 10.1007/s00122-011-1695-4.

This P., Jung A., Boccacci P., Borrego J., Botta R., Costantini L., Crespan M., Dangl G.S., Eisenheld C., Ferreira-Monteiro F., Grando S., Ibáñez J., Lacombe T., Laucou V., Magalhães R., Meredith C.P., Milani N., Peterlunger E., Regner F., Zulini L., Maul E. Development of a standard set of microsatellite references alleles for identification of grape cultivars. *Theor. Appl. Genet.* 2004;109:1048-1058. DOI 10.1007/s00122-004-1760-3.

van Heerden C.J., Burger P., Vermeulen A., Prins R. Detection of downy and powdery mildew resistance QTL in a 'Regent'×'RedGlobe' population. *Euphytica*. 2014;200:281-295. DOI 10.1007/s10681-014-1167-4.

VIVC. *Vitis* International Variety Catalogue. Last modified: 2021-03-14. Julius Kühn-Institut, 2021. Available at: https://www.vivc.de/index.php?r=site%2Findex.

Vokurka A., Maletic E., Benjak A., Karoglan-Kontic J., Pejic I. Application of molecular markers for analysis of presumed synonyms and homonyms within Croatian grapevine cultivars. *Acta Hortic.* 2003;603:581-584. DOI 10.17660/ActaHortic.2003.603.77.

Zendler D., Schneider P., Toepfer R., Zyprian E. Fine mapping of *Ren*3 reveals two loci mediating hypersensitive response against *Erysiphe necator* in grapevine. *Euphytica*. 2017;213(3):68. DOI 10.1007/s10681-017-1857-9.

ORCID ID

E.T. Ilnitskaya orcid.org/0000-0002-2446-0971
M.V. Makarkina orcid.org/0000-0002-3397-0666
I.V. Stepanov orcid.org/0000-0002-6251-300X
I.I. Suprun orcid.org/0000-0003-0355-8395
S.V. Tokmakov orcid.org/0000-0002-2092-7757
V.K. Kotlyar orcid.org/0000-0002-4865-0323

Acknowledgements. This study was supported by Russian Foundation for Basic Research and Academy of Sciences of Abkhazia within the framework of the scientific project No. 19-516-40001 Abkh_a.

Conflict of interest. The authors declare no conflict of interest.

Received March 15, 2021. Revised July 22, 2021. Accepted July 22, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):805-811
DOI 10.18699/VJ21.093

Original Russian text www.bionet.nsc.ru/vogis/

The relationship between the genetic status of the *Vrn-1* locus and the size of the root system in bread wheat (*Triticum aestivum* L.)

O.G. Smirnova, T.A. Pshenichnikova

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
wheatpsh@bionet.nsc.ru

Abstract. One of the main ways to fine-tune the adaptive potential of wheat cultivars is to regulate the timing of flowering using the genes of the *Vrn-1* locus, which determines the type and rate of development. Recently, with the use of introgression and isogenic lines of bread wheat, it was shown that this locus is involved in the genetic control of root length and weight both under irrigation and drought conditions. It turned out that the *VrnA1* gene is associated with a significant decrease in the size of the root system in a winter genotype. The *Vrn-A1* gene had the strongest effect on the reduction of the root system in comparison with the homoeoallelic genes *Vrn-B1* and *Vrn-D1*. The aim of this work was to determine whether the allelic composition of the genes at the *Vrn-1* locus affects the root size in seven spring cultivars and in two lines of bread wheat differing in flowering time under conditions of normal watering and drought. The research was carried out in a hydroponic greenhouse; drought was created at the tillering stage. In this work, we have shown that early flowering wheat cultivars with the dominant *Vrn-A1a* allele have more lightweight and shorter roots under normal watering conditions compared to the late flowering carriers of the dominant homoeoalleles *Vrn-B1* and *Vrn-D1*. In drought conditions, the root length decreased insignificantly, but the weight of the roots significantly decreased in all genotypes, with the exception of Diamant 2. It has been hypothesized that the level of the transcription factor VRN-1 at the onset of drought may affect the size of the root system. The large variability in root weight may indicate the participation, in addition to the *Vrn-1* locus, of other gene networks in the formation of this trait. Breeders working to develop early maturing varieties should consider the possibility of reducing the root size, especially in arid conditions. A significant increase in the root size of line 821 with introgressions into chromosomes 2A, 2B, and 5A from *T. timopheevii* indicates the possibility of using congeners as a source of increasing the trait in wheat.
Key words: bread wheat; root system; drought; *Vrn-1*; flowering dates.

For citation: Smirnova O.G., Pshenichnikova T.A. The relationship between the genetic status of the *Vrn-1* locus and the size of the root system in bread wheat (*Triticum aestivum* L.). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8):805-811. DOI 10.18699/VJ21.093

Взаимосвязь между генетическим статусом локуса *Vrn-1* и размерами корневой системы у мягкой пшеницы (*Triticum aestivum* L.)

О.Г. Смирнова, Т.А. Пшеничникова

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
wheatpsh@bionet.nsc.ru

Аннотация. Одним из главных способов тонкой настройки адаптационного потенциала сортов пшеницы является регулирование сроков цветения с использованием генов локуса *Vrn-1*, определяющего тип и скорость развития. Исследования, связывающие сроки цветения с формированием корневой системы в нормальных и засушливых условиях, единичны. Недавно с использованием интрогрессированных и изогенных линий пшеницы было показано, что локус *Vrn-1* участвует в генетическом контроле длины и массы корней и угла наклона корней в почве как на поливе, так и на засухе. Из трех гомеоаллельных генов – *Vrn-A1*, *Vrn-B1* и *Vrn-D1* – ген *Vrn A1* наиболее сильно уменьшал размеры корневой системы у озимого генотипа. Целью нашей работы было определить, влияет ли аллельный состав генов локуса *Vrn-1* на развитие корневой системы у различающихся по срокам цветения семи яровых сортов и двух линий мягкой пшеницы в условиях нормального полива и засухи. Исследования проведены в условиях гидропонной теплицы, засуха создавалась на стадии кущения. Мы показали, что раннецветущие сорта пшеницы с доминантным аллелем *Vrn-A1a* в нормальных условиях полива имеют корни меньшей массы и длины по сравнению с поздноцветущими носителями доминантных гомеоаллелей *Vrn-B1* и *Vrn-D1*. На засухе длина корней уменьшалась незначительно, а вот масса корней достоверно снижалась у всех генотипов, за исключением сорта Диамант 2. Мы предположили, что уровень транскрипционного фактора VRN-1 на момент наступления засухи может оказывать влияние на размер корневой системы. Большой размах изменчивости по массе корней может свидетельствовать об участии, помимо локуса *Vrn-1*, других генных сетей в фор-

мировании этого признака. Селекционерам, работающим над созданием скороспелых сортов, следует учитывать возможность уменьшения размеров корневой системы, особенно в засушливых условиях. Значительное увеличение массы корней у линии 821 с интрогрессиями в хромосомы 2A, 2B и 5A от вида *T. timopheevii* указывает на возможность использования сородичей в качестве источника увеличения размера корней у пшеницы.
Ключевые слова: мягкая пшеница; корневая система; засуха; *Vrn-1*; сроки цветения.

Introduction

Roots are an integral part of the plant organism. Their development begins at the first stages of ontogenesis. The architecture of the root system determines firm rooting of plants, efficient absorption of nutrients and water from the soil, and interaction with the soil biome. A well-developed shallow root system is able to absorb moisture even from small rains, while a longer root system gains access to moisture accumulated in the deeper layers of the soil. These properties of the root system are especially important in condition of droughts, which currently represent the most serious climate threat worldwide (Ahmad et al., 2017; IPCC, 2018). It was shown that the morphological and functional features of the root system are associated with the preservation of the yield under drought conditions (Comas et al., 2013). The accumulation and practical application of genetic knowledge about the formation of the root system in agricultural crops can lead to the second Green Revolution (Den Herder et al., 2010).

In rice and maize, genes responsible for the formation of the root system have been identified and cloned (Uga et al., 2013; Kitomi et al., 2018). Studies on the genetic control of the root system in bread wheat (*Triticum aestivum* L.) lag significantly behind the studies carried out on rice and maize. So far, in bread wheat, using bi-parental mapping populations and varietal associative mapping panels, QTLs have been localized on the chromosomes of almost all homoeological groups, which indicates a complex genetic control of this trait in wheat (Ehdaie et al., 2016; Lui et al., 2019). Researchers are also looking for genetic diversity for root size in relatives of bread wheat (Feng et al., 2018). It has been shown that the presence of rye introgression in wheat genotypes leads to a significant increase in root weight and an increase in plant productivity under normal and drought conditions (Ehdaie et al., 2003). A similar effect was found in the wheat line Pavon 76 with introgression from *Agropyron elongatum* to chromosome 7DL (Placido et al., 2013).

The dominant alleles of the *Vrn-A1*, *Vrn-B1*, and *Vrn-D1* genes located on chromosomes 5A, 5B, and 5D determine the spring type of wheat development (McIntosh et al., 2013). It was recently found that the *Vrn-1* locus is involved in the genetic control of root length, root weight, and root angle in wheat and barley (Voss-Fels et al., 2018). We were the first to show the relationship between the flowering time and the size of wheat root system under drought conditions (Pshenichnikova et al., 2020). Using monosomic lines of cv. Saratovskaya 29 for chromosomes 5A, 5B, and 5D, we discovered that the *Vrn-A1* gene most strongly affects the decrease in the size of the root system in comparison with other dominant genes of the *Vrn-1* locus. The main task of this work was to determine whether the allelic composition of genes at the *Vrn-1* locus affects the development of the root system in different cultivars and lines of bread wheat, differing in terms of flowering time under conditions of normal watering and drought.

Materials and methods

The studies were carried out on bread wheat accessions that are not related by origin and differ in the allelic composition of the *Vrn-1* genes. The set of genotypes included spring cultivars Saratovskaya 29 (S29), Novosibirskaya 67 (N67), Yanetskis Probat (YP), Diamant 2 (Dm2), Milturum 553 (M553), Duvanka, Chinese cultivar Chinese Spring (CS), line 821 with introgressions from *T. timopheevii* Tausch. into chromosomes 2A, 2B, and 5A of S29 (Leonova et al., 2001) and winter synthetic hexaploid wheat line Synthetic 6x (Syn6x) (AABBDD), obtained from the crossing of *T. dicoccoides* and *Ae. tauschii* and carrying introgression on chromosome 5D.

To equalize the vegetative period in the set of genotypes, two of them were vernalized at a temperature of +2 °C and a 12-hour light regime: winter Syn6x for 60 days, and the late-ripening cultivar CS for 30 days. Plants were grown in a hydroponic greenhouse of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (Center for Shared Use "Laboratory of Artificial Plant Cultivation") under 12–14-hour artificial light 45,000–50,000 lx, night temperature 18–20 °C and day temperature 24–26 °C. Expanded clay with a particle size of 5 to 15 mm was used as an artificial soil. Knop's solution served as a nutrient solution. The plants were grown in two identical bathtubs 500 × 100 × 35 cm in size and about 2 m^3 in volume. The distance between the plants was 12 cm.

Each genotype was grown under two irrigation regimes during three growing seasons. Before tillering stage, all plants were watered equally twice a day. After the beginning of tillering, two watering regimes were created. In the control variant, the previous irrigation regime was maintained until the end of the season. In the experimental variant, watering was stopped. The moisture level was measured in both baths once a week using a moisture meter. Humidity in the control experiment was 28–30 % throughout the season. Under drought conditions, the humidity gradually decreased and within a month was established at the level of 10–12 %. These experimental conditions simulate changes in soil moisture in the field during spring sowing in the sharply continental climate of Siberia. The flowering date was noted for each plant. The stage of grain waxy ripeness was considered the end of the experiment, after which watering in the control variant was stopped. After the soil dried out, the plants were dug up with the roots, and the aboveground part was cut off. After measuring the length of the roots, they were left in air until completely dry and weighed.

A.B. Shcherban performed molecular analysis of the allelic state of the genes at the *Vrn-1* locus in the studied genotypes according to the previously described method (Shcherban et al., 2012).

Phenotypic data obtained during three growing seasons were combined and analyzed by one-way analysis of variance separately for each trait and each irrigation regime. Comparative analysis between groups of genotypes and watering

regimes was performed using Student's *t*-test. The drought tolerance index was measured as a percentage and was calculated as the ratio of the mean value of the trait under drought to the mean value of the trait under irrigation, multiplied by 100. To study the relationship between three traits (days before flowering, root length and root weight), Pearson correlation analysis was performed. Analyzes were performed using the statistical package STATISTICA 6.

Results

Molecular analysis of the allelic composition of the *Vrn-1* locus in nine wheat genotypes showed the presence of the dominant allele *a* at the *Vrn-A1* locus in cultivars S29, YP, N67, Dm2 and line 821 with introgression from *T. timopheevii*, created on the basis of cultivar S29 (Table 1). The recessive allele *vrn-A1* was found in cultivars M553, Duvanka, CS, and Syn6x. The samples also differed in the allelic state of the *Vrn-B1* locus. The dominant allele *Vrn-B1c* was found in cultivars S29, YP, and line 821, the dominant allele *Vrn-B1a* was found in cultivars N67, Dm2, M553, and Duvanka, and the recessive allele *vrn-B1* was found in CS and Syn6x. All studied wheat samples, with the exception of CS, had the *vrn-D1* recessive allele. The CS had a dominant *Vrn-D1a* allele (see Table 1).

As a result of the analysis of the allelic state of the *Vrn-1* genes, the studied samples were divided into two groups. The first group included four cultivars (S29, N67, Dm2, YP) and line 821 – carriers of the dominant allele of the *Vrn-A1a* gene. The second group included four cultivars (M553, Duvanka, CS and Syn6x) – carriers of the *vrn-A1* recessive allele. Analysis of the flowering time showed that plants in the first group flowered 6 days earlier when irrigated and 7 days earlier in drought than plants in the second group ($t = 3.50$; $p < 0.001$). Among the carriers of the *Vrn-A1a* allele, under both irrigation conditions, cultivars S29 and N67 flowered earlier (see Table 1). Among the carriers of the recessive allele *vrn-A1*, the earliest cultivars were CS under irrigation and vernalized Syn6x under drought. The average number of days before flowering significantly increased under drought conditions in all studied genotypes (see Table 1). The tolerance index of this trait did not show a large range in the studied genotypes, with the exception of line 821 and cultivar CS, in which the delay in tillering during drought was the most significant (Table 2).

The root system was studied in all accessions under conditions of normal irrigation and drought. The range of variability in root length was 15 % under irrigation and 50 % under drought, and in root weight – 400 % under irrigation and 500 % under drought (see Table 1).

Line 821 (32 cm), Syn6x (30.2 cm) and cultivar Duvanka (29.7 cm) had the longest roots under irrigation. The roots of the cultivars YP, N67 and CS were several centimeters shorter. S29 and Dm2 formed the shortest roots (about 22 cm). A significant decrease in root length under drought was noted for N67 and Syn6x, and an increase was observed in Dm2. In the rest of the accessions, drought did not have a significant effect on this trait. It should be noted that the cultivars of the second group (carriers of the recessive *vrn-A1* allele) had significantly longer roots both during irrigation and drought compared to the cultivars carrying the *Vrn-A1a* allele. On average, root length slightly decreased during drought in all genotypes. The tolerance index of this trait did not show a large range (see Table 2).

Table 1. Average values of the number of days before flowering, length and weight of roots in bread wheat genotypes differing in the allelic composition of the *Vrn-1* locus under watering and drought

Genotypes	*Vrn-1* alleles	Days before flowering		Length, cm		Weight, g	
		Watering	Drought	Watering	Drought	Watering	Drought
Carriers of the dominant allele *Vrn-A1a*							
S29	*Vrn-A1a*, *Vrn-B1c*, *vrn-D1*	39.7$^{a\#}$	44.0^{a}	22.1^{a}	20.2^{a}	0.32^{a}	0.17^{a}
YP	*Vrn-A1a*, *Vrn-B1c*, *vrn-D1*	43.7^{b}	45.5^{a}	25.8^{b}	24.7^{b}	0.40^{a}	0.26^{a}
N67	*Vrn-A1a*, *Vrn-B1a*, *vrn-D1*	41.3^{a}	43.6^{a}	25.8^{b}	23.2^{a}	0.30^{a}	0.24^{a}
Dm2	*Vrn-A1a*, *Vrn-B1a*, *vrn-D1*	42.6^{b}	46.6^{b}	21.8^{a}	24.3^{b}	0.23^{a}	0.31^{b}
Line 821	*Vrn-A1a*, *Vrn-B1c*, *vrn-D1*	42.5^{b}	51.7^{c}	32.0^{c}	30.6^{c}	0.76^{b}	0.24^{a}
Average		41.9	46.3***	25.5	24.6	0.40	0.27***
Carriers of the recessive allele *vrn-A1*							
M553	*vrn-A1*, *Vrn-B1a*, *vrn-D1*	47.0^{c}	50.6^{c}	28.6^{c}	30.8^{c}	1.18^{c}	0.23^{a}
Duvanka	*vrn-A1*, *Vrn-B1a*, *vrn-D1*	50.8^{d}	51.4^{c}	29.7^{c}	29.5^{c}	0.80^{b}	0.30^{b}
CS	*vrn-A1*, *vrn-B1*, *Vrn-D1a*	45.5^{b}	55.8^{d}	27.2^{b}	24.9^{b}	1.09^{c}	0.68^{c}
Syn6x	*vrn-A1*, *vrn-B1*, *vrn-D1*	47.4^{c}	48.9^{b}	30.2^{c}	24.4^{b}	0.83^{b}	0.72^{c}
Average		47.7***	51.7***	28.9**	27.4**	1.00***	0.48***

$^{\#}$ The mean values followed by different letters in the column are significantly different according to the LSD at $p = 0.05$ within the entire set of genotypes. Differences between the mean values of traits in the groups of carriers of different alleles of the *Vrn-A1* gene are significant at ** $p < 0.01$, *** $p < 0.001$.

Table 2. Drought tolerance indices of wheat genotypes with a different allelic state of the genes at the *Vrn-1* locus

Genotypes	Tolerance indices, %		
	Days before flowering	Root length	Root weight
	Vrn-A1a		
S29	111	91	53
YP	104	95	63
N67	106	90	79
Dm2	109	111	135
Line 821	122	96	32
	vrn-A1		
M553	108	108	19
Duvanka	101	99	37
CS	123	92	62
Syn6x	103	81	87

Much greater variability was shown by root weight, both in absolute values and in the tolerance index to drought. Under normal irrigation, average root weight in the genotypes carrying the dominant *Vrn-A1a* allele was 0.6 g less than in the carriers of the recessive allele. Under these conditions, four wheat cultivars (S29, YP, N67, and Dm2) from the first group did not differ significantly from each other. The values ranged from 0.23 to 0.40 g (see Table 1). At the same time, the root weight in line 821 belonging to the same group turned out to be almost twice as high as in the cultivars mentioned above. In the group of genotypes – carriers of the recessive allele *vrn-A1*, cultivars M553 (*Vrn-B1a*) and CS (*Vrn-D1a*) had the largest root weight under irrigation, more than 1 g. The weight of roots in cultivars Duvanka and Syn6x was lower, about 0.8 g, and approximately the same as in line 821 from the first group.

Cultivar Dm2 differed from the other genotypes by the increase in root weight during drought. Dm2 showed the highest tolerance index of the trait, 135 % (see Table 2). In other genotypes, the weight of the roots decreased during drought. Cultivar M553 had the most significant 5-fold decrease in root weight with tolerance index 19 %. Cultivar Duvanka, which has the same allelic composition of the *Vrn-1* locus as M553, also showed a significant decrease in this trait during drought, tolerance index was 37 %. The same significant decrease in root weight was noted for line 821 from the first group of genotypes. Comparison of tolerance indices among the three studied traits showed that the root weight was the most sensitive to drought in comparison with the root length and the timing of flowering.

Correlation analysis was carried out for three traits for the entire set of genotypes and separately for each of the groups differing in the dominant composition of the *Vrn-A1* gene (Table 3). For the entire population, a correlation was found between the number of days before flowering and the root weight, both during irrigation and during drought. The root length correlated with the number of days before flowering only under drought conditions in the entire studied population and among the carriers of the dominant *Vrn-A1a* allele. The analysis showed that under irrigation conditions there is a correlation between the weight and length of roots for the entire population and for separate groups of genotypes. In drought conditions, this correlation was retained only for the genotypes of the second group with the recessive allele *vrn-A1*.

Discussion

Previously, the involvement of the *Vrn-1* locus in the formation of the size of the root system was established, but the study did not use genetic material with the dominant *Vrn-A1* allele (Voss-Fels et al., 2018). We have shown that the *Vrn-A1*

Table 3. Correlation coefficients between the number of days before flowering, root length and root weight under normal watering and drought among nine wheat genotypes differing in the allelic state of the genes at the *Vrn-1* locus

Traits	Watering		Drought	
	Number of days before flowering	Root length	Number of days before flowering	Root weight
	In the entire population			
Root length	0.12	–	0.31**	–
Root weight	0.36***	0.43***	0.24*	–0.02
	Among the carriers of the dominant allele *Vrn-A1a*			
Root length	–0.12	–	0.49***	–
Root weight	0.23	0.42**	0.06	0.23
	Among the carriers of the recessive allele *vrn-A1*			
Root length	–0.06	–	0.00	–
Root weight	–0.05	0.28*	0.06	0.35*

Note. Differences are significant at * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

gene, located on chromosome 5A, has the strongest effect on root development compared to the *Vrn-B1* and *Vrn-D1* genes (Pshenichnikova et al., 2020). This effect was discovered due to the use of a set of S29 monosomic lines. It should be noted that both of these studies were carried out on experimental genetic material using substitution, isogenic or recombinant wheat lines.

In this work, using seven spring bread wheat cultivars differing in the time of transition to flowering, and the lines with the introgressions from the tetraploid species *T. timopheevii* and *Ae. tauschii*, we estimated the dependence of the development of the root system on the allelic state of the *Vrn-1* locus. We also tried to establish how the length and weight of roots in different cultivars are related to the number of days before flowering under normal and drought conditions. The cultivars were divided into two groups. The first group included carriers of the dominant allele *Vrn-A1a*, the second – carriers of the recessive allele *vrn-A1*. At the same time, eight out of nine studied genotypes carried dominant alleles of other genes of the *Vrn-1* locus.

In this work, we have shown that early flowering cultivars with the dominant *Vrn-A1a* allele had, on average, significantly shorter roots compared to late flowering carriers of the recessive allele. Under irrigation, the difference was 3.4 cm, under drought – 2.8 cm. Cultivars with the recessive allele *vrn-A1* also significantly exceeded cultivars of the first group by root weight (see Table 1). Under irrigation, the difference was 0.6 g, and under drought – 0.24 g. Allelism for the *Vrn-B1* gene did not affect the size of the root system in carriers of the dominant *Vrn-A1a* allele.

Great variability in root weight between groups and within groups was found both under irrigation and under drought. The weight increased due to the intensive formation of secondary roots. Most likely, gene networks not associated with the *Vrn-1* locus control this process. This assumption was made earlier when the size of the root system was studied in lines with introgression from *Ae. tauschii* (Pshenichnikova et al., 2020). It was confirmed in the present study when line 821 – the carrier of the dominant *Vrn-A1a* allele and introgression into chromosomes 2A and 2B – was studied. The length of roots in the line 821 was comparable to the length of roots of the cultivars of the first group (carriers of *Vrn-A1a*). However, in terms of the weight of roots, line 821 was comparable to the accessions from the second group. Earlier, loci associated with root morphology and size were already identified in chromosomes 2A and 2B in bread wheat (Ehdaie et al., 2016; Liu et al., 2019).

In our experiment, drought, which occurred at an early stage of plant development, led to an increase in the number of days before flowering in all genotypes (see Table 1). This effect can be considered as the time spent by the plant on the adaptive restructuring of metabolism. Under drought conditions, among all studied genotypes, a correlation was observed between the number of days before flowering and the length and weight of roots. This may indirectly indicate the participation of the *Vrn-1* locus in the formation of the root system in response to drought. Under drought conditions, no correlation was observed between the weight and length of roots. Since under irrigation conditions the relationship between these characters was significant ($p < 0.001$, see Table 3), the lack of correlation under drought conditions may indicate a mismatch in the genetic pathways for the formation of the root system in unfavorable conditions. The correlation analysis carried out for individual groups showed that only among the carriers of the dominant *Vrn-A1a* allele the number of days before flowering correlated with the length of roots. No correlations were found among the carriers of the *vrn-A1* recessive allele. Perhaps this is due to the fact that by the time of the onset and development of drought (the beginning of tillering), the carriers of the dominant *Vrn-A1* allele accumulate the transcription factor VRN1 in the leaves in a larger amount than in the carriers of the recessive allele *vrn-A1* and the dominant genes *Vrn-B1* and *Vrn-D1* (Lukoianov et al., 2005). This leads to the interaction of VRN1 with the networks of hormonal and signaling responses at earlier stages and the arrest of root growth. In cultivars that carry only the dominant genes *Vrn-B1* and *Vrn-D1*, this response is delayed, and the roots continue to grow in length.

As has already been noted, the effect of drought on two traits, root length and weight, was different. For each genotype, variability in root length under two irrigation regimes was not significant (see Table 1). This is also evidenced by the drought tolerance indices, which, in general, were close to 100 % (see Table 2). The greatest decrease in root length during drought was observed in vernalized Syn6x. Earlier, we showed that the combined effect of vernalization and drought significantly inhibits the growth of the root system and revealed a weak dependence of the root length under the irrigation conditions (Pshenichnikova et al., 2020). In this work, we have shown that in the studied genotypes the length of the roots depends on the allelic state of the *Vrn-1* locus to a greater extent than on the irrigation regime.

The weight of roots, in contrast, showed great diversity of variability under drought. This is reflected in the indicators of the tolerance indices (see Table 2). The greatest decrease in the root weight and low tolerance index (53 %) among the cultivars carrying the dominant allele of the *Vrn-A1* gene was found in the drought-tolerant cultivar S29. Line 821, which has a genetic basis of cultivar S29, showed the greatest decrease (by 3.2 times) in the weight of roots during drought and the lowest tolerance index in the first group – 32 %. These values are comparable to the decrease in weight in the cultivars from the second group, which form a large root system under favorable irrigation conditions. In Duvanka, the weight of roots decreased 2.7 times with a tolerance index of 37 %, and in M553 – 5 times with the lowest tolerance index – 19 %. Cultivar Dm2 turned out to be the only one of the entire population in which weight of roots and their length were increased. Dm2 was characterized by high tolerance index of the root weight. In general, cultivars, carriers of the dominant *Vrn-A1* allele, and vernalized cultivar CS (*Vrn-D1* gene) and winter Syn6x had similar dynamics of root weight reduction under drought. Vernalization induces intensive production of the transcription factor VRN-1 (Trevaskis et al., 2007) and thus equalizes the vegetation status of vernalized accessions and cultivars carrying the dominant allele of the *Vrn-A1* gene. The level of the transcription factor VRN-1 at the onset of drought may be insufficient for the effective functioning of the gene networks for drought resistance in the non-vernalized cultivars M553 and Duvanka, carriers of the recessive allele

of the *Vrn-A1* gene and the dominant allele of the *Vrn-B1* gene. Thus, the dynamics of the decrease in root weight under drought may indirectly indicate the relationship of the *Vrn-1* locus with the gene networks of the response to drought.

The lack of correlation between weight and length of the roots in the entire population under drought indicates the disconnection of the processes of root growth and the accumulation of its biomass through the formation of secondary roots. Each cultivar can use different individual adaptive mechanisms under drought conditions. In particular, cultivar Tincurrin, which has a small root system, has been shown to use soil water more efficiently under drought (Figueroa-Bustos et al., 2020). This was achieved by reducing photosynthetic processes and accelerating grain filling before the onset of serious consequences of water stress. In our experiment, we used cultivar S29, which also has a small root system, but is considered drought-tolerant (Ilyina, 1989). Previously, it was shown that drought resistance of S29 is provided by diverse physiological mechanisms (Osipova et al., 2020).

The mechanisms of the formation of the root system have been studied most fully in the model diploid plant *Arabidopsis thaliana*. It was found that in Arabidopsis, abscisic acid negatively affected the number and elongation of lateral roots during irrigation (De Smet et al., 2006). Cytokinin has an inhibitory effect on branching of lateral roots, while cytokinin biosynthesis mutants exhibit an increased number of lateral roots (Smith, de Smet, 2012). These and other studies have identified individual components of the development of the root system. At the same time, root development was found to be integrated into systemic signaling that, through sugar metabolism, coordinates whole plant growth during flowering induction (Bouché et al., 2016). Experiments on transgenic barley have shown that the transcription factor VRN1 has, in addition to the main binding sites in the promoters of the flowering initiation genes, secondary binding sites. These sites have been found in genes that play a central role in both hormonal responses and hormone metabolism, which include abscisic acid and cytokinins (Deng et al., 2015). Thus, taking part in the regulation of hormonal pathways, VRN1 can influence the formation of the root system under drought conditions.

Conclusion

The allelic composition of the *Vrn-1* locus determines the time required for wheat plants to enter the generative phase of development. Studies linking the timing of flowering with the formation of the root system under normal and arid conditions are sporadic. In this work, we have shown that cultivars with a dominant allele of the *Vrn-A1* gene under normal watering conditions have roots of smaller mass and length compared to carriers of the dominant homoeoalleles *Vrn-B1* and *Vrn-D1*. Drought, which occurs at the tillering stage, led to later flowering of the studied genotypes. At the same time, the length of roots decreased insignificantly, but the weight of roots significantly decreased during drought in all genotypes with the exception of Dm2. The large range of variability in root weight may indicate the participation of additional gene networks in the formation of this trait under drought. Introgressions from *T. timopheevii* and *Ae. tauschii* led to an increase in the size of the root system. This indicates the possibility of using congeners as a source of increasing root size in wheat.

Regulation of the flowering time of cultivars in different growing conditions using the *Vrn-1* locus is considered one of the main ways to fine-tune the adaptive potential. As our work has shown, one should also take into account the possible relationship of this locus with the size of the root system. Breeders working to develop early maturing cultivars may experience a reduction in the size of the root system, especially in arid conditions. One of the ways to maintain the size of the root system in wheat can be the use of introgressions from species-congeners.

References

Ahmad I., Ali N., Ahmad H., Inamullah. Association mapping of root traits for drought tolerance in bread wheat. In: Wanyera R., Owuoche J. (Eds.) Wheat Improvement, Management and Utilization. London: InTech, 2017;39-57. DOI 10.5772/67242.

Bouché F., D'Aloia M., Tocquin P., Lobet G., Detry N., Périlleux C. Integrating roots into a whole plant network of flowering time genes in *Arabidopsis thaliana*. *Sci. Rep.* 2016;6:29042. DOI 10.1038/srep29042.

Comas L.H., Becker S.R., Cruz V.M.V., Byrne P.F., Dierig D.A. Root traits contributing to plant productivity under drought. *Front. Plant Sci*. 2013;4:442. DOI 10.3389/fpls.2013.00442.

De Smet I., Zhang H., Inze D., Beeckman T. A novel role for abscisic acid emerges from underground. *Trends Plant Sci*. 2006;11(9):434-439. DOI 10.1016/j.tplants.2006.07.003.

Den Herder G., Van Isterdael G., Beeckman T., De Smet I. The roots of a new green revolution. *Trends Plant Sci.* 2010;15(11):600-607. DOI 10.1016/j.tplants.2010.08.009.

Deng W., Casao M., Wang P., Sato K., Hayes P.M., Finnegan E.J., Trevaskis B. Direct links between the vernalization response and other key traits of cereal crops. *Nat. Commun.* 2015;6:5882. DOI 10.1038/ncomms6882.

Ehdaie B., Mohammadi S.A., Nouraein M. QTLs for root traits at mid-tillering and for root and shoot traits at maturity in a RIL population of spring bread wheat grown under well-watered conditions. *Euphytica*. 2016;211(1):17-38. DOI 10.1007/s10681-016-1670-x.

Ehdaie B., Whitkus R.W., Waines J.G. Root biomass, water-use efficiency and performance of wheat–rye translocations of chromosomes 1 and 2 in spring bread wheat 'Pavon'. *Crop Sci*. 2003;43(2):710-717. DOI 10.2135/cropsci2003.7100.

Feng K., Cui L., Lv S., Bian J., Wang M., Song W., Nie X. Comprehensive evaluating of wild and cultivated emmer wheat (*Triticum turgidum* L.) genotypes response to salt stress. *Plant Growth Regul*. 2018;84(2):261-273. DOI 10.1007/s10725-017-0337-5.

Figueroa-Bustos V., Palta J.A., Chen Y., Stefanova K., Siddique K.H.M. Wheat cultivars with contrasting root system size responded differently to terminal drought. *Front. Plant Sci*. 2020;11:1285. DOI 10.3389/fpls.2020.01285.

Ilyina L.G. Breeding of Spring Bread Wheat in Southeastern Regions. Saratov: Saratov Univ. Publ., 1989. (in Russian)

IPCC. Special report: Global Warming of 1.5 °C. Summary for Policymakers. 2018. Available at: https://www.ipcc.ch/sr15/ (Date of application Aug. 30, 2021).

Kitomi Y., Nakao E., Kawai S., Kanno N., Ando T., Fukuoka S., Irie K., Uga Y. Fine mapping of *QUICK ROOTING 1* and *2*, quantitative trait loci increasing root length in rice. *G3: Genes Genomes Genetics* (*Bethesda*). 2018;8(2):727-735. DOI 10.1534/g3.117.300147.

Leonova I.N., Kalinina N.P., Budashkina E.B., Röder M.S., Salina E.A. Comparative molecular and genetic analysis of *Triticum aestivum* × *Triticum timopheevii* hybrid lines resistant to leaf rust. In: Proc. 11th EWAC Conf., Novosibirsk, 2001.

Liu P., Jin Y., Liu J., Liu C., Yao H., Luo F., Guo Z., Xia X., He Z. Genome-wide association mapping of root system architecture traits

in common wheat (*Triticum aestivum* L.). *Euphytica*. 2019;215(7): 121. DOI 10.1007/s10681-019-2452-z.

Loukoianov A., Yan L., Blechl A., Sanchez A., Dubcovsky J. Regulation of *VRN-1* vernalization genes in normal and transgenic polyploid wheat. *Plant Physiol*. 2005;138(4):2364-2373. DOI 10.1104/pp.105.064287.

McIntosh R.A., Yamazaki Y., Dubcovsky J., Roger J., Morris C., Appels R., Xia X.C. Catalogue of Gene Symbols for Wheat: 12th Int. Wheat Genetics Symp., 8–13 Sept. 2013, Yokohama, Japan, 2013.

Osipova S., Permyakov A., Permyakova M., Rudikovskaya E., Pomortsev A., Verkhoturov V., Pshenichnikova T. Drought tolerance evaluation of bread wheat (*Triticum aestivum* L.) lines with the substitution of the second homoeological group chromosomes. *Cereal Res. Comm.* 2020;48:267-273. DOI 10.1007/s42976-020-00043-y.

Placido D.F., Campbell M.T., Folsom J.J., Cui X., Kruger G.R., Baenziger P.S., Walia H. Introgression of novel traits from a wild wheat relative improves drought adaptation in wheat. *Plant Physiol*. 2013;161:1806-1819. DOI 10.1104/pp.113.214262.

Pshenichnikova T.A., Smirnova O.G., Simonov A.V., Shchukina L.V., Morozova E.V., Lohwasser U., Börner A. The relationship between root system development and vernalization under contrasting irrigation in bread wheat lines with the introgressions from a synthetic hexaploid. *Plant Growth Regul.* 2020;92(3):583-595. DOI 10.1007/s10725-020-00666-5.

Shcherban A.B., Emtseva M.V., Efremova T.T. Molecular genetical characterization of vernalization genes *Vrn-A1*, *Vrn-B1* and *Vrn-D1* in spring wheat germplasm from Russia and adjacent regions. *Cereal Res. Comm.* 2012;40(3):351-361. DOI 10.1556/CRC.40.2012.3.4.

Smith S., De Smet I. Root system architecture: insights from *Arabidopsis* and cereal crops. *Philos. Trans. R. Soc. Lond. B. Biol Sci*. 2012;367(1595):1441-1452. DOI 10.1098/rstb.2011.0234.

Trevaskis B., Hemming M.N., Dennis E.S., Peacock W.J. The molecular basis of vernalization-induced flowering in cereals. *Trends Plant Sci.* 2007;12(8):352-357. DOI 10.1016/j.tplants.2007.06.010.

Uga Y., Sugimoto K., Ogawa S., Rane J., Ishitani M., Hara N., Kitomi Y., Inukai Y., Ono K., Kanno N., Inoue H., Takehisa H., Motoyama R., Nagamura Y., Wu J., Matsumoto T., Takai T., Okuno K., Yano M. Control of root system architecture by *DEEPER ROOTING 1* increases rice yield under drought conditions. *Nat. Genet.* 2013;45:1097-1102. DOI 10.1038/ng.2725.

Voss-Fels K.P., Robinson H., Mudge S.R., Richard C., Newman S., Wittkop B., Stahl A., Friedt W., Frisch M., Gabur I., Miller-Cooper A., Campbell B.C., Kelly A., Fox G., Christopher J., Christopher M., Chenu K., Franckowiak J., Mace E.S., Borrell A.K., Eagles H., Jordan D.R., Botella J.R., Hammer G., Godwin I.D., Trevaskis B., Snowdon R.J., Hickey L.T. *VERNALIZATION1* modulates root system architecture in wheat and barley. *Mol. Plant*. 2018;11(1):226-229. DOI 10.1016/j.molp.2017.10.005.

ORCID ID

O.G. Smirnova orcid.org/0000-0002-3023-767X
T.A. Pshenichnikova orcid.org/0000-0001-5639-916X

Acknowledgements. This work was supported by the Russian Foundation for Basic Research grant No. 19-416-540001 and contract No. p-94 (June 21, 2019) with the Administration of the Novosibirsk region. Wheat plants were grown in the Center for Shared Use "Laboratory of Artificial Plant Cultivation" of the Institute of Cytology and Genetics of the SB RAS with the support of the budget project No. 0259-2021-0012. We are grateful to Andrey B. Shcherban for determining the allelic state of the *Vrn-1* locus.

Conflict of interest. The authors declare no conflict of interest.

Received August 31, 2021. Revised October 11, 2021. Accepted October 12, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):812-821
DOI 10.18699/VJ21.094

Original Russian text www.bionet.nsc.ru/vogis/

Haploid biotechnology as a tool for creating a selection material for sugar beets

E.O. Kolesnikova ✉, E.I. Donskikh, R.V. Berdnikov

Breeding and Genetic Center "UnionSeedsBeet", Ltd., VNIISS, Ramonsky district, Voronezh region, Russia
✉ kolelkbn@mail.ru

Abstract. Since the discovery of the phenomenon of haploidy, biotechnology has become an integral part in the successful creation of new varieties and hybrids of various plant species. In particular, these technologies are actively used in agriculture, which is concerned with increasing the volume and improving the quality of products. The integration of haploid production techniques together with other available biotechnological tools such as marker selection (MAS), induced mutagenesis and genetic engineering technologies can significantly accelerate crop breeding. This article shows the main stages in the development of biotechnology since 1921. Now they are successfully used to create doubled haploids to accelerate the selection process of various plants and, in particular, sugar beet, which is the most important sugar crop in regions with a temperate climate. There are several methods for obtaining forms with a single set of chromosomes. For sugar beets, the use of gynogenesis turned out to be expedient, since in this case the other methods turned out to be ineffective in the mass production of haploids. The article considers the stages of obtaining the H and DH lines of *Beta vulgaris* L., as well as the main stages of biotechnological production of homozygous breeding material of this culture. These stages include selecting parental forms – donor explants, sterilizing buds and introducing non-pollinated ovules *in vitro*, obtaining haploids, doubling their chromosome set, creating doubled haploids, determining ploidy at different stages, relocating the obtained plants to greenhouses and growing stecklings. A number of advantages that the technology of creating doubled haploids *in vitro* has in comparison with traditional methods of selection are described. It has been shown that the use of these approaches is relevant when obtaining new highly productive hybrids and varieties of agricultural plants; however, the methods for the production of homozygous forms in sugar beet still require additional research aimed at increasing the efficiency and reproducibility of each stage of the process.

Key words: sugar beet; haploid; doubled haploid; gynogenesis; biotechnology; *in vitro*; DH lines.

For citation: Kolesnikova E.O., Donskikh E.I., Berdnikov R.V. Haploid biotechnology as a tool for creating a selection material for sugar beets. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021; 25(8):812-821. DOI 10.18699/VJ21.094

Биотехнологии гаплоидов как инструмент создания селекционного материала сахарной свеклы

Е.О. Колесникова ✉, Е.И. Донских, Р.В. Бердников

Селекционно-генетический центр ООО «СоюзСемСвекла», пос. ВНИИСС, Рамонский район, Воронежская область, Россия
✉ kolelkbn@mail.ru

Аннотация. С момента открытия явления гаплоидии биотехнологии стали неотъемлемой частью в процессе успешного создания новых сортов и гибридов различных видов растений. Особенно активно данные технологии применяются в сельском хозяйстве, которое заинтересовано в увеличении объемов и повышении качества производимой продукции. Интеграция приемов получения гаплоидов вместе с другими имеющимися биотехнологическими инструментами, такими как маркерная селекция (MAS), индуцированный мутагенез и генно-инженерные технологии, может значительно ускорить селекцию сельскохозяйственных культур. В статье показаны основные этапы развития биотехнологий начиная с 1921 г. Теперь они успешно используются при создании удвоенных гаплоидов для ускорения селекционного процесса различных растений, и в частности сахарной свеклы – важнейшей сахароносной культуры в регионах с умеренным климатом. Существует несколько методов получения форм с одинарным набором хромосом. Для сахарной свеклы целесообразным оказалось применение гиногенеза, поскольку остальные приемы были малоэффективны при массовом получении гаплоидов. В публикации рассматриваются этапы получения H- и DH-линий *Beta vulgaris* L., а также основные этапы биотехнологического производства гомозиготного селекционного материала этой культуры. К ним относятся: отбор родительских форм – доноров эксплантов; стерилизация бутонов и введение неопыленных семяпочек *in vitro*; получение гаплоидов;

удвоение их хромосомного набора; создание удвоенных гаплоидов; определение плоидности на разных этапах; перевод полученных растений в закрытый грунт и выращивание штеклингов. Описан ряд преимуществ, которые имеет технология создания удвоенных гаплоидов *in vitro* по сравнению с традиционными методами селекции. Показано, что применение данных подходов является актуальным при получении новых высокопродуктивных гибридов и сортов сельскохозяйственных растений, однако приемы производства гомозиготных форм у сахарной свеклы все еще требуют проведения дополнительных исследований, направленных на увеличение эффективности и воспроизводимости каждого этапа процесса.
Ключевые слова: сахарная свекла; гаплоид; удвоенный гаплоид; гиногенез; биотехнологии; *in vitro*; DH-линии.

Introduction

Due to the increase in the consumption of agricultural products, an urgent need arose for the development of technologies that accelerate the selection processes of cultivated crops. Sugar beet, which is one of the main industrial crops, is not an exception since sugar consumption is growing every year. *Beta vulgaris* L. accounts for a significant part of the world's total sugar production and about 80 % of the obtained beet sugar comes from European countries. At the same time, Russia is the leader in terms of volume in sugar beet cultivation.

The main task of breeding this crop under conditions of intensification of agricultural production is the creation of highly productive domestic hybrids on a linear basis. The process of obtaining promising hybrids is carried out more efficiently when the biotechnological production of new breeding material is introduced. The great economic importance of *Beta vulgaris* determines considerable interest in optimizing the processes of micropropagation *in vitro*, creating haploid forms and obtaining homozygous forms on their basis for breeding work. Modern agro-industry should have specialists who are skilled at using biotechnological methods in order to intensify agricultural production and improve its quality.

Homozygous lines are a unique genetic material for accelerating the process of creating new hybrids and reducing its labor intensity, for mapping the populations, for use in functional genomics and molecular selection. All genes of haploid plants are represented by a single allele; due to this fact, all unfavorable recessive signs can be detected at early stages of the selection process. Homozygous lines based on haploids can be obtained within two years whereas classical methods for breeding for heterosis in cross-pollinated crops make it possible to achieve homozygosity only after 6–7 years of inbreeding. Since *Beta vulgaris* has a 2-year development cycle, in this case the whole process takes on average 12 years. Besides, this culture is characterized by self-compatibility and by the occurrence of inbred depression (Urazaliyev et al., 2013). Nowadays, the production of doubled haploids has become a tool in the selection programs of world research laboratories as an alternative to the classical method of obtaining homozygous lines. Due to the growing understanding of the importance of using such an approach in selection, interest in research in this field is constantly increasing (Datta, 2005). A significant number of new haploid-derived plant varieties are recorded worldwide every year. The EU, Canada, Australia, the USA and China have been leaders in the field of haploid technologies so far (Dunwell, 2010).

Haploidy

The phenomenon of haploidy has been the object of scientific attention since the beginning of the 20th century and is currently widely known in many angiosperms. The development of experimental haploidy methods began a little later, when the potential of using plants with a single set of chromosomes in creating pure lines for breeding needs was discovered. There are several ways to form haploids:

1. Plants are pollinated by pollen of plants of the same species (inducers of haploids), which are classified as paternal or maternal inducers based on the genetic constitution of the obtained haploids. Respectively, paternal and maternal haploids carry the genome from male and female parents. At the same time, chromosomes-inducers are eliminated in haploid embryonic cells within the first week after pollination (Chaikam et al., 2019).
2. Plants are pollinated by pollen of an unallied species, for example, crossing of wheat with corn. This method is very effective for the production of haploids in most *Triticum* spp. genotypes, including difficult-to-respond forms in *in vitro* anther culture. The data obtained by scientists showed that the yield of haploid embryos in individual crosses reached up to 53 % (Djatchouk et al., 2019).
3. Plants are pollinated by pollen of a wild allied species. This method is used in barley selection to obtain haploids when crossing *Hordeum vulgare* × *H. bulbosum* (the so-called "bulbosum" method). Elimination of *H. bulbosum* chromosomes occurs during mitosis and during interphase and is accompanied by the formation of micronuclei and heterochromatinization. Complete elimination of bulb barley chromatin occurs within 5–9 days after pollination. The use of embryo rescue technology (embryo rescue) provides an increase in the efficiency of the method at the next stage and the possibility of its use in selection (Sahijram, Rao, 2015).
4. Pollination by irradiated pollen is a well-documented method of induction of cucumber haploids, which can be obtained from various material such as selection lines, hybrids and varieties. Haploid plants are genetically stable, but it is necessary to double the number of chromosomes in them before further use in selection as it is a very important step. The technology for obtaining doubled

haploids of cucumber using pollination with irradiated pollen is considered to be the most developed, convenient, providing a stable yield of DH lines, compared with technologies for obtaining in anthers culture or microspores and non-pollinated ovules *in vitro* (Gałązka et al., 2015).

5. Several protocols for CENH3-mediated induction of haploids are currently known. However, an effective haploid technology based on CENH3 is not available for crops yet, with the exception of maize, where the haploid level reached 3.6 % (Kelliher et al., 2016). This goal is actively studied now. The expectations for the application of this approach to crops are high, while the creation of viable methods of haploidization based on CENH3 is a difficult task and there is no mass introduction into practice yet (Watts et al., 2016).
6. Obtaining haploids by androgenesis (that means obtaining haploid plants on an artificial nutrient medium from isolated anthers and microspores) and gynogenesis (that means obtaining haploid plants on a nutrient medium from isolated ovules).

During androgenesis, the male gametophyte, under the influence of *in vitro* inducing factors, switches from the gametophytic path of development to the sporophytic pathway with the formation of an embryoid or androgenic callus, from which regeneration of haploids or doubled haploids is possible. Almost all biotechnological laboratories of breeding and genetic companies in Europe and the USA use anther culture (Touraev et al., 2009; Basu et al., 2010).

When cultivating anthers, callus formation is more common. As a result of further morphogenesis, plants are regenerated from callus cells. The cases of direct embryogenesis, when pro-embryonic structures are formed from immature pollen grains, developing into embryoids that give a rise to haploid plants, are quite rare.

A more promising method for obtaining doubled haploids is the culture of isolated microspores, which can be isolated in large quantities, providing potentially embryogenic single haploid cells. This technology is quite simple to implement, cost-effective and provides a high yield of DH plants while optimizing the technology for a specific species and genotype. The absence of somatic tissues in the culture of microspores *in vitro* makes it possible not to question the origin of the obtained plants (Domblides et al., 2019).

Despite the success achieved in the development of androgenetic methods and the creation of a number of varieties of the most important types of grain crops on their basis, their effective use is hindered by a number of reasons, the main of which is the reproducibility of the results obtained in different seasons for different genotypes in combination with a decrease in the cost of obtaining. A rather high yield of chlorophyll-free seedlings, which affects breeding programs by reducing the frequency of regeneration of green seedlings, is also a problem (Kasha et al., 2001).

In vitro induction of maternal haploids – gynogenesis is used mainly in plants for which androgenesis and pollination induction are ineffective. In contrast to induced parthenogenesis, in gynogenesis prior to *in vitro* pollination, ovaries or isolated ovules, rather than defective embryos from seeds, are introduced. When cultured on nutrient media, the haploid cells of the embryo sac form embryoids (direct embryogenesis) or morphogenic callus, from which a plant is formed (indirect embryogenesis). The process of induction of gynogenesis is also influenced by a large number of factors: genotype, growing conditions of the donor plant, stage of gametophyte development, composition of the nutrient medium, and stress. In male-sterile plants, the cultivation of non-pollinated ovules is the only way to obtain haploids. In some plants, such as barley and rice, the induction of green plants is much higher during gynogenesis than in androgenesis. The frequency of the accidental appearance of plants with a single set of chromosomes is insignificant, up to 0.01 %, that is, it is a rare event and has limited practical value (Bohanec, 2009; Kiełkowska et al., 2014).

History of the development of haploid plant technologies

The first plant identified as a haploid, according to one version, was discovered by A.D. Bergner in 1921. The use of such plants in breeding was suggested by A.F. Blakeslee and J. Belling. The scientists obtained the plants with a single set of chromosomes while trying to mutate *Datura stramonium* L. by applying cold as a stimulus. The obtained haploid immediately became an innovation among such flowering plants as a sporophyte having a set of chromosomes characteristic of the gametophyte (Blakeslee et al., 1922). Later, the haploid form was identified in the F_1 progeny when crossing the species *Nicotiana tabacum* and *N. sylvestris*. The obtained plant had some morphological differences from the parental forms, such as long narrow leaves, smaller flowers, sterile pollen, and inability to form mature seeds, which was confirmed by cytological examination (Clausen, Mann, 1924). According to E.F. Gains, a haploid wheat plant that had 21 chromosomes instead of 42 chromosomes as in the parental individuals was discovered in 1925. The haploid form was practically indistinguishable from the diploid form, but it had greater tillering. Distinct differences became noticeable at the time of flowering and during the formation of immature seeds (Gains, Aase, 1926). Later, in the course of research, the induction of haploids in wheat was achieved by pollination of plants with pollen that underwent X-ray irradiation. As a result, male gametes were inactivated and lost the ability to merge with the egg, but stimulated its division and development of the embryo (Katayama, 1934). Studies using X-ray irradiation did not give significant quantitative results and posed a danger to humans.

At the initial stages, haploid forms of plants were obtained by traditional selection methods using distant hybridization. Thus, when *Triticum aestivum* L. and *Secale cereale* L. were crossed, two haploids were obtained (Sears, 1988). Later, other methods using various chemicals appeared. With the development of biotechnological techniques, the creation of haploids has become possible for many plant crops.

The haploid form *Beta vulgaris* L. was first identified by a Swedish scientist A. Levan in the greenhouses of the Swedish Sugar Co Hilleshog breeding station, where in 1942 the scientist treated sugar beet plants with colchicine (Levan, 1945). The collected seeds were germinated and in 1943 the number of chromosomes in the obtained plants was examined. In addition to diploid, triploid and tetraploid plants, one haploid plant was obtained with the number of chromosomes equal to 9, in contrast to the parental forms with 18 chromosomes. A. Levan suggested that colchicine treatment caused damage to gametes: one pollen grain could stimulate the development of the embryo, but was not capable of fertilization. In his study, the author paid special attention to the morphology and cytology of the haploid plant. According to his description, the haploid possessed a large number of narrow leaves, which were smaller in size compared to diploid, triploid and tetraploid plants. According to external signs, the haploid was clearly weaker and lower than diploid plants, but it could form well-developed inflorescences and fertile pollen, which eventually underwent degradation. On the basis of cytogenetic studies, A. Levan concluded that in haploids meiosis was as close as possible to that in diploids, but due to the absence of homologous chromosome pairs the whole mechanism ended in failure. Further research in this field made it possible to obtain experimentally haploid forms of plants with a frequency exceeding the natural level.

In 1964, employees of the Department of Botany of the University of Delhi (India) S. Guha and S.C. Maheswari published data on the results of biochemical studies of meiosis of *Datura* anthers in culture *in vitro* (Guha, Maheswari, 1964, 1966). When cultivating mature anthers on nutrient media, scientists discovered embryoids developing from immature microspores. Some of the embryoids that regenerated during the experiment turned into normal seedlings. In further studies, it was found that some of them had a haploid number of chromosomes. Later, the method by which hundreds of haploid plants of various types of tobacco were obtained from pollen grains *in vitro* was presented. When grown on a nutrient medium, some pollen grains grew into embryonic structures, which, gradually developing, were capable of abundant flowering, but did not form seeds (Nitsch J., Nitsch C., 1969). It also became known that rice haploids (*Oryza sativa* L.) (Niizeki, Oono, 1968), wheat haploids (*Triticum aestivum* L.) obtained in anther culture (Ouyang et al., 1973) had been successfully produced *in vitro.*

In 1982, the culture of isolated microspores became known, which was more effective in the production of haploids (Lichter, 1982). Later, haploids of *Triticum aestivum* L. were produced in the culture of isolated microspores (Datta, Wenzel, 1987; Tuvesson, Öhlund, 1993), by distant hybridization with wild barley (*Hordeum bulbosum* L.) and corn (*Zea mays* L.) (Barclay, 1975; Laurie, Bennett, 1986; Inagaki, Tahir, 1990).

The extraction of anthers from buds and their subsequent opening to release microspores was a rather laborious procedure that was improved by M. Zheng in the course of obtaining haploids and doubled haploids from wheat microspores. The developed technology included the stages of homogenization, filtration, and centrifugation of the obtained sample in a density gradient. As a result, M. Zheng was able to collect a fraction of viable embryogenic microspores for culturing on nutrient media under *in vitro* conditions, which allowed to optimize the method of wheat microspore extraction (Zheng, 2003). Further, the work of L. Cistué and Z. Labbani with co-authors on the DH protocol for durum wheat was published (Cistué et al., 2006; Labbani et al., 2007). The main improvements that they applied in their work were pretreatment with mannitol and the use of colchicine *in vitro*.

So, with the acquisition of more and more data on the possibility of creating haploid forms of higher plants *in vitro*, the value of their use in breeding and the importance of the development of biotechnological methods were revealed. By now, for many plant crops, such as wheat, triticale, barley, rice, corn, cabbage, carrots, etc., effective techniques have been developed that make it possible to obtain haploid plants to create clean lines.

Development of sugar beet haploid biotechnologies

In 1971, N. Bosemark reported the production of five haploids by pollinating plants with wild beet pollen and irradiated sugar beet pollen. In addition to the emergence of haploid forms, it was possible to create homozygous diploid and tetraploid lines after treating previously germinated seeds with colchicine (Bosemark, 1971). In 1983, it became known that haploids could be obtained by the method of distant hybridization. When sterile sugar beet plants were pollinated with red table beet pollen, the incidence of haploids was 0.013 % (Seman, 1983). These methods were aimed at stimulating an unfertilized egg to develop, however, they showed insignificant results in the number of haploids obtained, which was insufficient for large-scale breeding work.

The androgenesis pathway for producing sugar beet haploids has generally been found to be ineffective. Anthers most often induced callus, proembryogenic structures, and roots on the mineral media used; however, their percentage depended on the combination of growth substances used. According to the results of the study by J. Rogozinska and M. Goska, the Linsmeier and Skoog medium with the addition of zeatin 6-(4-hydroxy-3-methyl-trans-2-butenylamino) purine) or zeatin and NAA (1-naphthaleneacetic acid) was recognized as the best medium for differentiation acid, and the addition of PFP (*p*-fluorophenylalanine) increased the percentage of anther differentiation (Rogozinska, Goska, 1982). In addition to callus and roots, vegetative buds were formed on one anther out of approximately 140,000 tested, from which numerous diploid plants were obtained. Cytological analyzes showed the formation of multicellular structures, which subsequently degenerated. In other similar experiments, whole plants developed, but their tissue was

more often diploid, and the gametophytic origin of the regenerants was not confirmed (Gürel E. et al., 2008). In recent studies, conditions have been developed for the direct induced androgenesis of sugar beet in *in vitro* culture. The process of obtaining haploids included cold pretreatment of explants, which was a necessary factor in initiating the transition of microspores from the gametophytic to the sporophytic pathway. For the cultivation of anthers and developed embryoids, the composition of the medium was modified, which included 2,4-D and 6-BAP. As a result of the experiment, from 0.15 to 1.32 % of microclones of androgenic origin were obtained (Hontarenko, Herasymenko, 2018). However, it was not effective enough for mass production of haploid forms. A lot of studies in this field gave unsatisfactory results. The morphogenetic response of *in vitro* cultivation of elements of the male generative system of *B. vulgaris*, according to data available today, is considered very low.

D. Hosemans and D. Bossoutrot were the first to succeed in obtaining sugar beet haploids by cultivating non-pollinated ovules. In their experiment, they identified 0.17 % of haploids formed (Hosemans, Bossoutrot, 1983; Bossoutrot, Hosemans, 1985). Their further histological study showed that the regenerated embryoids could have originated from an unfertilized egg or antipode. However, the resulting gynogenetic plants showed the phenomenon of endopolyploidy at the level of the root meristem, while the shoot meristem remained haploid. From the moment when it became clear that this approach may be the only effective method for producing haploid sugar beet plants, numerous *in vitro* studies have begun to optimize this process.

Interest in sugar beet gynogenesis increased every year, and in 1987 J. Van Geyt and his co-authors reported the production of haploids from ovules with a frequency of up to 6.1 % (Van Geyt et al., 1987). The results of histological studies in the experiment confirmed that the obtained plants originated from haploid cells of the embryo sac, but spontaneous polyploidization was observed at the root tips, as in the study of D. Hosemans and D. Bossoutrot (Hosemans, Bossoutrot, 1983; Bossoutrot, Hosemans, 1985). According to J. Van Geyt, the shape of the extracted ovules was of great importance when introduced into tissue culture. It was found that the loss of the shape of the comma by the explant was accompanied by the death of the egg. It was also reported that plant regeneration was inhibited by the formation of callus from the mother's tissue, but after removing it and transferring the ovule to a new nutrient medium containing charcoal, its reappearance could be suppressed. Further studies of gynogenesis revealed the dependence of this process on various conditions.

In the course of M. Doctrinal's studies, it was found that factors such as the nature and concentration of hormones used, cultivation temperature, seasonal effects, and genotype had a great influence on the development of haploid plants (Doctrinal et al., 1989). According to the results of observations, for the initiation of embryoidogenesis in non-pollinated ovules of sugar beet, the most preferable temperature was 27 °C, while the seasonal effect had a significant impact. The most active regeneration was observed in July. It was also found that the hormonal composition of nutrient media influenced the pathways of female gametophyte morphogenesis and the formation of morphological structures. The best qualitative and quantitative indicators of the gynogenic response in the studied genotypes were observed on nutrient media containing 2.85 μM 3-indoleacetic acid and 0.88 μM 6-benzylaminopurine, as well as on a medium containing 2.3 μM kinetin. Depending on the genotype, when using these media, from 6 to 10 % of viable plants were obtained, 81 % of which turned out to be haploids. Research by M. Doctrinal confirmed the promise of obtaining sugar beet haploids by gynogenesis. The development of biotechnological methods continued, taking into account many factors affecting the gynogenesis of *Beta vulgaris* L. *in vitro*, aimed at optimizing the cultivation conditions and increasing the efficiency of the corresponding methods.

In the works of H. Lux and co-authors, it was noted that the yield of regenerants from the ovules decreased from the most active regeneration in September to the lowest in January (Lux et al., 1990). These data suggested that the effectiveness of gynogenesis was seasonal. According to the authors, despite the laboriousness of the process, gynogenesis turned out to be a more suitable method for obtaining sugar beet haploids *in vitro*, as the species was not amenable to androgenesis. Depending on the genotype, from 0 to 13 % of haploid plants were obtained, while in 10 % of plants during cultivation and reproduction, spontaneous doubling of chromosomes was observed while 90 % remained haploids.

For the successful application of gynogenesis in practical problems, the number of formed haploid plants is of great importance. S. Gürel and co-authors confirmed that cold pretreatment and the action of activated charcoal can increase the incidence of embryo formation (Gürel S. et al., 2000). In the experiments of scientists, embryos that developed from an egg formed shoots with a haploid number of chromosomes. However, when developing optimal conditions for obtaining haploids, the problem of genotypic dependence of the response to cultivation conditions arose. The yield of the obtained embryoids differed in sugar beet lines, as well as the response of developed microclones to different growth conditions. Genotypic differences in the response to cultivation conditions are a serious problem not only in the culture of ovules (Hansen et al., 1995), but also in the work with other tissues of sugar beet (Mikami et al., 1989; Gürel E., 1997); therefore, it is recommended to select the composition of the nutrient medium, as well as the cultivation conditions, for each genotype individually.

Studies conducted by many scientists have indicated the importance of pre-cold treatment of the material for increasing the rate of gynogenesis of *Beta vulgaris* (Lux et al., 1990; Gürel S. et al., 2000; Svirshchevskaya, Dolezel, 2000; Pazuki et al., 2018a). H. Lux and co-authors showed that cold treatment (4 °C) for 4–5 days was able to significantly

increase the rate of sugar beet gynogenesis. An increase in the period of cold exposure to inflorescences over one week reduced the gynogenic response of ovules, while preliminary cold treatment for 7 days and the use of BAP at a concentration of 1 mg/L had a stimulating effect when switching the development of explants from the gametophytic to the sporophytic pathway. It was emphasized that the influence of the season and genotype on the rate of regeneration was significant. In addition to cold pretreatment, the percentage of regeneration also increased when the spectrum of light irradiation, used in the cultivation of ovules in a thermal room, was shifted towards the red region of the spectrum (D'Halluin, Keimer, 1986).

Researchers E. Weich and M. Levall (2003) proposed a protocol for obtaining doubled sugar beet haploids, in which all stages of creating breeding material were considered: growing donor plants, collecting inflorescences, surface sterilization of the material, isolating ovules, cultivating haploid embryoids, reproduction of haploids, obtaining doubled haploids, their rooting, transfer to greenhouse conditions, and acclimatization. The conditions for each stage were described, recommendations were given for manipulating plant material, the compositions of the optimal nutrient media were described. This protocol was of a recommendatory nature; due to the diversity of the genotypic response to cultivation conditions, it should be modified for different genotypes of sugar beet.

To obtain haploids efficiently, M. Tomaszewska-Sowa described a two-stage cultivation process, in which the explants were non-pollinated ovules isolated after sterilization from the buds of the generative shoots of sugar beet plants. The reproductive structures were kept in a liquid nutrient medium for 12 weeks (Tomaszewska-Sowa et al., 2017). The regenerated explants were transferred to a solid nutrient medium with a modified composition, after which the formation of shoots was observed. It was found that the organogenesis of the ovules in the two-phase method was not direct, but proceeded through the formation of callus tissue. The efficiency of regeneration depended on the type and origin of the explant. Differentiation processes in somatic embryoids were enhanced by the presence of 6-BAP and 2,4-D in the medium, which, in turn, increased the number of formed specialized tissue structures.

The ratio of hormones in the substrate has a major influence on the pathways of development of the embryonic structure. At present, the protocols for obtaining sugar beet haploids require further development and improvement, therefore, studies on the selection of the optimal composition of media and the search for other factors stimulating gynogenesis are still ongoing.

Stages of creating homozygous material *in vitro*

Selection of parental forms

So, the selection of donor plants can be considered the first stage with which the work on the introduction of plant material *in vitro* begins. To reduce the level of infection of explants, the collection of inflorescences should be carried out in dry weather, with a prolonged absence of precipitation. This procedure is best done at the beginning of the flowering period. Depending on the region of growth, it can be May–July. The collected inflorescences of each genotype are placed in plastic bags, labeled and stored in the cold until the stage of sterilization of plant material under laboratory conditions.

Sterilization of material with the introduction of ovules *in vitro*

Plants are easily affected by various epiphytic and endophytic microorganisms and viruses. That is why the stage that determines the success of the *in vitro* material introduction process is the high-quality sterilization of the starting material. Previously developed techniques for sterilizing explants using a number of mercury-containing preparations (mercury chloride, diocide, merthiolate) were found to be effective (Granda, 2009), but very toxic to both humans and plants. Over time, they were supplanted by techniques using other substances. For surface sterilization, plant tissues can be treated with chlorine-containing substances (calcium or sodium hypochlorite, bleach, chloramine), hydrogen peroxide, ethanol. O. Jones suggested using solutions containing sodium hypochlorite (Domestos) as a disinfectant for sterilization (Jones et al., 1979). E. Weich and M. Levall used the commercial product Klorin or 3 % sodium hypochlorite for sterilization. After sterilization, the material was thoroughly washed with distilled water and stored in a refrigerator at 8±2 °C (Weich, Levall, 2003). When choosing a sterilization method, it is necessary to take into account both its effectiveness against bacterial and fungal infections, and the prevention of damage to plant tissues. The use of new sterilizing agents increases the likelihood of an effective explant sterilization process.

Introduction of ovules into culture *in vitro*

By now, despite the emergence of high-tech devices and devices that facilitate some manipulations in laboratories, the creation of haploid material depends on the delicate manual work of the operator. The process of extraction of sugar beet ovules when introduced into *in vitro* culture has been described in detail by A. Pazuki et al. (Pazuki et al., 2018b). The scientists performed this process under sterile conditions under a stereomicroscope using tweezers and a scalpel. The first closed and subsequent buds were opened towards the top of the inflorescence, introduced into Petri dishes containing the nutrient medium. The cultivation was carried out at a temperature of 27±2 °C with an 18-hour photoperiod. Regenerated ovules were transferred to MS proliferation and propagation medium containing 0.2 mg/L of kinetin, 10 g/L of sucrose. Due to the different reaction of genotypes, the composition of the nutrient media used and the cultivation conditions may be different for each genotype at all stages of creating the breeding material. The regeneration of the

ovules can begin as early as 2–3 weeks after injection. The developed full-fledged microclones are further maintained in *in vitro* culture.

Determination of ploidy of the obtained regenerant

Phenotypically, haploid plants differ from diploid plants in height, size and number of organs and have narrower leaf plates; however, visual determination of ploidy does not give accurate results. After the formation of normally developed regenerants from non-pollinated ovules, their ploidy is checked, since there is a possibility of obtaining seedlings from their somatic cells that have a diploid or mixoploid chromosome set.

One of the reliable methods for determining ploidy is the counting of chromosomes which are at the stage of mitosis of actively growing tissues (growth zones of the root or young leaves). The method is laborious and requires lengthy preparation. The process of counting chromosomes in sugar beet cells is described in detail by A. Pazuki and coauthors (Pazuki et al., 2018a). Young leaves of seedlings were treated *in vitro* for 3 hours with a solution of 8-hydroxyquinoline (0.002 M), followed by their fixation in a freshly prepared solution of 96 % ethanol and hydrochloric acid (2:1). Then they were washed and stored in distilled water. Then a small piece of leaf tissue was transferred into a drop of 3 % orsein in 45 % acetic acid onto a glass slide and crushed under a cover glass. Then the chromosomes were counted under a light microscope.

An alternative to the laborious method of counting chromosomes under a microscope has become flow cytometry of cell nuclei, which is a more convenient and a faster way to determine ploidy. This hardware method is based on measuring the amount of DNA in the nuclei of cells in the mode of one-by-one analysis in a liquid flow using signals of light scattering and fluorescence, and has high accuracy and productivity (Galbraith, 2010). The flow cytometry method allows to determine the ploidy of microclones in a short time without causing significant damage to the studied plants, which is important when working with a limited volume of regenerated material.

Doubling the number of chromosomes

The main goal of induction of gynogenesis is to obtain haploids to create clean lines. For this, at the next stage, the number of chromosomes in normally developed haploids should be doubled *in vitro* or *in vivo* by methods. A. Hansen and co-authors studied the effectiveness of antimitotic agents directly in the culture of sugar beet ovules (Hansen et al., 1998). According to the results of the experiment, amiprofosmethyl showed relatively low toxicity to the embryo and contributed to obtaining an average of 4.7 diploid plants per 100 injected explants. According to S. Gürel, the most effective method for creating doubled haploids is the treatment of plants with antimitotic agents such as colchicine, oryzalin, trifluralin or short-term cultivation of shoots on nutrient media containing these substances. Colchicine is the most commonly used alkaloid, which is able to inhibit the formation of the fission spindle at the prophase stage and to stop the separation of chromosomes to the poles of daughter cells. Violation of this process leads to a doubling of the number of chromosomes in the mother's cell. S. Gürel and co-authors (2000) showed in their studies that treatment of haploids with colchicine and trifluralin gave similar results, and both agents were more effective when used in the form of liquid solutions than when added to an agar medium. For successful diploidization, scientists immersed the grown haploids in a liquid MS medium containing 150 mg/L of an antimitotic agent (colchicine), 1 mg/L of BAP, and 3 % sucrose for a period of 48 hours at a temperature of 27 °C. After treatment, the shoots were washed with sterile distilled water and transferred onto solid MS medium supplemented with 1 mg/L BAP. The number of plants with a doubled chromosome set reached 29.1 %.

To obtain DH plants, E. Weich and M. Levall first removed the tips from the roots of haploid plants, and then incubated them for 5 hours in a solution of 0.2 % colchicine and 0.25 % DMSO (Weich, Levall, 2003). A similar technique was also used with a 0.3 % colchicine solution, in which the regenerants were immersed in the root system for 24 hours, and then they were planted in the soil mixture. At the same time, 19 % of the treated haploid plants doubled the set of chromosomes (Svirshchevskaya, Dolezel, 2000). It has also been reported that applying a 0.1 % colchicine solution with 2 % DMSO to the meristem of a sugar beet haploid once a day for 3 days would result in the doubling of the chromosome set (D'Halluin, Keimer, 1986). M. Ragot and P. Steen (1992) placed a cotton swab moistened with 0.2 % colchicine solution on the apical buds of potted haploid plants for three days and got up to 50 % doubled haploids.

It should be noted that after the treatment of haploids with antimitotic agents, some time later, repeated analysis of the ploidy of microclones and careful selection of the obtained doubled haploids from the entire volume of experimental material will be necessary.

In modern breeding programs, molecular genetic methods are additionally used to determine the valuable agricultural properties of the obtained forms. These methods are usually aimed at identifying stress resistance genes and target alleles responsible for encoding a certain trait. This makes it possible to significantly reduce the breeding process and accumulate the desired alleles in one genotype. Marker assisted selection (MAS) has become the main molecular selection method, which is widely used in the breeding of many crops (Muranty et al., 2014). Genetic markers can be used at various stages of the biotechnological process and breeding programs in order to select promising genotypes.

Rooting of doubled haploids and obtaining stecklings

The next stage is the rooting of the obtained plants of doubled haploids *in vitro* in order to increase their survival when adapting to outdoor conditions. During the period of preparation of doubled haploids for planting in greenhouse

conditions, one may also encounter a low frequency of root formation. This process under *in vitro* conditions can be delayed by bright lighting, required by the green part of the microclones, by a high concentration of salts and carbohydrates or by the presence of hormones and a low concentration of oxygen in the nutrient medium. Due to the different genotypic response, the search for optimal nutrient substrate composition and microclone culture conditions continues.

Normally developed DH microclones with a developed root system are planted in the greenhouses. Due to the transplantation from *in vitro* conditions, the stage of adaptation of the planted plants lasts up to 4 weeks; meanwhile, a gradual decrease in air humidity is carried out in the greenhouse. Subsequently, after 2–3 months of growing, the stecklings are formed and the plants will be ready for harvesting. The biotechnological cycle of creating a new homozygous material ends with the stage of artificial vernalization of the stecklings at low temperatures. After that, the homozygous material is sent to further stages of the breeding process, it is planted in experimental field conditions in order to grow flowering plants and carry out crosses.

Conclusion

Thanks to the introduction of biotechnology, the process of creating new sugar beet hybrids can be significantly accelerated. Obtaining doubled haploids can significantly reduce the time and resources spent on creating clean lines. The most successful method for obtaining *Beta vulgaris* L. haploids is the method of induced gynogenesis that implies the cultivation of non-pollinated ovules *in vitro* with the subsequent formation of plants with a haploid set of chromosomes. To create DH lines, doubling of the number of haploid chromosomes using antimitotic agents, ploidy control of the created material, and growing of microclones in greenhouse conditions are used. The biotechnological stage ends with the production of doubled haploid plant plugs.

Processes in tissue culture of sugar beet, in particular the induction of H and DH forms, still require additional research. Analysis of the scientific literature shows that in order to maximize efficiency and reproducibility, the production of doubled haploids *Beta vulgaris* L. needs to be studied more thoroughly and be improved using new approaches. Biotechnology laboratories need to be able to obtain haploids in a sufficiently large amount. Therefore, the methods of cultivating explants, diploidization, rooting, and adaptation of regenerants when transplanted from sterile *in vitro* conditions into soil, require an increase in efficiency. Whereas improving each stage of the process is still an urgent task, the ultimate goal will be to increase the quality and volume of the output of the finished homozygous material.

References

Barclay I.R. High frequencies of haploid production in wheat (*Triticum aestivum*) by chromosome elimination. *Nature*. 1975;256:410-411.

Basu S.K., Eudes F., Kovalchuk I. Role of recA/RAD51 gene family in homologous recombination repair and genetic engineering of transgenic plants. In: Kumar A., Sopory S. (Eds.) Applications of Plant Biotechnology: In vitro propagation, plant transformation and secondary metabolite production. Ch. 12. India, New Delhi: I K Int. Publ. House Pvt. Ltd., 2010;231-255.

Blakeslee A.F., Belling J., Farnham M.E., Bergner A.D. A haploid mutant in the Jimson Weed, "*Datura stramonium*". *Science.* 1922;55(1433):646-647. DOI 10.1126/science.55.1433.646.

Bohanec B. Doubled haploids *via* gynogenesis. In: Touraev A., Foster B.P., Jain E.M. (Eds.) Advances in Haploid Production in Higher Plants. Dordrecht: Springer Science + Business Media BV, 2009;35-46. DOI 10.1007/978-1-4020-8854-4.

Bosemark N.O. Haploids and homozygous diploids, triploids and tetraploids in sugar beet. *Hereditas*. 1971;69(2):193-204. DOI 10.1111/j.1601-5223.1971.tb02433.x.

Bossoutrot D., Hosemans D. Gynogenesis in *Beta vulgaris* L.: from *in vitro* culture of unpollinated ovules to the production of doubled haploid plants in soil. *Plant Cell Rep.* 1985;4:300-303. DOI 10.1007/BF00269883.

Chaikam V., Molenaar W., Melchinger A.E., Boddupalli P.M. Doubled haploid technology for line development in maize: technical advances and prospects. *Theor. Appl. Genet*. 2019;132:3227-3243. DOI 10.1007/s00122-019-03433-x.

Cistué L., Soriano M., Castillo A.M., Vallés M.P., Sanz J.M., Echavarri B. Production of doubled haploid in durum wheat (*Triticum turgidum* L.) through isolated microspore culture. *Plant Cell Rep.* 2006;25(4):257-264. DOI 10.1007/s00299-005-0047-8.

Clausen R.E., Mann M.C. Inheritance in *Nicotiana tabacum*: V. The occurrence of haploid plants in interspecific progenies. *Proc. Natl. Acad. Sci. USA.* 1924;10(4):121-124. DOI 10.1073/pnas.10.4.121.

Datta S.K. Androgenic haploids: factors controlling development and its application in crop improvement. *Curr. Sci.* 2005;89(11):1870-1878. DOI 10.1007/978-94-017-1203-319.

Datta S.K., Wenzel G. Isolated microspore derived plant formation via embryogenesis in *Triticum aestivum* L. *Plant Sci.* 1987;48:49-54.

D'Halluin K., Keimer B. Production of haploid sugarbeets (*Beta vulgaris* L.) by ovule culture. In: Horn W., Jensen C.J., Odenbach W., Schieder O. (Eds.) Genetic Manipulation in Plant Breeding. Berlin: Walter de Gruyter & Co., 1986;307-309.

Djatchouk T.I., Akinina V.N., Khomyakova O.V., Kalashnikova E.V. Distant hybridization as a method of haploid production in cereals. *Biotekhnologiya i Seletsiya Rasteniy = Plant Biotechnology and Breeding*. 2019;2(2):44-52. DOI 10.30901/2658-6266-2019-2-44-52. (in Russian)

Doctrinal M., Sangwan R.S., Sangwan-Norreel B.S. *In vitro* gynogenesis in *Beta vulgaris* L.: effects of plant growth regulators, temperature, genotypes and season. *Plant Cell Tissue Organ Cult.* 1989;17:1-12. DOI 10.1007/BF00042276.

Domblides E.A., Belov S.N., Soldatenko A.V., Pivovarov V.F. Production of doubled haploids in cucumber. *Ovoshchi Rossii = Vegetable Crops of Russia.* 2019;5:3-14. DOI 10.18619/2072-9146-2019-5-3-14. (in Russian)

Dunwell J.M. Haploids in flowering plants: origins and exploration. *Plant Biotechnol. J.* 2010;8(4):377-424. DOI 10.1111/j.1467-7652.2009.00498.x.

Gains E.F., Aase H.C. A haploid wheat plant. *Am. J. Bot.* 1926;13(6):373-385.

Gałązka J., Słomnicka R., Góral-Radziszewska K., Niemirowicz-Szczytt K. From pollination to DH lines – verification and optimization of protocol for production of doubled haploids in cucumber. *Acta Sci. Pol. Hortorum Cultus.* 2015;14(3):81-92.

Galbraith D.W. Flow cytometry and fluorescence-activated cell sorting in plants: the past, present, and future. *Biomédica*. 2010; 30(Suppl.):65-70. DOI 10.7705/biomedica.v30i0.824.

Granda H.R.K. Micropropagation of chrysanthemums. *Izvestiya Timiryazevskoy Selskokhozyaystvennoy Akademii = Bulletin of the Timiryazev Agricultural Academy.* 2009;1:145-148. (in Russian)

Guha S., Maheshwari S. *In vitro* production of embryos from anthers of *Datura*. *Nature*. 1964;204:497. DOI 10.1038/204497a0.

Guha S., Maheshwari S. Cell division and differentiation of embryos in the pollen grains of *Datura in vitro*. *Nature*. 1966;212:97-98. DOI 10.1038/212097a0.

Gürel E. Callus and root development from leaf explants of sugar beet (*Beta vulgaris* L.): variability at variety, plant and organ level. *Turk. J. Botany.* 1997;21:131-136.

Gürel E., Gürel S., Lemaux P.G. Biotechnology applications for sugar beet. *Crit. Rev. Plant Sci*. 2008;27(2):108-140. DOI 10.1080/07352680802202000.

Gürel S., Gürel E., Kaya Z. Doubled haploid plant production from unpollinated ovules of sugar beet (*Beta vulgaris* L.). *Plant Cell Rep*. 2000;19(12):1155-1159. DOI 10.1007/s002990000248.

Hansen A.L., Gertz A., Joersbo M., Andersen S.B. Short-duration colchicine treatment for *in vitro* chromosome doubling during ovule culture of *Beta vulgaris* L. *Plant Breed*. 1995;114(6):515-519. DOI 10.1111/j.1439-0523.1995.tb00847.x.

Hansen A.L., Gertz A., Joersbo M., Andersen S.B. Antimicrotubule herbicides for *in vitro* chromosome doubling in *Beta vulgaris* L. ovule culture. *Euphytica*. 1998;101(2):231-237. DOI 10.1023/a:1018380103304.

Hontarenko S.M., Herasymenko H.M. Direct induced androgenesis in culture *in vitro* in sugar beet (*Beta vulgaris*). *Plant Varieties Studying and Protection*. 2018;14(4):375-381. DOI 10.21498/2518-1017. 14.4.2018.151900. (in Ukranian)

Hosemans D., Bossoutrot D. Induction of haploid plants from *in vitro* culture of unpollinated beet (*Beta vulgaris* L.). *Z. Pflanzenzücht*. 1983;91(1):74-77.

Inagaki M.N., Tahir M. Comparison of haploid production frequencies in wheat varieties crossed with *Hordeum bulbosum* L. and maize. *Jpn. J. Breed*. 1990;40(2):209-216. DOI 10.1270/jsbbs1951.40.209.

Jones O.P., Pontikis C.A., Hopgood M.E. Propagation *in vitro* of five apple scion cultivars. *J. Hortic. Sci*. 1979;54(2):155-158. DOI 10.1080/00221589.1979.11514863.

Kasha K.J., Simion E., Oro R., Yao Q.A., Hu T.C., Carlson A.R. An improved *in vitro* technique for isolated microspore culture of barley. *Euphytica*. 2001;120:379-385. DOI 10.1023/A:1017564100823.

Katayama Y. Haploid formation by X-rays in *Triticum monococcum*. *Cytologia.* 1934;5(2):235-237.

Kelliher T., Starr D., Wang W., McCuiston J., Zhong H., Nuccio M.L., Martin B. Maternal haploids are preferentially induced by CENH3-tailswap transgenic complementation in maize. *Front. Plant Sci.* 2016;7:414. DOI 10.3389/fpls.2016.00414.

Kiełkowska A., Adamus A. Baranski R. An improved protocol for carrot haploid and doubled haploid plant production using induced parthenogenesis and ovule excision *in vitro*. *In Vitro Cell. Dev. Biol. Plant*. 2014;50(3):376-383. DOI 10.1007/s11627-014-9597-1.

Labbani Z., Buyser J.D., Picard E. Effect of mannitol pretreatment to improve green regeneration on isolated microspore culture in *Triticum turgidum* ssp. *durum* cv. 'Jennah Khetifa'. *Plant Breed*. 2007;126(6):565-568. DOI 10.1111/j.1439-0523.2007.01399.x.

Laurie D.A., Bennett M.D. Wheat × maize hybridization. *Can. J. Genet. Cytol*. 1986;28(2):313-316. DOI 10.1139/g86-046.

Levan A. A haploid sugar beet after colchicine treatment. *Hereditas*. 1945;31:399-410. DOI 10.1111/j.1601-5223.1945.tb02760.x.

Lichter R. Induction of haploid plants from isolated pollen of *Brassica napus*. *Z. Pflanzenphysiol.* 1982;105(5):427-434. DOI 10.1016/S0044-328X(82)80040-8.

Lux H., Herrmann L., Wetzel C. Production of haploid sugar beet (*Beta vulgaris* L.) by culturing unpollinated ovules. *Plant Breed.* 1990;104:177-183. DOI 10.1111/j.1439-0523.1990.tb00420.x.

Mikami T., Sudoh R. Nagao E., Kinoshita T. Genotypic variation in the *in vitro* morphogenesis from leaf explants of *Beta vulgaris* L. and *Beta maritima* L. *Euphytica*. 1989;40:271-273. DOI 10.1007/BF00024523.

Muranty H., Jorje V., Bastien C., Lepoittevin C., Bouffier L., Sanchez L. Potential for marker-assisted selection for forest tree breeding: lessons from 20 years of MAS in crops. *Tree Genet. Genomes.* 2014;10:1491-1510. DOI 10.1007/s11295-014-0790-5.

Niizeki H., Oono K. Induction of haploid rice plant from anther culture. *Proc. Jpn. Acad.* 1968;44(6):554-557. DOI 10.2183/pjab1945.44.554.

Nitsch J.P., Nitsch C. Haploid plants from pollen grains. *Science*. 1969;163(3862):85-87. DOI 10.1126/science.163.3862.85.

Ouyang Y.W., Hu C.C., Chuang C.C., Tseng C.C. Induction of pollen plants from anthers of *Triticum aestivum* L. cultured *in vitro*. *Sci. Sin.* 1973;16:79-95.

Pazuki A., Aflaki F., Gürel E., Ergül A., Gürel S. Gynogenesis induction in sugar beet (*Beta vulgaris*) improved by 6-benzylaminopurine (BAP) and synergized with cold pretreatment. *Sugar Tech*. 2018a;20:69-77. DOI 10.1007/s12355-017-0522-x.

Pazuki A., Aflaki F., Gürel S., Ergül A., Gürel E. Production of doubled haploids in sugar beet (*Beta vulgaris*): an efficient method by a multivariate experiment. *Plant Cell Tissue Organ Cult*. 2018b;132(1):85-97. DOI 10.1007/s11240-017-1313-5.

Ragot M., Steen P. Genetic and environmental effects on chromosome doubling of sugarbeet (*Beta vulgaris* L.) haploids. *Euphytica*. 1992;63:233-237. DOI 10.1007/BF00024549.

Rogozinska J.H., Goska M. Attempts to induce haploids in anther culture of sugar, fodder and wild species of beet. *Acta Soc. Bot. Pol.* 1982;51(1):91-105.

Sahijram S., Rao B.M. (Eds.). Hybrid Embryo Rescue in Crop Improvement. Plant Biology and Biotechnology. Division of Biotechnology, Indian Institute of Horticultural Research (IIHR). India: Bangalore, 2015;2:363-383. DOI 10.1007/978-810322-2283-5_18.

Sears E.R. History of Chinese Spring aneuploids. In: Proc. 7th Int. Wheat Genetics Symp. 1988;1:4-6.

Seman I. Possibilities of detection and induction of haploids in *Beta vulgaris* L. *Biologia* (*Bratislava*). 1983;38(11):1113-1122.

Svirshchevskaya A.M., Dolezel FJ. Production and performance of gynogenetic sugarbeet lines. *J. Sugar Beet Res.* 2000;37(4):117-133. DOI 10.5274/jsbr.37.4.117.

Tomaszewska-Sowa M., Figas A.S., Gatz A. Histological analysis of organogenesis and somatic embryogenesis during shoot formation in sugar beet (*Beta vulgaris* L.) via gynogenesis. *Polish J. Nat. Sci.* 2017;32(4):705-717.

Touraev A., Forster B.P., Jain S.M. (Eds.) Advances in Haploid Production in Higher Plants. Dordrecht: Springer Science + Business Media BV, 2009;347. DOI 10.1007/978-1-4020-8854-4.

Tuvesson I.K.D., Öhlund R.C.V. Plant regeneration through culture of isolated microspores of *Triticum aestivum* L. *Plant Cell Tissue Organ Cult.* 1993;34:163-167.

Urazaliyev K.R., Orsini J.M., Abekova A.M., Bazylova T.A., Daniyarova A.K. Speeding wheat breeding using dihaploids obtained by microspore culture. *Vestnik KazNU. Seriya Ekologiche-*

skaya = Bulletin of the KNU: Environmental Series. 2013;38(2/2): 369-374.

Van Geyt J., Speckmann G.J., D'Halluin K., Jacobs M. *In vitro* induction of haploid plants from unpollinated ovules and ovaries of the sugarbeet (*Beta vulgaris* L.). *Theor. Appl. Genet.* 1987;73(6): 920-925. DOI 10.1007/BF00289399.

Watts A., Kumar V., Bhat S.R. Centromeric histone H3 protein: from basic study to plant breeding applications. *J. Plant Biochem. Biotechnol.* 2016;25:339-348. DOI 10.1007/s13562-016-0368-4.

Weich E.W., Levall M.W. Doubled haploid production of sugar beet (*Beta vulgaris* L.). In: Maluszynski M., Kasha K.J., Forster B.P., Szarejko I. (Eds.) Doubled Haploid Production in Crop Plants. New York: Springer Science + Business Media, 2003;255-263. DOI 10.1007/978-94-017-1293-4_38.

Zheng M.Y. Microspore culture in wheat (*Triticum aestivum*) – doubled haploid production via induced embryogenesis. *Plant Cell Tissue Organ Cult.* 2003;73:213-230. DOI 10.1023/A:10230 76213639.

ORCID ID
E.O. Kolesnikova orcid.org/0000-0002-7926-5129

Conflict of interest. The authors declare no conflict of interest.
Received May 19, 2021. Revised September 30, 2021. Accepted October 8, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):822-830
DOI 10.18699/VJ21.095

Original Russian text www.bionet.nsc.ru/vogis/

Genetic diversity of the Central European wild boar (*Sus scrofa scrofa*) population and domestic pig (*Sus scrofa domesticus*) breeds based on a microsatellite DNA locus

E.A. Snegin[1] ✉, V.M. Makeeva[2], A.P. Kaledin[3], A.M. Ostapchuk[3], I.D. Alazneli[2], A.V. Smurov[2]

[1] Belgorod National Research University, Belgorod, Russia
[2] Lomonosov Moscow State University, Moscow, Russia
[3] Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, Russia
✉ snegin@bsu.edu.ru

Abstract. The results of studies of the genetic structure of the Central European wild boar (*Sus scrofa scrofa*) population and four breeds of domestic pigs (Duroc, Yorkshire, Large White and Landrace) bred in the Central Black Earth region of Russia are presented in this work. Based on 12 microsatellite loci, a significant (p <0.05) decrease in the level of genetic variability in bred breeds was shown. The expected heterozygosity and Shannon index were as follows: in the wild boar, $Ho = 0.763 \pm 0.026$, $I = 1.717 \pm 0.091$; in the Duroc breed, $Ho = 0.569 \pm 0.068$, $I = 1.191 \pm 0.157$; in the Landrace, $Ho = 0.618 \pm 0.062$, $I = 1.201 \pm 0.147$; in the Large White, $Ho = 0.680 \pm 0.029$, $I = 1.362 \pm 0.074$; and in the Yorkshire, $Ho = 0.642 \pm 0.065$, $I = 1.287 \pm 0.156$. The results of checking genotypic Hardy–Weinberg equilibrium based on the G-test of maximum likelihood demonstrated that the overwhelming majority of loci in the wild boar population were in the state of said equilibrium. By contrast, in pig breed populations, some loci demonstrated a significant deviation from the indicated equilibrium. In addition, the Yorkshire, Large White, and Landrace populations had loci, for which the hypothesis of neutrality was reliably rejected based on the results of the Ewens–Watterson test. The revealed private alleles, characteristic of the wild boar and breeds, can later be used to identify them. The ordination of the centroids of different herds in the space of the first two principal coordinates based on the matrix of pairwise estimates of Nei's genetic distances showed that the most distant populations are the Duroc and Boar breeds, and the most genetically close are the Yorkshire and Landrace breeds. The closest to the wild boar population was the Large White breed. The assessment of the effective size, carried out using the method based on the linkage disequilibrium and the molecular coancestry method, showed that in all studied groups, including the wild boar population, the effective size was less than 100 individuals. The low effective size of the wild boar population (Ne = 21.8, Neb = 4.0) is probably caused by the death and shooting of animals due to *Pestis africana suum*.

Key words: wild boar; pig breeds; microsatellite loci; genetic structure; effective population size.

For citation: Snegin E.A., Makeeva V.M., Kaledin A.P., Ostapchuk A.M., Alazneli I.D., Smurov A.V. Genetic diversity of the Central European wild boar (*Sus scrofa scrofa*) population and domestic pig (*Sus scrofa domesticus*) breeds based on a microsatellite DNA locus. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8): 822-830. DOI 10.18699/VJ21.095

Генетическое разнообразие популяции центральноевропейского кабана (*Sus scrofa scrofa*) и пород домашних свиней (*Sus scrofa domesticus*) на основе микросателлитных локусов ДНК

Э.А. Снегин[1] ✉, В.М. Макеева[2], А.П. Каледин[3], А.М. Остапчук[3], И.Д. Алазнели[2], А.В. Смуров[2]

[1] Белгородский государственный национальный исследовательский университет, Белгород, Россия
[2] Московский государственный университет им. М.В. Ломоносова, Москва, Россия
[3] Российский государственный аграрный университет – МСХА им. К.А. Тимирязева, Москва, Россия
✉ snegin@bsu.edu.ru

Аннотация. В работе приведены результаты исследований генетической структуры популяции центральноевропейского кабана (*Sus scrofa scrofa*) и четырех пород домашних свиней (дюрок, йоркшир, крупная белая и ландрас), разводимых в Центрально-Черноземном регионе России. На основе 12 микросателлитных локусов установлено достоверное (p < 0.05) снижение генетической изменчивости в разводимых породах. Ожидаемая гетерозиготность и индекс Шеннона были равными: у кабана – $Ho = 0.763 \pm 0.026$, $I = 1.717 \pm 0.091$;

у пород дюрок – *Ho* = 0.569 ± 0.068, *I* = 1.191 ± 0.157; ландрас – *Ho* = 0.618 ± 0.062, *I* = 1.201 ± 0.147; крупная белая – *Ho* = 0.680 ± 0.029, *I* = 1.362 ± 0.074; йоркшир – *Ho* = 0.642 ± 0.065, *I* = 1.287 ± 0.156. Результаты проверки генотипического равновесия Харди–Вайнберга на основе G-теста максимального правдоподобия показали, что в популяции кабана большинство локусов находилось в состоянии генотипического равновесия Харди–Вайнберга. Напротив, в популяциях различных пород свиней часть локусов демонстрирует достоверное отклонение от отмеченного равновесия. Кроме того, в популяциях йоркшир, крупная белая и ландрас присутствовали локусы, для которых достоверно отвергалась гипотеза о нейтральности на основании результатов теста Эвенса–Ваттерсона (Ewens–Watterson test). Обнаруженные приватные аллели, характерные для кабана и различных пород, в дальнейшем могут быть использованы для их идентификации. Ординация центроидов разных стад в пространстве первых двух главных координат на основании матрицы попарных оценок генетических дистанций M. Nei показала, что наиболее удаленные популяции – породы дюрок и кабан, а самые генетически близкие – йоркшир и ландрас. Ближе всех к популяции кабана была порода крупная белая. Оценка эффективной численности, проведенная с использованием метода, основанного на неравновесии по сцеплению (linkage disequilibrium) и MC-метода (the molecular coancestry method), продемонстрировала, что во всех изученных группах, включая и популяцию кабана, эффективный размер оказался меньше 100 особей. Низкое значение эффективного размера популяции кабана (*Ne* = 21.8, *Neb* = 4.0), вероятно, является следствием падежа и отстрела животных из-за африканской чумы свиней (*Pestis africana suum*).
Ключевые слова: кабан; породы свиней; микросателлитные локусы; генетическая структура; эффективная численность популяции.

Introduction

According to various estimates, the domestication of the wild boar began 7–9 thousand years ago. During this time, more than 730 breeds of these animals have been created by man. It is obvious that for such a long "cultural" evolution, various breeds, being so-called clean lines, have largely lost the natural genetic potential that provides homeostatic mechanisms. As a result, maintaining the stability of existing breeds, like any artificially created systems, requires significant financial investments. In this regard, the study of the genetic potential of natural populations of wild boars for a possible increase in the resistance of pig breeds (for example, by methods of genome editing, etc.) is highly relevant.

In this regard, the study of wild boar genetics is now receiving much attention, both in Russia and other countries (Gladyr et al., 2009; Zinovieva et al., 2013; Rębała et al., 2016; Mihalik et al., 2020). In addition, due to constant outbreaks of the African swine fever (*Pestis africana suum*) in Russia, the wild boars are regularly shot as potential carriers of this disease. At the same time, the population structure of this animal is not taken into account, which can cause depletion of the gene pool and, against the background of increasing anthropogenic pressure, lead to the extinction of some groups. Examples of a significant reduction in the population of commercial species occurring due to the data on the state of the gene pool being neglected are well known (Altukhov, 2003).

Also, in the practice of molecular genetic laboratories, for forensic purposes, it is often necessary to diagnose tissue samples from illegally caught wild boars and prove their belonging to a wild species, and not to domesticated forms of pigs (Kipen et al., 2016; Lorenzini et al., 2020), or identify wild boar tissue in food (Szemethy et al., 2021). In this regard, the identification of private alleles for natural populations is also an urgent problem.

Microsatellite DNA loci (STR markers), which are tandem repeats of the noncoding part of nuclear DNA, are very convenient markers for studying genetic processes in populations. There are many works on the assessment of population gene pools of both domestic pigs and wild boars in different regions (Vernesi et al., 2003; Ferreira et al., 2009; Nikolov et al., 2009; da Silva et al., 2011; Choi et al., 2014; Sahoo et al., 2016; Ryabtseva et al., 2018; Han et al., 2021; Snegin et al., 2021).

The purpose of this work is to assess the genetic diversity of microsatellite loci in the population of the Central European wild boar (*Sus scrofa scrofa*) and the most common pig breeds (*Sus scrofa domesticus*) bred in the Central Black Earth region of Russia. It should be noted that such studies have not previously been conducted in this region.

Materials and methods

A total of 320 animals were involved in the study. A sample of 30 wild boars was taken from the populations of the Oryol region (districts: Korsakovsky, Zalegoshchennsky, Novosilsky, Pokrovsky, Shablykinsky), Russia. The wild boars were caught during the hunting season of 2018. For comparison, samples from four populations of different breeds of domestic pigs bred on the farms of the Central Black Earth Region of Russia were used: Durok – 67 individuals (Belgorod region), Yorkshire – 108 individuals (Kursk region), Landrace – 50 individuals (Belgorod region), large white – 65 individuals (Voronezh region). All analyzed animals are Canadian breeds.

All 12 microsatellite loci recommended by ISAG-FAO (International Society for Animal Genetics, Food and Agriculture Organization) (FAO SoW-AnGR…, 2006) and arranged in one multiplex panel (*S0101*, *S0155*, *S0228*, *S0355*, *S0386*, *SW24*, *SW240*, *SW72*, *SW857*, *SW911*, *SW936*, *SW951*) were used as DNA markers (Table 1). Primers for PCR were selected taking into account the amplification of all 12 loci in one tube. The size of all amplified PCR products, taking into account all known alleles, was < 300 base pairs.

In the domestic pigs, DNA extraction was carried out from ear pits, and in the wild boars – from muscle tissue samples. For this purpose, we used kits with proteinase K DNA-Extran-2 (Syntol, Russia). The PCR reaction was carried out on a Verity amplifier (Applied Biosystems, USA) in 20 μL of a mixture containing 20 ng of genomic DNA, PCR buffer (10 mmol Tris-HCl (pH 8.3), 50 mmol KCl, 2 mmol $MgCl_2$), 0.25 mmol dNTP , 0.5 μmol primer, 1 unit *Taq* DNA polymerase (inhibited for hot start).

Table 1. The characteristics of the microsatellite loci recommended by ISAG for determining the reliability of the origin pigs

Locus	Allele length, bp	Fluorescent Dye	Primers (5′--------3′)
S0101	193–221	R6G	F: GAATGCAAAGAGTTCAGTGTAGG R: GTCTCCCTCACACTTACCGCAG
S0155	142–166	TAMRA	F: TGTTCTCTGTTTCTCCTCTGTTTG R: AAAGTGGAAAGAGTCAATGGCTAT
S0228	218–270	TAMRA	F: GGCATAGGCTGGCAGCAACA R: AGCCCACCTCATCTTATCTACACT
S0355	223–277	FAM	F: TCTGGCTCCTACACTCCTTCTTGATG R: TTGGGTGGGTGCTGAAAAATAGGA
S0386	164–182	FAM	F: GAACTCCTGGGTCTTATTTTCTA R: GTCAAAAATCTTTTTATCTCCAACAGTAT
SW24	95–124	ROX	F: CTTTGGGTGGAGTGTGTGC R: ATCCAAATGCTGCAAGCG
SW240	92–124	R6G	F: AGAAATTAGTGCCTCAAATTGG R: AAACCATTAAGTCCCTAGCAAA
SW72	97–125	TAMRA	F: ATCAGAACAGTGCGCCGT R: TTTGAAAATGGGGTGTTTCC
SW857	137–161	R6G	F: TGAGAGGTCAGTTACAGAAGACC R: GATCCTCCTCCAAATCCCAT
SW911	149–177	ROX	F: CTCAGTTCTTTGGGACTGAACC R: CATCTGTGGAAAAAAAAAGCC
SW936	81–117	FAM	F: TCTGGAGCTAGCATAAGTGCC R: GTGCAAGTACACATGCAGGG
SW951	124–134	FAM	F: TTTCACAACTCTGGCACCAG R: GATCGTGCCCAAATGGAC

PCR parameters: 94 °C – 3 min; (98 °C – 30 sec, 59 °C – 120 sec, 72 °C – 90 sec) – 4 cycles; (94 °C – 30 sec, 59 °C – 120 sec, 72 °C – 90 sec) – 6 cycles; (90 °C – 30 sec, 59 °C – 120 sec, 72 °C – 75 sec) – 20 cycles; 68 °C – 30 min. The heating rate from 59 to 72 °C was no more than 0.3 °C/1 sec.

Fragment analysis of the PCR products was carried out on an ABI PRISM 3500 automatic capillary DNA sequencer (Applied Biosystems, USA) using 50 cm capillaries and a POP-7™ polymer matrix. Fragment size analysis was performed using GeneMapper R Software v. 4.1 (Applied Biosystems).

For statistical processing of the data obtained we used the GenAIEx v. 6.5 (Peakall, Smouse, 2012) and PorGene 1.32 (Yeh et al., 1999).

Results

In Table 2, the allele frequencies of the microsatellite loci used for analysis are presented. The data provide insight into the distribution of various alleles among populations of domestic pigs and wild boars.

The presence of private alleles in different populations is shown in Table 3. The results show a high content of unique alleles in the wild boar population (16 alleles). The highest frequency of private alleles was found at the loci *SW24* and *SW72* (0.25 in each). At the same time, the allele 97 at the *SW24* locus, and the allele 99 at the *SW72* locus are found more often than others among private alleles. The pig of the Duroc breed is slightly inferior to the wild boar in the number of private alleles, while the 105 and 111 alleles at the *SW936* locus (0.246 and 0.276, respectively) had the highest frequency of occurrence. The Large White breed is almost three times inferior in the number of private alleles to the population of wild boar and Duroc pigs. However, some of the original alleles are found in this group of animals with noticeable frequency. For example, the allele 141, unique for this breed, at the *SW857* locus was observed in half of the animals analyzed (frequency 0.5). No private alleles were found in the Landrace population, and only one private allele was noted in the Yorkshire population.

The wild boar population has significantly high values of the indicators of genetic variability in comparison with the pig breeds. The comparison was carried out using the Pearson χ^2 test ($p < 0.05$) (Table 4).

The level of inbreeding (F) in the studied groups turned out to be low, and the populations of Yorkshire, Landrace and wild boar received a negative value due to the prevalence of observed heterozygosity over the theoretically expected heterozygosity (see Table 4).

The results of checking the genotypic balance of Hardy–Weinberg based on the G-test of maximum likelihood demonstrated that in the wild boar population the overwhelm-

Table 2. Frequencies of the microsatellite loci alleles in the populations of various pig breeds and wild boars

Locus	Population					Locus	Population					Locus	Population				
S0101	1	2	3	4	5	*S0386*	1	2	3	4	5	*SW72*	1	2	3	4	5
193	0	0	0.169	0	0	164	0.022	0	0.008	0	0.017	119	0.500	0.361	0	0.250	0.017
195	0	0.055	0	0.090	0.017	166	0.045	0	0.054	0.140	0.133	121	0.097	0	0	0	0.050
197	0	0	0	0.020	0.150	168	0	0.231	0.015	0.060	0.150	123	0.052	0	0	2	0
199	0	0.097	0	0.030	0.017	170	0.007	0	0.131	0.090	10	*SW857*	1	2	3	4	5
201	0	0.028	0	0	0.083	172	0.060	0.014	0.023	0	0.117	137	0.007	0	0	0	0
203	0	0	0	0	0.050	174	0.112	0.14	0	0.010	0.017	141	0	0	0.500	0	0
205	0.007	0	0.015	0	0	176	0.015	0	0.215	0	0.033	143	0	0	0.008	0	0
207	0	0.319	0.008	0.310	0.317	178	0.022	0.630	0.346	0.630	10	147	0.007	0	0	0	0
209	0	0.241	0.061	0.280	0.200	180	0.649	0.083	0.208	0.040	0.017	149	0	0	0.231	0	0.317
211	0.851	0.014	0.746	0.210	0.100	182	0.067	0.028	0	0.030	11	151	0	0.352	0.100	0.370	0.150
213	0.142	0.176	0	0.060	0	*SW24*	1	2	3	4	5	153	0.052	0.194	0.077	0	0.333
215	0	0.069	0	0	2	95	0	0	0	0	0.050	155	0.627	0.014	0	0	0.050
217	0	0	0	0	2	97	0	0	0	0	0.183	157	0	0.241	0.054	0.470	0.067
S0155	1	2	3	4	5	99	0	0	0	0	0.017	159	0.298	0.143	0.031	0.160	0.083
144	0.015	0	0	0	0	101	0.037	0	0	0	0.017	161	0.007	0.056	0	0	0
146	0.022	6	0.061	0	0.433	105	0	0	0.046	0.410	0.050	*SW911*	1	2	3	4	5
148	0.007	0.352	0.008	0.250	0.117	107	0	0.245	0	0	0.050	149	0	0	0.046	0	0.017
150	0	0	0	0	0.017	109	0.455	0.157	0	0	0.050	151	0.007	0	0	0	0
152	0.007	0	0	0	0	111	0	0	0.261	0.010	0.017	153	0.022	3	0	0	0
154	0	0	0	0	0.133	113	0	0.018	0.046	0	0	155	0	0	0.208	0	0
156	0.589	0	0.208	0	0.017	115	0.134	0.111	0.069	0.150	0.067	157	0.187	0.217	0.100	0.350	0
158	0	0.264	0.346	0.420	0.100	117	0	0.060	0.354	0.310	0.267	159	0	0	0.485	0	0.200
160	0.022	0.356	0.223	0.330	0.083	119	0.052	0.380	0.215	0.070	0.067	161	0	0.444	0.054	0.430	0.117
162	0.328	0	0.154	0	0.083	121	0.216	0.028	0.008	0.030	0.133	163	0.537	0	0.008	0	0.117
164	0.007	0	0	0	0	123	0.097	0	0	0.020	0.033	165	0	0	0.015	0	0.333
166	0	0	0	0	0.017	125	0.007	0	0	0	0	167	0	0.264	0.085	0.130	0.017
S0228	1	2	3	4	5	*SW240*	1	2	3	4	5	*SW911*	1	2	3	4	5
220	0.015	0.236	0.008	0.240	0	93	0	0.305	0	0	0	169	0.119	0.060	0	0.090	11
222	0.134	0.278	0	0.060	0.017	97	0.007	0	0.231	0	0.017	171	0.112	0	0	0	0.017
224	0.157	0.018	0	0.040	0	99	0	0	0.008	0	0.367	177	0.015	0	0	0	0
226	0	0	0.061	0.030	0	101	0.060	0	0.185	0	0.033	*SW936*	1	2	3	4	5
228	0	0.056	0.023	0.050	0.433	103	0.082	0	0	0	0.017	85	0	0	0	0.500	0.033
230	0.037	0	0	0.020	0.017	105	0.254	0	0.200	0	0	87	0.015	0.500	0	0	0
232	0.022	0.014	0	0.090	0.017	107	0.164	0	0	0.030	0	91	0	0	0	0.500	0.383
234	0.022	0.167	0	0.330	0	109	0.246	0.046	0	0.120	0	93	0	0.500	0.346	0	0.150
236	0.007	0.060	0	0.040	0.267	111	0	0.028	0	0.030	0	95	0	0	0.254	0	0
238	0	0.028	0.061	0.010	0.050	113	0	0.018	0.315	0	0.233	97	0	0	0	0	0.150
240	0.007	0.055	0.015	0.010	0.033	115	0	0.055	0	0	0.133	99	0	0	0	0	0.050
242	0	0.028	0.015	0.020	0.083	117	0.030	0.088	0.054	0.430	0	101	0	0	0	0	0.017
244	0	0	0.515	0.020	0.050	119	0.127	0.296	0	0	0	103	0.007	0	0	0	0.033
246	0.366	0.032	0.031	0	0	121	0.015	0.162	0.008	0.390	0.033	105	0.246	0	0	0	0
248	0.231	0.028	0.269	0.040	0.033	123	0.015	0	0	0	0.167	107	0.022	0	0.231	0	0.183
S0355	1	2	3	4	5	*SW72*	1	2	3	4	5	109	0.045	0	5	0	0
243	0	0	0.415	0	0.667	99	0	0	0	0	0.233	111	0.276	0	0	0	0
245	0	0.125	0	0.110	0.083	101	0.015	0	0	0	0.033	113	0.246	0	0.061	0	0
247	1.0	0	0.300	0	0	103	0	0	0	0	0.017	115	0.097	0	0.069	0	0
249	0	0.088	0	0.190	0	105	0	0	0.162	0.420	0.133	117	0.045	0	0	0	0
255	0	0	0	0	0.033	107	0	0.245	0.115	0	0.017	*SW951*	1	2	3	4	5
259	0	0.301	0	0.210	0.100	109	0.298	0.171	0	0	0.050	126	0.134	0	0.154	0	0.383
261	0	0	0	0	0.050	111	0.007	0	0	0	0	128	0.045	1.0	0.246	1.0	0.200
269	0	0	0	0	0.067	113	0.015	0	0.531	0	0.133	130	0.619	0	0.054	0	0.067
271	0	0	0.285	0.160	0	115	0	0.181	0.192	0.220	0.150	132	0.134	0	0.546	0	0.350
273	0	0.486	0	0.330	0	117	0.015	0.042	0	0.090	0.167	134	0.067	0	0	0	0

Note. Population: 1 – Duroc; 2 – Yorkshire; 3 – Large White; 4 – Landrace; 5 – Wild Boar.

Table 3. Private alleles in the studied populations of various breeds of pigs and wild boars

Population	Locus									
	S0101	*S0155*	*S0355*	*SW24*	*SW240*	*SW72*	*SW857*	*SW911*	*SW936*	*SW951*
Duroc		144, 152, 164		125		111	137, 147	151, 177	105, 111, 117	134
Yorkshire					93					
Large White	193						141, 143	155	95	
Wild Boar	203, 217	150, 154, 166	255, 261, 269	95, 97, 99		99, 103			97, 99, 101	

Note. No private alleles were found in the Landrace pig population.

Table 4. The indicators of the genetic diversity in the studied groups of pigs and wild boars (Mean ± SE)

Population	*N*	*P*, %	*Aa*	*Ae*	*I*	*Ho*	*He*	*F*	N_{pa}
Duroc	67	91.7	6.917 ± 0.802	2.913 ± 0.396	1.191 ± 0.157	0.525 ± 0.079	0.569 ± 0.068	0.076 ± 0.076	1.083 ± 0.336
Yorkshire	108	91.7	5.667 ± 0.847	3.452 ± 0.384	1.287 ± 0.156	0.716 ± 0.086	0.642 ± 0.065	–0.128 ± 0.104	0.083 ± 0.083
Large White	65	100.0	6.167 ± 0.534	3.350 ± 0.241	1.362 ± 0.074	0.660 ± 0.060	0.680 ± 0.029	0.022 ± 0.075	0.417 ± 0.193
Landrace	50	91.7	5.250 ± 0.978	3.124 ± 0.336	1.201 ± 0.147	0.713 ± 0.081	0.618 ± 0.062	–0.175 ± 0.101	0.000
Wild Boar	30	100.0	8.583 ± 0.712	4.702 ± 0.444	1.717 ± 0.091	0.844 ± 0.038	0.763 ± 0.026	–0.106 ± 0.033	1.333 ± 0.414

Note. *N* – the number of individuals; *P* – the percentage of polymorphic loci; *Aa* – the average number of alleles; *Ae* – the effective number of alleles; *I* – Shannon's index; *Ho* – observed heterozygosity; *He* – expected heterozygosity; *F* – the coefficient of inbreeding; N_{pa} – the average number of private alleles per locus; Mean ± SE – mean ± standatd error.

Table 5. The results of checking the genotypic balance of Hardy–Weinberg for 12 loci of MS-DNA in herds of pigs of different breeds and wild boar based on the G-test of maximum likelihood (likelihood ratio G-test)

Locus	Population				
	Duroc	Yorkshire	Large White	Landrace	Wild Boar
S0101	ns	< 0.001	ns	ns	ns
S0155	ns	ns	ns	0.019	ns
S0228	0.014	< 0.001	ns	ns	ns
S0355	mono	ns	ns	< 0.001	ns
S0386	< 0.001	< 0.001	< 0.001	< 0.001	ns
SW24	ns	< 0.001	< 0.001	0.004	ns
SW240	ns	< 0.001	< 0.001	< 0.001	ns
SW72	ns	< 0.001	ns	0.001	ns
SW857	ns	< 0.001	ns	< 0.001	ns
SW911	ns	< 0.001	ns	0.032	ns
SW936	ns	< 0.001	0.014	< 0.001	0.002
SW951	< 0.001	mono	0.017	mono	< 0.001

Note. ns – correspondence to Hardy–Weinberg equilibrium; $p > 0.05$; mono – monomorphic locus.

ing majority of the loci were in the state of the genotypic equilibrium of Hardy–Weinberg (Table 5). On the contrary, in the populations of the different breeds of pigs, some of the loci demonstrated a significant deviation from the indicated equilibrium. In addition, the monomorphic loci were noted in three groups (Duroc, Yorkshire, and Landrace). Also, in the populations of Yorkshire, Large White, and Landrace, the MC-loci were present, for which the hypothesis of neutrality was reliably rejected based on the results of the Ewens–Watterson test (Ewens, 1972; Watterson, 1978) (Table 6). Most of these loci were found in the Yorkshire and Landrace breeds.

The degree of similarity of the genetic structure of the pigs was assessed using the principal component analysis, which resulted in the ordination of the centroids of different herds

Table 6. The results of the Ewens-Watterson test for 12 MS-DNA loci for various breeds of pigs and boars (only the loci for which the hypothesis of neutrality is reliably rejected are shown)

Population	Locus	Obs. F	L95–U95
Yorkshire	*S0155*	0.321	0.332–0.906
	SW72	0.254	0.282–0.893
	SW936	0.500	0.502–0.991
Large White	*S0355*	0.344	0.367–0.970
Landrace	*S0155*	0.348	0.371–0.961
	S0355	0.227	0.267–0.831
	SW936	0.500	0.503–0.980

Note. Obs. F – the actual sum of squares of allele frequencies; L95, U95 – lower and upper values of the 95 % confidence interval of the Obs F estimate, calculated from 1000 simulations.

Table 7. The values of genetic distances (according to M. Nei) between the studied groups of pigs and wild boars

Population	Duroc	Yorkshire	Large White	Landrace	Wild Boar
Duroc	0.000				
Yorkshire	1.831	0.000			
Large White	0.988	1.209	0.000		
Landrace	1.976	0.318	1.087	0.000	
Wild Boar	1.803	1.091	0.707	0.968	0.000

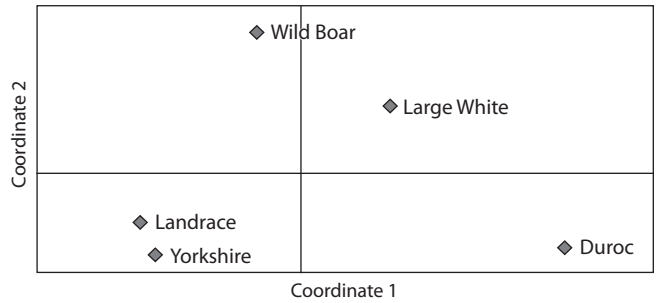
Table 8. The estimates of the effective population size (*Ne, Neb*), calculated using the LD- and MC-method between 12 loci of MS-DNA

Population	Linkage disequilibrium		Molecular coancestry	
	Ne	95 % CI	*Neb*	95 % CI
Duroc	86.1	54.9–164.2	19.4	2.3–54.0
Yorkshire	7.4	5.9–8.9	7.6	4.3–11.7
Large White	44.6	34.0–60.9	15.5	6.7–28.0
Landrace	9.0	6.6–11.8	2.1	1.3–3.5
Wild Boar	21.8	17.1–28.9	4.0	2.7–5.6

Note. 95 % CI – limits of 95 % confidence interval.

in the space of the first two Main Coordinates based on the matrix of pairwise estimates of genetic distances by M. Nei (Table 7, see the Figure). According to the data obtained, the most distant populations are the Duroc and Boar breeds, and the most genetically close are the Yorkshire and Landrace breeds. The closest to the wild boar population was the Large White breed.

The effective population size was estimated using the linkage disequilibrium (LD) method (Hill 1981; Waples, 2006; Waples, Do, 2010), as well as the molecular coancestry method (Nomura, 2008). The calculations were performed using the NeEstimator V2 software (Do et al., 2014). The results are shown in Table 8.



The ordination of centroids based on a matrix of pairwise estimates of M. Nei's genetic distances based on the distribution of allele frequencies of 12 MS-DNA loci.

Table 9. Level of actual heterozygosity for microsatellite markers in different wild boar populations

Country	*Ho*	References	Country	*Ho*	References
Portugal	0.627	Ferreira et al., 2009	Japan	0.473	Choi et al., 2014
South Korea	0.682	Han et al., 2021	Indonesia	0.658	
	0.422–0.673	Choi et al., 2014	Russia		
Bulgaria	0.620	Nikolov et al., 2009	Primorsky region	0.710	Choi et al., 2014
Germany	0.460		Kirov region	0.463	Gladyr et al., 2009
Italy	0.520-0.720	Vernesi et al., 2003	Yaroslavl region	0.535	
Hungary	0.750		Orenburg region	0.722	
China	0.845	Choi et al., 2014	Krasnodar region	0.614	
Vietnam	0.859		Khabarovsk region	0.670	

Discussion

The reliably high values of genetic variability ($p < 0.05$), noted in the wild boar population, are quite an expected phenomenon. This is despite the fact that the sample from the analyzed group was smaller than the samples from herds of domestic pigs. This clearly demonstrates the consequences of panmixis and genetic-automatic processes, which, in combination with the natural selection, form the gene pools of natural populations. Moreover, in the wild boar population we analyzed, the level of actual heterozygosity turned out to be significantly higher than in the populations of this animal both in Europe and Asia. At the same time, the Oryol group, in terms of the level of genetic diversity, turned out to be similar to the Chinese and Vietnamese wild boar populations (Table 9).

On the contrary, the transition of a number of loci to the monomorphic state and the lack of equilibrium according to Hardy–Weinberg, noted by us for most loci in domestic pig breeds, is a consequence of long-term artificial selection work, as a result of which many alleles of the "wild" type were lost, which led to the loss of genetic variety. This is also evidenced by the significant deviation from neutrality of some microsatellite loci in the Landrace, Yorkshire, and Large White breeds, which was revealed by us using the Evens–Watterson test (see Table 6).

Allelic diversity results in a large number of private alleles observed in the wild boar population. At the same time, the Duroc population, despite the large number of private alleles, turned out to be the most monomorphic among the studied groups. This probably indicates a long-term selection of this red color breed in the conditions of the North American continent, isolated from crossing with other pig breeds, including white breeds of European origin (Large White, Yorkshire and Landrace). This can explain the significant genetic distance of the Duroc breed, both from the wild boar and the European pig breeds. The presence of original alleles in the studied populations may be used in the future for identifying both the breed of pigs and belonging to a wild boar population.

It should be noted that the results obtained are in part consistent with the data obtained in the work of E.A. Glydyr et al. (2009). In this study, the level of actual heterozygosity in three out of five wild boar populations was found to be higher than in domestic pigs. However, the average number of the effective alleles was the same (*Ae* = 2.6). In terms of the number of private alleles, the wild boar population also surpassed the domestic pigs (21 versus 10, respectively).

The calculation of the effective population size based on the LD method showed that in almost all the studied groups, including the wild boar population, the effective size was less than 50 individuals. The only exception was the Duroc breed (*Ne* = 86.1). The data indicate the observed linkage disequilibrium, probably caused by closely related mating in the analyzed groups of domestic pigs. This genetic drift has generated a non-random association between alleles at different loci.

Calculations carried out using the MC method showed even lower values. This phenomenon is most likely associated with a small number of parental individuals (primarily males) who founded the studied populations. It is worth noting that a similar picture was obtained in the work on assessing the effective population size of the endangered *Gochu Asturcelta* pig breed (Menendez et al., 2016).

If the results for breeds of domestic pigs were comparable with data obtained in other studies, where *Ne* varied in general from 20 to 92 individuals (Šveistienė, Razmaitė, 2013; Krupa et al., 2015; Zanella et al., 2016; Lugovoy et al., 2018), then in relation to the wild boar population, the result was somewhat unexpected, since in previous studies *Ne* ranged from 180 to 1477 animals in natural populations (Cowled et al., 2008; Herrero-Medrano et al., 2013).

Such a low *Ne* value noted in the wild boar population of the Oryol region can be partly explained by a small sample, however, reasons behind it may be more serious. In particular, it is known that outbreaks of African swine fever (*Pestis africana suum*) (https://www.kommersant.ru/doc/4236233), resulting in the death of wild boars are often recorded on the territory of the Central Black Earth Region, which includes the specified area. In addition, in order to prevent the spread of infection, hunting farms are forced to shoot a significant part of the animals. It is likely that these phenomena affect the effective size of the wild boar populations.

Conclusion

Thus, on the basis of the studies carried out, a clear reduction in the genetic diversity of the domestic pig breeds in comparison with the natural wild boar population was demonstrated. The presence of private alleles can further aid in the identification of wild boar and different breeds of pigs. Low values of the effective size of the studied groups require attention from breeders in relation to the breeds of pigs. In particular, the pig breeding companies in the region under study need to use a larger number of producers (primarily males) to obtain replacement livestock. With regard to wild boar populations, prophylactic shooting and harvesting should be carried out under the control of environmental authorities with mandatory monitoring of the state of the population gene pools.

References

Altukhov Yu.P. Genetic Processes in Populations. Moscow: Akademkniga Publ., 2003. (in Russian)

Choi S.K., Lee J.-E., Kim Y.-J., Min M.-S., Voloshina I., Myslenkov A., Oh J.G., Kim T.-H., Markov N., Seryodkin I., Ishiguro N., Yu L., Zhang Y.-P., Lee H., Kim K.S. Genetic structure of wild boar (*Sus scrofa*) populations from East Asia based on microsatellite loci analyses. *BMC Genet*. 2014;15:85. DOI 10.1186/1471-2156-15-85.

Cowled B.D., Aldenhoven J., Odeh I.O.A., Garrett T., Moran Ch., Lapidge S.J. Feral pig population structuring in the rangelands of eastern Australia: applications for designing adaptive management units. *Cons. Genet.* 2008;9:211-224. DOI 10.1007/s10592-007-9331-1.

Do C., Waples R.S., Peel D., Macbeth G.M., Tillett B.J., Ovenden J.R. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (*Ne*) from genetic data. *Mol. Ecol. Resour.* 2014;14(1):209-214. DOI 10.1111/1755-0998.12157.

Ewens W. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* 1972;3:87-112.

FAO SoW-AnGR. The State of the World's Animal Genetic Resources. Rome, 2006.

Ferreira E., Souto L., Soares A.M.V.M., Fonseca C. Genetic structure of the wild boar population in Portugal: Evidence of a recent bottleneck. *Mamm. Biol.* 2009;74:274-285. DOI 10.1016/j.mambio.2008.05.009.

Gladyr E.A., Ernst L.K., Kostyunina O.V. Study of *Sus scrofa* genome with the use of DNA-markers. *Selskokhozyaystvennaya Biologiya = Agricultural Biology.* 2009;2:16-26. (in Russian)

Han S.W., Park H.Ch., Kim J.H., Suh J.H., Lee H., Min M.S. Population genetic structure of wild boar and dispersal performance based on kinship analysis in the Northern Region of South Korea. *Research Square*. 2021. DOI 10.21203/rs.3.rs-368091/v1.

Herrero-Medrano J.M., Megens H.J., Groenen M.A.M., Ramis G., Bosse M., Perez-Enciso M., Crooijmans R.P. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. *BMC Genet.* 2013;14(1):106. DOI 10.1186/1471-2156-14-106.

Hill W.G. Estimation of effective population size from data on linkage disequilibrium. *Genet. Res*. 1981;38(3):209-216.

Kipen V.N., Ryabtseva A.O., Kotova S.A., Tsybovskiy I.S. Modern approaches to the differentiation of the European wild boar (*Sus scrofa scrofa*) and its domestic variety (*Sus scrofa domesticus*) using molecular genetic methods. *Voprosy Kriminologii, Kriminalistiki i Sudebnoy Ekspertizy = Questions of Criminology, Forensic Science and Forensic Examination*. 2016; 2(40):127-134. (in Russian)

Krupa E., Žáková E., Krupová Z. Evaluation of inbreeding and genetic variability of five pig breeds in Czech Republic. *Asian-Australas. J. Anim. Sci.* 2015;28(1):25-36. DOI 10.5713/ajas.14.0251.

Lorenzini R., Fanelli R., Tancredi F., Siclari A., Garofalo L. Matching STR and SNP genotyping to discriminate between wild boar, domestic pigs and their recent hybrids for forensic purposes. *Sci. Rep.* 2020;10:3188. DOI 10.1038/s41598-020-59644-6.

Lugovoy S.I., Kharzinova V.R., Kramarenko S.S., Lykhach A.V., Kramarenko A.S., Lykhach V.Ya. Genetic polymorphism of microsatellite loci and their association with reproductive traits in Ukrainian meat breed pigs. *Cytol. Genet.* 2018;52(5):360-367. DOI 10.3103/S0095452718050079.

Menendez J., Alvarez I., Fernandez I., Menendez-Arias N.A., Goyache F. Assessing performance of single-sample molecular genetic methods to estimate effective population size: empirical evidence from the endangered Gochu Asturcelta pig breed. *Ecol. Evol.* 2016;6(14):4971-4980. DOI 10.1002/ece3.2240.

Mihalik B., Frank K., Astuti P.K., Szemethy D., Szendrei L., Szemethy L., Kusza S., Stéger V. Population genetic structure of the wild boar (*Sus scrofa*) in the Carpathian Basin. *Genes*. 2020;11:1194. DOI 10.3390/genes11101194.

Nikolov I.S., Gum B., Markov G., Kuehn R. Population genetic structure of wild boar *Sus scrofa* in Bulgaria as revealed by microsatellite analysis. *Acta Theriol.* 2009;54(3):193-205. DOI 10.4098/j.at.0001-7051.049.2008.

Nomura T. Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evol. Appl.* 2008;1(3):462-474. DOI 10.1111/j.1752-4571.2008.00015.x.

Peakall R., Smouse P.E. GenAIEx V6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*. 2006;6(1):288-295.

Rębała K., Rabtsava A.A., Kotova S.A., Kipen V.N., Zhurina N.V., Gandzha A.I., Tsybovsky I.S. STR profiling for discrimination between wild and domestic swine specimens and between main breeds of domestic pigs reared in Belarus. *PLoS One*. 2016:11(11): e0166563. DOI 10.1371/journal.pone.0166563.

Ryabtseva A.O., Tsybovskiy I.S., Kotova S.A. Microsatellite markers in the study of polymorphism of wild boar (*Sus scrofa*) and domestic pig (*Sus scrofa domesticus*) inhabiting the territory of the Republic of Belarus. *Molekulyarnaya i Prikladnaya Genetika = Molecular and Applied Genetics*. 2018;25:56-66. (in Russian)

Sahoo N.R., Naskar S., Banik S., Pankaj P.K. Microsatellite based diversity analysis of native pigs of North-Eastern India. *Indian J. Anim. Res*. 2016;50(6):831-838. DOI 10.18805/ijar.7086.

Silva P.V.C., Lui J.F., Band G.O., Regitano L.C.A., Grossi S.F., Sollero B.P., Nunes C.S. Genetic variability among the wild boars (*Sus scrofa scrofa*), crossbred animals and pigs using microsatellite markers (STRs). *Braz. Arch. Biol. Technol*. 2011;54(2):301-306.

Snegin E.A., Kramarenko A.S., Artemchuk O.Y., Kramarenko S.S. Intra- and interbreed genetic heterogeneity and divergence in four commercial pig breeds based on microsatellite markers. *Regul. Mech. Biosyst.* 2021;12(1):128-135. DOI 10.15421/022120.

Szemethy D., Mihalik B., Frank K., Nagy T., Újváry D., Kusza S., Szemethy L., Barta E., Stéger V. Development of wild boar species-specific DNA markers for a potential quality control and traceability method in meat products. *Food Anal. Methods*. 2021;14:18-27. DOI 10.1007/s12161-020-01840-1.

Šveistienė R., Razmaitė V. Animal genetic resources in Lithuania. *Slovak J. Anim. Sci.* 2013;46(4):131-136.

Vernesi C., Crestanello B., Pecchioli E., Tartari D., Caramelli D., Hauffe H., Bertorelle G. The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*): a microsatellite analysis. *Mol. Ecol.* 2003;12(3):585-595. DOI 10.1046/j.1365-294x.2003.01763.x.

Waples R.S. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv. Genet.* 2006;7(2):167-184. DOI 10.1007/s10592-005-9100-y.

Waples R.S., Do C. Linkage disequilibrium estimates of contemporary *Ne* using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evol. Appl.* 2010; 3(3):244-262. DOI 10.1111/j.1752-4571.2009.00104.x.

Watterson G. The homozygosity test of neutrality. *Genetics*. 1978;88: 405-417.

Yeh F.C., Yang R., Boyle T. POPGENE Version 1.32: Microsoft Window-Based Freeware for Population Genetic Analysis. Edmonton: University of Alberta, Center of International Forestry Research, 1999.

Zanella R., Peixoto J.O., Cardoso F.F., Cardoso L.L., Biegelmeyer P., Cantao M.E., Otaviano A., Freitas M.S., Caetano A.R., Ledur M.C. Genetic diversity analysis of two commercial breeds of pigs using genomic and pedigree data. *Genet. Sel. Evol.* 2016;48(24). DOI 10.1186/s12711-016-0203-3.

Zinovieva N.A., Kostyunina O.V., Ekonomov A.V., Shevnina M.S., Domskij E.A., Gladyr E.A., Brem G. Polymorphism of genes associated with the quantitative trait loci in wild boar (*Sus scrofa* L., 1758) in Russia. *Selskokhozyaystvennaya Biologiya = Agricultural Biology.* 2013;2:77-82. DOI 10.15389/agrobiology.2013.2.77rus (in Russian)

ORCID ID

E.A. Snegin orcid.org/0000-0002-7574-6910
V.M. Makeeva orcid.org/0000-0002-4360-5371
A.P. Kaledin orcid.org/0000-0002-1769-5043
A.M. Ostapchuk orcid.org/0000-0002-9202-8611
I.D. Alazneli orcid.org/0000-0001-9305-8030
A.V. Smurov orcid.org/0000-0001-5143-1634

Conflict of interest. The authors declare no conflict of interest.

Received March 21, 2021. Revised September 18, 2021. Accepted September 27, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):831-838
DOI 10.18699/VJ21.096

Original Russian text www.bionet.nsc.ru/vogis/

Assessment of genetic diversity and phylogenetic relationships in Black Pied cattle in the Novosibirsk Region using microsatellite markers

R.B. Aitnazarov, T.M. Mishakova, N.S. Yudin ✉

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ yudin@bionet.nsc.ru

Abstract. There are currently over a thousand indigenous cattle breeds well adapted to local habitat conditions thanks to their long history of evolution and breeding. Identification of the genetic variations controlling the adaptation of local cattle breeds for their further introduction into the genome of highly productive global breeds is a matter of great relevance. Studying individual populations of the same breed with the use of microsatellite markers makes it possible to assess their genetic diversity, relationships, and breed improvement potential. Although the Black Pied breed is the most common dairy cattle breed in Russia, there are only a few studies on genetic diversity in local Black Pied populations in some Russian regions. The goal of the present study was to analyze the genetic diversity in Black Pied cattle populations in the Novosibirsk Region and compare them with other Russian populations; to identify significantly divergent populations with a view to preserving them under the programs aimed at maintaining the genetic diversity of the domestic Black Pied breed. DNA samples from 4788 animals of the Black Pied breed from six breeding enterprises in the Novosibirsk Region have been studied using 11 microsatellite markers. No significant differences in genetic variability parameters were found between individual populations. Private alleles have been identified in five out of six populations. Five populations have shown inbreeding coefficient values (F_{IS}) below zero, which indicates heterozygosity excess. The population distribution test, principal component analysis, F_{ST} and D_{EST} values, cluster analysis, and phylogenetic analysis have revealed two populations genetically distinct from the others. Essentially, the genetic diversity parameters of the six studied Black Pied cattle populations from the Novosibirsk Region show no significant differences from other Russian populations of the breed. Excess heterozygosity is observed in most breeding enterprises, which is a sign of a low inbreeding rate. To maintain the genetic diversity of the Russian Black Pied cattle, we recommend focusing on the two populations with significant genetic distinctions from the others.
Key words: cattle; Black Pied breed; Novosibirsk Region; microsatellite; genetic diversity; diversity preservation.

For citation: Aitnazarov R.B., Mishakova T.M., Yudin N.S. Assessment of genetic diversity and phylogenetic relationships in Black Pied cattle in the Novosibirsk Region using microsatellite markers. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8):831-838. DOI 10.18699/VJ21.096

Оценка генетического разнообразия и филогенетических отношений черно-пестрого скота Новосибирской области с использованием микросателлитных маркеров

Р.Б. Айтназаров, Т.М. Мишакова, Н.С. Юдин ✉

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ yudin@bionet.nsc.ru

Аннотация. В настоящее время известно более 1000 аборигенных пород крупного рогатого скота, которые хорошо приспособлены к местным условиям среды благодаря длительной адаптации и селекции. Крайне актуально выявление генетических вариантов, контролирующих адаптацию местного скота, для переноса этих вариантов в геномы высокопродуктивных глобальных пород. Исследования отдельных популяций внутри одной породы с помощью микросателлитных маркеров позволяют оценить их генетическое разнообразие, родственные взаимоотношения и перспективы их использования для улучшения породы. Черно-пестрая порода – наиболее массовая порода крупного рогатого скота молочного направления на территории России. Однако имеются лишь единичные работы, посвященные изучению генетического разнообразия местных популяций этой породы в отдельных областях России. Целью работы являются: анализ генетического разнообразия популяций черно-пестрого скота Новосибирской области и их сравнение с другими российскими популяциями; идентификация популяций, существенно отличающихся от всех остальных, для их дальнейшего использования в программах по сохранению генетического разнообразия отечественной черно-пестрой породы. Образцы ДНК от 4788 животных черно-пестрой породы из шести племенных хозяйств Новосибирской области

были исследованы по 11 микросателлитным маркерам. Значения всех показателей генетической изменчивости достоверно не различались между отдельными популяциями. Приватные аллели были обнаружены в пяти популяциях из шести. В пяти популяциях коэффициент инбридинга F_{IS} был достоверно ниже нуля, что говорит об избытке гетерозигот. Результаты теста распределения по популяциям, анализа методом главных компонент, анализа показателей F_{ST} и D_{EST}, а также кластерного и филогенетического анализов свидетельствуют о генетической обособленности двух популяций от всех остальных. Таким образом, параметры генетического разнообразия исследованных нами шести популяций черно-пестрого скота Новосибирской области существенно не отличаются от других российских популяций данной породы. В большинстве этих хозяйств наблюдается избыток гетерозигот, что говорит о низкой степени инбридинга. При разработке мероприятий, направленных на сохранение генетического разнообразия отечественного черно-пестрого скота, мы рекомендуем использовать животных из двух популяций, которые по генетическим характеристикам существенно отличаются от других.
Ключевые слова: крупный рогатый скот; черно-пестрая порода; Новосибирская область; микросателлит; генетическое разнообразие; сохранение биоразнообразия.

Introduction

There are currently over a thousand indigenous cattle breeds well adapted to local habitat conditions thanks to their long history of evolution and breeding (Buchanan, Lenstra, 2015). All these breeds are of high economic, scientific, historical, and cultural value (Stolpovskiy, Zakharov-Gezekhus, 2017). Meanwhile, we can see a global economically-driven replacement of local breeds by several high-productivity global breeds (Stolpovskiy, 2013). However, these breeds are typically poorly adapted to local habitats and are thus unable to reveal their outstanding qualities (Mokhov, Shabalina, 2011). As a result, identification of the genetic variations controlling the adaptation of local cattle breeds for their further introduction into the genome of highly productive global breeds is a matter of great relevance (Madan, 2005). Eventually, it will make the development of new breeds combining great productivity traits with adaptability to various geographical regions possible. For example, the H100Q mutation in gene *NRAP* discovered recently in Yakutian cattle seems to affect its adaptation to extreme cold (Buggiotti et al., 2021). This approach has become increasingly effective since the promising CRISPR/Cas genome editing technology was introduced into animal husbandry (Bevacqua et al., 2016; Ikeda et al., 2017).

The Black Pied breed is the most common dairy cattle in Russia (Breeds and Types of Farm Animals…, 2013). Intense breeding efforts involving the four approved Black Pied cattle types (Irmen, Priobsky, Krasnoyarsk, and Pribaikalsky) well adapted to extreme climatic conditions and local feeds are currently ongoing in Siberia (Klimenok et al., 2014). Composite cross-breeding of Black Pied cows with Holstein breeding bulls in 12 breeding enterprises in Western and Eastern Siberia has recently produced Sibiryachka, a brand new high-productivity dairy breed (Yarantseva et al., 2019).

Highly polymorphic microsatellite loci have been widely used as genetic markers in population and conservation genetics for relationship identification and other purposes (Guichoux et al., 2011; Städele, Vigilant, 2016; Galinskaya et al., 2019). In particular, microsatellites are used to analyze the origin and phylogenetic relationships of local cattle breeds (Olschewsky, Hinrichs, 2021). Studying individual populations of the same breed makes it possible to assess their genetic diversity, relationships, and breed improvement potential (Zsolnai et al., 2014; Agung et al., 2016; Szucs et al., 2019). However, studies on genetic diversity in local Black Pied breed populations are very few (Smaragdov, 2018; Modorov et al., 2021); this is especially true for the Novosibirsk Region, which remains poorly studied in this regard.

The goals of the present study were: to analyze the genetic diversity of six Black Pied cattle populations from the Novosibirsk Region and to compare them with other Russian populations; to identify significantly divergent populations to be preserved under the programs aimed at maintaining the genetic diversity of the Russian Black Pied breed.

Materials and methods

Blood samples were taken from 4788 Black Pied cows and bulls from six breeding enterprises located in the Novosibirsk Region (referred to below as populations A–F). To analyze the populations' structure and phylogenetic relationships, the Holstein cattle breed was used as a control group (referred to below as HOL) (van de Goor et al., 2011).

The total DNA was isolated using the COrDIS SPRINT reagent (Gordiz, Moscow, Russia) as per the manufacturer's instructions. PCRs for 11 microsatellite markers (ETH3, INRA023, TGLA227, TGLA126, TGLA122, SPS115, ETH225, BM2113, BM1824, ETH10, BM1818) were performed using the COrDIS Cattle kit (Gordiz, Moscow, Russia) as per the manufacturer's protocol. Fragment analysis of amplified DNA was carried out using an automated genetic analyzer NANOPHORE-05 (Syntol, Moscow, Russia). The sizes of microsatellite DNA markers were calculated in GeneMapper Software 5 (Thermo Fisher Scientific, USA).

The genetic diversity parameters, *F*-statistics, population distribution test, and the significance of genotype distribution deviation from expected Hardy–Weinberg equilibrium (HWE) were calculated using the GenAlEx 6.5 software

Table 1. Genetic variability parameters of the microsatellite loci in the Black Pied cattle of the Novosibirsk Region (N = 4788)

Locus	N_A	N_E	H_O	H_E	Private alleles*
BM1818	7.000 ± 0.447	2.658 ± 0.040	0.648 ± 0.011	0.623 ± 0.005	–
BM1824	5.833 ± 0.307	3.321 ± 0.259	0.726 ± 0.021	0.690 ± 0.024	185
BM2113	7.500 ± 0.500	4.212 ± 0.087	0.808 ± 0.009	0.762 ± 0.005	153
ETH3	7.333 ± 0.333	3.047 ± 0.124	0.733 ± 0.023	0.669 ± 0.013	–
ETH10	8.000 ± 0.258	3.905 ± 0.129	0.777 ± 0.009	0.743 ± 0.009	–
ETH225	6.667 ± 0.333	3.031 ± 0.151	0.681 ± 0.017	0.666 ± 0.015	–
TGLA122	12.833 ± 1.195	5.760 ± 0.165	0.879 ± 0.020	0.826 ± 0.005	139, 145, 179
TGLA126	5.333 ± 0.211	2.348 ± 0.045	0.569 ± 0.006	0.573 ± 0.008	113, 125
TGLA227	9.667 ± 0.211	5.324 ± 0.262	0.864 ± 0.010	0.810 ± 0.009	95, 101
INRA023	7.167 ± 0.477	3.876 ± 0.098	0.776 ± 0.010	0.741 ± 0.006	198, 216
SPS115	6.333 ± 0.211	2.160 ± 0.058	0.554 ± 0.017	0.535 ± 0.013	–
Mean	7.606 ± 0.281	3.604 ± 0.143	0.729 ± 0.013	0.694 ± 0.011	–

Note. Here and elsewhere, the scores are given as M ± m, where M is the arithmetical mean; m is the standard error; N_A is the average number of alleles per locus; N_E is the number of effective alleles per locus; H_O is the observed heterozygosity; H_E is the expected heterozygosity; * is the unique alleles typical for a certain population.

(Peakall, Smouse, 2012). The allelic richness (AR) was assessed via the rarefaction algorithm in HP-Rare software (Kalinowski, 2005). Calculation of the pairwise F_{ST} values and significance check of the nonzero F_{IS} values were performed using the bootstrap method adjusted for multiple comparisons in the FSTAT software (Goudet, 2003). Cluster analysis was carried out in the STRUCTURE software (Hubisz et al., 2009). The significance of differences between populations was analyzed using Student's *t*-test or one-way ANOVA with post hoc Bonferroni correction in Statistica 8.0.

The phylogenetic tree was built using the UPGMA approach based on Nei's genetic distances in the POPTREE2 software (Takezaki et al., 2010). The statistical reliability of the phylogenetic tree was analyzed using bootstrap values based on 1000 permutations (Szucs et al., 2019). The confidence threshold was set at 70 (Lukashov, 2009).

Results

The results of genetic variability analysis for the Black Pied cattle populations of the Novosibirsk Region can be seen in Table 1. All the microsatellite loci turned out to be highly polymorphic and contained 105 alleles in general. The average number of alleles per locus was 7.606, and the effective number – 3.604. The observed heterozygosity (0.729) was statistically similar to the expected one (0.694).

The pairwise comparison of genetic differences for breeding enterprises, performed using Fischer's exact test in the Genepop software, demonstrated that the cattle of each enterprise could be considered as a separate population statistically different from the others (Supplementary Material 1)[1]. The genetic variability parameters for each of the populations can be seen in Table 2. The maximum number of alleles per locus (8.455) was observed in population A, and the minimum one (6.273) – in population B. The allele enrichment and the effective number of alleles between populations varied between 6.087 (C) – 6.863 (F) and 3.437 (B) 3.873 (D), respectively. The observed and expected heterozygosities varied from 0.701 (F) to 0.755 (B) and from 0.682 (C) to 0.714 (D), respectively. The values of all the above indicators did not significantly differ between individual populations.

The private alleles, i. e. the unique alleles characteristic for a particular animal population, were found in five of the six populations. In populations A–E, the inbreeding coefficient F_{IS} was statistically below zero. These were the populations where in particular loci statistically significant genotype deviations from HWE were observed (Suppl. Material 2). The highest number of such loci (six) was spotted in populations A and D. Meanwhile, the genotype distribution of the *ETH225* and *TGLA126* loci matched HWE in all the populations investigated.

Analysis of the results of a population distribution test demonstrated that, on average, 45.7 % of the animals had been properly assigned (Suppl. Material 3). However, for population B this score reached 70.8 %, which evidences

[1] Supplementary Materials 1–4 are available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/Suppl_Aitnazarov_Engl.pdf

Table 2. Genetic variability parameters of microsatellite loci in some populations of Black Pied cattle of the Novosibirsk Region

Index	Population					
	A	B	C	D	E	F
Number of animals	2408	65	1065	630	459	161
Number of alleles (N_A)	8.455 ± 0.835	6.273 ± 0.524	8.000 ± 1.000	7.818 ± 0.536	7.455 ± 0.562	7.636 ± 0.472
Allele enrichment (AR)	6.452 ± 0.422	6.273 ± 0.524	6.087 ± 0.474	6.650 ± 0.409	6.450 ± 0.393	6.863 ± 0.438
Number of effective alleles (N_E)	3.611 ± 0.368	3.437 ± 0.340	3.457 ± 0.350	3.873 ± 0.375	3.587 ± 0.325	3.658 ± 0.408
Number of private alleles	4	0	3	1	2	1
Number of animals with one or more private alleles	6	0	7	1	5	1
Observed heterozygosity (H_O)	0.724 ± 0.035	0.755 ± 0.035	0.706 ± 0.035	0.749 ± 0.034	0.737 ± 0.032	0.701 ± 0.030
Expected heterozygosity (H_E)	0.692 ± 0.032	0.683 ± 0.027	0.682 ± 0.030	0.714 ± 0.030	0.699 ± 0.026	0.696 ± 0.029
Inbreeding coefficient (F_{IS})	–0.046 ± 0.011*	–0.103 ± 0.016*	–0.032 ± 0.010*	–0.047 ± 0.010*	–0.052 ± 0.016*	–0.007 ± 0.014

* Fixation index ($p < 0.05$) is statistically different from zero.

this population being significantly different from the others. Principle component analysis (PCA) of the F_{ST} values showed that population B was significantly different from the others in the first component reflecting 36.73 % of the genetic variability of the whole dataset (Fig. 1). As for the second-component distribution responsible for 27.61 % of genetic variability, it was most prominent for population D.

The highest degree of genetic differentiation, both for Jost's differentiation and the F_{ST} fixation indices, was between populations B and D (Suppl. Material 4). The closest populations in this respect were A and C.

The results of genetic clustering in STRUCTURE demonstrated that at $k = 2$, the population of Black Pied and HOL breeds was distributed between two different clusters (Table 3), where HOL had the highest values of similarity coefficient Q in one of the clusters. The Q values for all the Black Pied populations (except for D) were statistically lower than those for the HOL animals.

In the UPGMA phylogenetic tree built using the Nei distances, populations B and D belonged to different branches, which was statistically confirmed (Fig. 2). All the other populations including the control HOL population formed a single cluster.

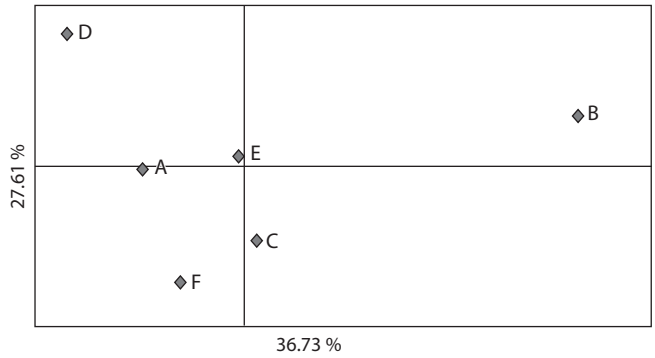


Fig. 1. Principle component analysis results of the fixation index F_{ST} values.

The axes indicate the degree of explained dispersion.

Table 3. Genetic clustering results for the Black Pied (A–F) and Holstein (HOL) populations

Population	Similarity coefficient Q	*N*
A	0.456 ± 0.004*	2408
B	0.507 ± 0.019*	65
C	0.481 ± 0.005*	1065
D	0.604 ± 0.007	630
E	0.549 ± 0.008*	459
F	0.551 ± 0.013*	161
HOL	0.622 ± 0.009	254

Note. Similarity coefficient Q (Pritchard et al., 2000) was calculated for $k = 2$ (Q1 and Q2). Data for the HOL population were the courtesy of van de Goor et al. (2011); * is $p < 0.001$ for Student's pairwise comparison against the HOL population.

Discussion

Analysis of 11 microsatellite loci from the whole sample of Black Pied cattle of the Novosibirsk Region revealed 105 alleles, which is lower than the number obtained after investigating 13 224 Holsteinized Black Pied animals in the Sverdlovsk Region (Modorov et al., 2021). The 15 loci that included the microsatellites investigated in our study contained 164 alleles, but the frequency of 38 of them did not exceed 0.1 %. On the other hand, a study of 36 animals

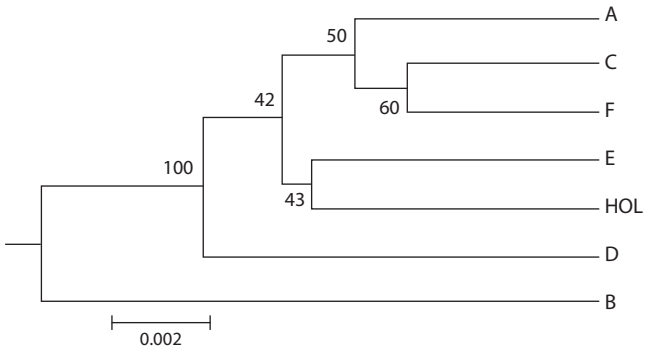


Fig. 2. UPGMA phylogenetic tree to reveal the genetic relations between the Black Pied (A–F) (collected data) and the HOL populations (van de Goor et al., 2011).

The nodes indicate the bootstrap values.

from Poland produced just 76 alleles for a similar set of loci (10 out of 11 markers matched) (Radko et al., 2005). Thus, the observed differences may be related to the sample size and/or the number of microsatellite loci used.

The *TGLA122* locus was characterized by the highest average number of alleles per locus (12.833). A similar score for this locus (14 alleles) was obtained in a study investigating the Black Pied breed from the Pskov Region (Arzhankova et al., 2015). The lowest average number of alleles per locus (5.333) was found in *TGLA126*, which correlates with the analogous parameter in the Black Pied breed from the Sverdlov Region (7 alleles, the frequency of 2 of them does not exceed 0.1 %) (Modorov et al., 2021). The highest (5.760) and lowest (2.160) numbers of effective alleles were detected in the *TGLA122* and *SPS115* loci, which also correlates with the results obtained by M.V. Modorov et al. (2021). The values of observed and expected heterozygosity (0.729 and 0.694) obtained for our sample were similar to those for the Black Pied cattle from the Sverdlovsk Region (0.73 and 0.72) (Modorov et al., 2021) but lower than the numbers for the pedigree bulls of the same breed (0.779 and 0.751) (Zinovieva et al., 2015).

It is known that the genetic data of 25–30 randomly selected animals from a population are sufficient for reliable estimation of the population's allele frequency, expected heterozygosity and genetic distances (Hale et al., 2012). In our study, the sampling size significantly exceeded the mentioned threshold. The results of Fischer's exact test demonstrated that all the six samples investigated could be regarded as separate populations (see Suppl. Material 1), which enabled us to shift to a more detailed analysis of their genetic differences.

Such parameters as the number of effective alleles, allele enrichment, observed/expected heterozygosity are widely used to estimate genetic variations between populations since they do not depend on a sampling size (Leberg et al., 2002; Galinskaya et al., 2019). In our study, these parameters did not have statistically significant differences between any population pairs investigated (see Table 2), which may be since all the considered breeding enterprises rely upon semen production from the same sources.

In this respect, our results are in good correlation with those of M.V. Modorov et al. (2021) who investigated 29 herds of Holsteinized Black Pied cattle from the Sverdlovsk Region and found no statistically significant genetic differences for 27 of them. Unfortunately, using microsatellite markers within a single breed for cattle population monitoring, the authors, as a rule, ignore statistical methods when comparing the genetic variability parameters (Glinskaya, 2013; Kuznetsov, 2014; Zsolnai et al., 2014; Agung et al., 2016; Szucs et al., 2019). In our study, private alleles were found in all populations, except population B, which is probably due to the size of the population ($N = 65$). In this respect, the Black Pied cattle from the Novosibirsk Region significantly outmatched the Black Pied animals from the Republic of Belarus, where private alleles were detected only in three populations out of nine (Glinskaya, 2013).

Inbreeding coefficient F_{IS} is known to indicate heterozygosity reduction due to nonrandom coupling (Kuznetsov, 2014). At $F_{IS} > 0$, there is a deficiency of heterozygous individuals (inbreeding); while at $F_{IS} < 0$, such individuals are in excess (outbreeding). At $F_{IS} = 0$, mating becomes HWE-random. In our study for most of the populations (A–E), the inbreeding coefficient was significantly below zero, meaning the heterozygotes were excessive. Consequently, populations A–E demonstrated statistically significant deviations from HWE in some of the locus genotypes (see Suppl. Material 2), which is a good correlation with the result presented above. The most probable reason for this effect might be implementation of a mating system (outbreeding; disassortative mating, etc.) aimed to reduce inbreeding (Kuznetsov, 2014). At the same time, such factors as population's finite size, nonrandom mating, selection effect, etc. can not be completely excluded (Galinskaya, 2019).

In the population distribution test performed in our study, on average 45.7 % of animals were correctly distributed in their original groups (see Suppl. Material 3), which matched well with the 48 % distribution in a study of 16 herds of the Limousin breed in Hungary (Szucs et al., 2019). However, for population B, the distribution parameter was 70.8 %, which evidenced this population being significantly different from the others.

The results of subpopulation fixation index (F_{ST}) PCA analysis demonstrated that populations A, C, E and F formed a compact group (see Fig. 1), for which the F_{ST} values varied from 0.004 to 0.008 (see Suppl. Material 4). Populations B and D are further from this group for the first and second components, respectively. The longest genetic distance, as per fixation index, was observed between populations B and D and comprised 0.013. The obtained genetic distances range was in good correlation with the data for single nucleotide polymorphisms (SNPs)

obtained with an Illumina BovineSNP50 chip assay for the Holsteinized Black Pied cattle of six breeding enterprises in the Leningrad Region (0.002–0.012) (Smaragdov, 2018) and the populations of Jersey cattle in the USA, Canada and the UK (0.006–0.016) (Cooper et al., 2016).

According to S. Write's classification, genetic differentiation is considered insignificant, if F_{ST} does not exceed 0.05 (Wright, 1978). However, V.M. Kuznetsov states that it is a differentiation of less than 0.01 that can be regarded as 'insignificant and negligible', so interpretation of the above-mentioned results can be a complex issue (Kuznetsov, 2020). Nevertheless, T.V. Galinskaya et al. assume that interpretation of the F_{ST} value is more complex than just referring to the mentioned authors (Galinskaya et al., 2019). In their opinion 'what is more important is whether we could detect a statistically significant genetic differentiation ($F_{ST} > 0$) or not'.

The permutation test in our study demonstrated that all the obtained F_{ST} values were statistically valid ($p < 0.01$) (see Suppl. Material 4), which confirms the genetic isolation of populations B and D.

Although F_{ST} is widely used to assess genetic population differentiation, its application for multiallelic and multilocus markers such as microsatellites is often criticized (Meirmans, Hedrick, 2011; Kuznetsov, 2021). For these markers, several alternative statistical methods have been suggested such as Jost's differentiation index (D_{EST}), which accounts for changes in effective number of alleles (Jost et al., 2018). F_{ST} and D_{EST} are believed to complement each other and be applied jointly (Meirmans, Hedrick, 2011; Kuznetsov, 2021). In our study, the D_{EST} distances statistically correlated with the F_{ST} estimations ($r = 0.92$, $p < 0.0001$). For both parameters, populations B and D were genetically the most distant.

Cluster analysis revealed that the Black Pied and Holstein (HOL) populations were distributed between two different clusters (see Table 3), which confirms their genetic affinity (Yurchenko et al., 2018; Yudin, Larkin, 2019). The similarity coefficient values for all the populations except D turned out to be much lower than that of HOL, which evidences different HOL pedigree levels in the investigated Black Pied populations (Zinovieva et al., 2015).

Phylogenetic analysis distributed the Black Pied populations into three groups (see Fig. 2). One group included populations A, C, E and F that were close to HOL. Populations B and D formed two independent statistically verified branches. The result confirmed the genetic insulation of populations B and D, which we also confirmed with the results of population distribution test, PCA, F_{ST}/D_{EST} index analysis and the results of cluster analysis presented above.

It is generally believed that to preserve a breed as a selection material, one has to sustain all its genetic pool because in most cases it is unknown which particular genes and their combination determine the economic characteristics of the breed (Stolpovskiy, 2013; Stolpovskiy, Zakharov-Gezekhus, 2017). According to the authors, the purpose of a biodiversity preservation program is to 'sustain the diversity of the alleles a species (breed) has as well as to support the process of accumulation and potential preservation of newly appearing mutant alleles as an important source of constant evolution and improvement in animals'.

The results of the tests performed in our study confirm the genetic insulation of populations B and D from the other populations investigated, so these two populations, above all else, have to be used to preserve the genetic pool of the Black Pied breed.

Conclusion

Thus, the genetic variability parameters of the six populations of Black Pied cattle from the Novosibirsk Region have had no significant differences from other Russian populations of this breed. Most of the breeding enterprises involved in the study have heterozygote excess due to low-level inbreeding. Our recommendation to those developing the programs aimed at preserving the genetic diversity of the Russian Black Pied cattle is to use animals of two populations, the genetic characteristics of which differ significantly from all others.

References

Agung P.P., Saputra F., Septian W.A., Lusiana Zein M.S., Sulandari S., Anwar S., Wulandari A.S., Said S., Tappa B. Study of genetic diversity among Simmental cross cattle in West Sumatra based on microsatellite markers. *Asian-Australas. J. Anim. Sci.* 2016;29(2):176-183. DOI 10.5713/ajas.15.0155.

Arzhankova Yu.V., Mosachikhina I.A., Kharitonov A.V. Genetic features of Black Pied and cross-breeding cattle by microsatellite loci. *Izvestiya Velikolukskoy Gosudarstvennoy Sel'skokhozyaystvennoy Akademii = Bulletin of the Velikie Luki State Agricultural Academy.* 2015;1:7-11. (in Russian)

Bevacqua R.J., Fernandez-Martín R., Savy V., Canel N.G., Gismondi M.I., Kues W.A., Carlson D.F., Fahrenkrug S.C., Niemann H., Taboga O.A., Ferraris S., Salamone D.F. Efficient edition of the bovine PRNP prion gene in somatic cells and IVF embryos using the CRISPR/Cas9 system. *Theriogenology.* 2016;86(8):1886-1896.e1. DOI 10.1016/j.theriogenology.2016.06.010.

Breeds and Types of Farm Animals in the Russian Federation. Dictionary of Terms Concerning the Industry, Genetics, Breeding, and Reproduction Technology of Farm Animals. List of Russian and International Organizations in the Field of Animal Husbandry. Moscow: All-Russia Research Institute of Animal Breeding, 2013. (in Russian)

Buchanan D.S., Lenstra J.A. Breeds of cattle. In: Garrick D.J., Ruvinsky A. (Eds.) The Genetics of Cattle. 2nd ed. CABI: Wallingford, UK, 2015;33-66. DOI 10.1079/9781780642215.0033.

Buggiotti L., Yurchenko A.A., Yudin N.S., Vander Jagt C.J., Vorobieva N.V., Kusliy M.A., Vasiliev S.K., Rodionov A.N., Boronetskaya O.I., Zinovieva N.A., Graphodatsky A.S., Daetwyler H.D., Larkin D.M. Demographic history, adaptation, and NRAP convergent evolution at amino acid residue 100 in the world northernmost cattle from Siberia. *Mol. Biol. Evol.* 2021;38(8):3093-3110. DOI 10.1093/molbev/msab078.

Cooper T.A., Eaglen S.A.E., Wiggans G.R., Jenko J., Huson H.J., Morrice D.R., Bichard M., Luff W.G.L., Woolliams J.A. Genomic evaluation, breed identification, and population structure of Guernsey cattle in North America, Great Britain, and the Isle of Guernsey. *J. Dairy Sci.* 2016;99(7):5508-5515. DOI 10.3168/jds.2015-10445.

Galinskaya T.V., Schepetov D.M., Lysenkov S.N. Prejudices against microsatellite studies and how to resist them. *Russ. J. Genet.* 2019; 55(6):657-671. DOI 10.1134/S1022795419060048.

Glinskaya N.A. The STR analysis of genetic differentiation in Black-Pied cattle populations bred in Belarus. *Vestnik Polesskogo Gosudarstvennogo Universiteta = Bulletin of the Polessky State University.* 2013;2:21-26. (in Russian).

Goudet J. Fstat (ver. 2.9.4). a program to estimate and test population genetics parameters. 2003. Available from: http://www.unil.ch/izea/softwares/fstat.html.

Guichoux E., Lagache L., Wagner S., Chaumeil P., Léger P., Lepais O., Lepoittevin C., Malausa T., Revardel E., Salin F., Petit R.J. Current trends in microsatellite genotyping. *Mol. Ecol. Resour.* 2011;11(4):591-611. DOI 10.1111/j.1755-0998.2011.03014.x.

Hale M.L., Burg T.M., Steeves T.E. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *PLoS One.* 2012; 7(9):e45170. DOI 10.1371/journal.pone.0045170.

Hubisz M.J., Falush D., Stephens M., Pritchard J.K. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* 2009;9(5):1322-1332. DOI 10.1111/j.1755-0998.2009.02591.x.

Ikeda M., Matsuyama S., Akagi S., Ohkoshi K., Nakamura S., Minabe S., Kimura K., Hosoe M. Correction of a disease mutation using CRISPR/Cas9-assisted genome editing in Japanese Black cattle. *Sci. Rep.* 2017;7(1):17827. DOI 10.1038/s41598-017-17968-w.

Jost L., Archer F., Flanagan S., Gaggiotti O., Hoban S., Latch E. Differentiation measures for conservation genetics. *Evol. Appl.* 2018; 11(7):1139-1148. DOI 10.1111/eva.12590.

Kalinowski S.T. HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Mol. Ecol. Notes*. 2005; 5(1):187-189. DOI 10.1111/j.1471-8286.2004.00845.x.

Klimenok I.I., Yaranceva S.B., Shishkina M.A. Productive and breeding qualities of black and motley cattle Siberian region. *Genetika i Razvedeniye Zhivotnykh = Genetics and Breeding of Animals.* 2014;2:30-33. (in Russian)

Kuznetsov V.M. Wright's F-statistics: evaluation and interpretation. *Problemy Biologii Productivnykh Zhivotnykh = Problems of Productive Animal Biology*. 2014;4:80-104. (in Russian)

Kuznetsov V.M. Nei's methods for analyzing genetic differences between populations. *Problemy Biologii Productivnykh Zhivotnykh = Problems of Productive Animal Biology.* 2020;1:91-110. DOI 10.25687/1996-6733.prodanimbiol.2020.1.91-110. (in Russian)

Kuznetsov V.M. Assessment of genetic differentiation of populations by analysis of molecular variance (analytical review). *Agrarnaya Nauka Evro-Severo-Vostoka = Agricultural Science Euro-North-East.* 2021;22(2):167-187. DOI 10.30766/2072-9081.2020.21.2.167-187. (in Russian)

Leberg P.L. Estimating allelic richness: effects of sample size and bottlenecks. *Mol. Ecol.* 2002;11(11):2445-2449. DOI 10.1046/j.1365-294x.2002.01612.x.

Lukashov V.V. Molecular Evolution and Phylogenetic Analysis. Moscow: Binom Publ., 2009. (in Russian)

Madan M.L. Animal biotechnology: applications and economic implications in developing countries. *Rev. Sci. Tech.* 2005;24(1): 127-139. DOI 10.20506/rst.24.1.1555.

Meirmans P.G., Hedrick P.W. Assessing population structure: F_{ST} and related measures. *Mol. Ecol. Resour.* 2011;11(1):5-18. DOI 10.1111/j.1755-0998.2010.02927.x.

Modorov M.V., Tkachenko I.V., Grin A.A., Sevost'yanov M.Yu., Zezin N.N. Population genetic structure of Ural Black Pied cattle. *Russ. J. Genet.* 2021;57(4):453-459. DOI 10.1134/S1022795421040104.

Mokhov B.P., Shabalina E.P. Influence of heredity and ecogenesis for adaptation and productivity of dairy cows. *Izvestiya Ul'yanovskoy Gosudarstvennoy Sel'skokhozyaystvennoy Akademii = Bulletin of the Ulyanovsk State Agricultural Academy.* 2011;2(14):90-96. (in Russian)

Olschewsky A., Hinrichs D. An overview of the use of genotyping techniques for assessing genetic diversity in local farm animal breeds. *Animals (Basel).* 2021;11(7):2016. DOI 10.3390/ani11072016.

Peakall R., Smouse P.E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics.* 2012;28(19):2537-2539. DOI 10.1093/bioinformatics/bts460.

Pritchard J.K., Stephens M., Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2): 945-959.

Radko A., Zyga A., Zabek T., Słota E. Genetic variability among Polish Red, Hereford and Holstein-Friesian cattle raised in Poland based on analysis of microsatellite DNA sequences. *J. Appl. Genet.* 2005;46(1):89-91.

Smaragdov M.G. Genome wide assessment of inter-herd genetic differences of cattle. *Dostizheniya Nauki i Tekhniki APK = Achievements of Science and Technology of AIC.* 2018;32(4):47-49. DOI 10.24411/0235-2451-2018-10411. (in Russian)

Städele V., Vigilant L. Strategies for determining kinship in wild populations using genetic data. *Ecol. Evol*. 2016;6(17):6107-6120. DOI 10.1002/ece3.2346.

Stolpovskiy Yu.A. Population genetics studies underlying preservation of domesticated animal species gene pools. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2013;17(4/2):900-915. (in Russian)

Stolpovskiy Yu.A., Zakharov-Gezekhus I.A. The problem of conservation of gene pools of domesticated animals. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2017;21(4):477-486. DOI 10.18699/VJ17.266. (in Russian)

Szucs M., Szabó F., Bán B., Józsa C., Rózsa L., Zsolnai A., Anton I. Assessment of genetic diversity and phylogenetic relationship of Limousin herds in Hungary using microsatellite markers. *Asian-Australas. J. Anim. Sci.* 2019;32(2):176-182. DOI 10.5713/ajas.18.0164.

Takezaki N., Nei M., Tamura K. POPTREEW: web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities. *Mol. Biol. Evol.* 2014; 31(6):1622-1624. DOI 10.1093/molbev/msu093.

van de Goor L.H., Koskinen M.T., van Haeringen W.A. Population studies of 16 bovine STR loci for forensic purposes. *Int. J. Legal. Med.* 2011;125(1):111-119. DOI 10.1007/s00414-009-0353-8.

Wright S. Evolution and the Genetics of Population. Vol. 4. Variability within and Among Natural Populations. Chicago: Univ. Chicago Press, 1978.

Yarantseva S.B., Gerasimchuk L.D., Shishkina M.A. New breed of dairy cattle Sibiryachka. *Sibirskii Vestnik Sel'skokhozyaistvennoy Nauki = Siberian Herald of Agricultural Science*. 2019;49(6): 62-70. DOI 10.26898/0370-8799-2019-6-7. (in Russian)

Yudin N.S., Larkin D.M. Whole genome studies of origin, selection and adaptation of the Russian cattle breeds. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2019;23(5):559-568. DOI 10.18699/VJ19.525. (in Russian)

Yurchenko A., Yudin N., Aitnazarov R., Plyusnina A., Brukhin V., Soloshenko V., Lhasaranov B., Popov R., Paronyan I.A., Plemyashov K.V., Larkin D.M. Genome-wide genotyping uncovers genetic profiles and history of the Russian cattle breeds. *Heredity.* 2018;120(2):125-137. DOI 10.1038/s41437-017-0024-3.

Zinovieva N.A., Gladyr E.A., Bagirov V.A., Brem G. Dynamics of the biodiversity of Black and White cattle influenced by cross-breeding. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2015;19(2):222-225. (in Russian).

Zsolnai A., Kovács A., Anton I., Rátky J., Brüssow K.P., Józsa C., Bán B., Gyurmán A. Comparison of different Hungarian Grey herds as based on microsatellite analysis. *Anim. Sci. Pap. Rep.* 2014; 32(2):121-130.

ORCID ID

R.B. Aitnazarov orcid.org/0000-0001-6929-7648
N.S. Yudin orcid.org/0000-0002-1947-5554

Acknowledgements. This work was supported by State Budgeted Project 0259-2021-0015.

Conflict of interest. The authors declare no conflict of interest.

Received September 6, 2021. Revised October 17, 2021. Accepted October 17, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):839-846
DOI 10.18699/VJ21.097

Original Russian text www.bionet.nsc.ru/vogis/

The role of the *KIBRA* and *APOE* genes in developing spatial abilities in humans

A.V. Kazantseva[1, 2] ✉, R.F. Enikeeva[1], Yu.D. Davydova[1], R.N. Mustafin[3], Z.R. Takhirova[4], S.B. Malykh[5, 6], M.M. Lobaskova[5], T.N. Tikhomirova[5, 6], E.K. Khusnutdinova[1, 6]

[1] Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
[2] Ufa State Petroleum Technological University, Department of molecular technologies, Ufa, Russia
[3] Bashkir State Medical University, Department of medical genetics and fundamental medicine, Ufa, Russia
[4] Bashkir State University, Department of genetics and fundamental medicine, Ufa, Russia
[5] Psychological Institute of the Russian Academy of Education, Moscow, Russia
[6] Lomonosov Moscow State University, Department of psychology, Russia
✉ Kazantsa@mail.ru

Abstract. In the contemporary high-tech society, spatial abilities predict individual life and professional success, especially in the STEM (Science, Technology, Engineering, and Mathematics) disciplines. According to neurobiological hypotheses, individual differences in cognitive abilities may be attributed to the functioning of genes involved in the regulation of neurogenesis and synaptic plasticity. In addition, genome-wide association studies identified rs17070145 located in the *KIBRA* gene, which was associated with individual differences in episodic memory. Considering a significant role of genetic and environmental components in cognitive functioning, the present study aimed to estimate the main effect of *NGF* (rs6330), *NRXN1* (rs1045881, rs4971648), *KIBRA* (rs17070145), *NRG1* (rs6994992), *BDNF* (rs6265), *GRIN2B* (rs3764030), *APOE* (rs7412, rs429358), and *SNAP25* (rs363050) gene polymorphisms and to assess the effect of gene-environment interactions on individual differences in spatial ability in individuals without cognitive decline aged 18–25 years (N = 1011, 80 % women). Spatial abilities were measured using a battery of cognitive tests including the assessment of "3D shape rotation" (mental rotation). Multiple regression analysis, which was carried out in the total sample controlling for sex, ethnicity and the presence of the "risk" *APOE* ε4 allele, demonstrated the association of the rs17070145 T-allele in the *KIBRA* gene with enhanced spatial ability (β = 1.32; p_{FDR} = 0.037) compared to carriers of the rs17070145 CC-genotype. The analysis of gene-environment interactions revealed that nicotine smoking (β = 3.74; p = 0.010) and urban/rural residency in childhood (β = –6.94; p = 0.0002) modulated the association of *KIBRA* rs17070145 and *APOE* (rs7412, rs429358) gene variants with individual differences in mental rotation, respectively. The data obtained confirm the effect of the *KIBRA* rs17070145 T-allele on improved cognitive functioning and for the first time evidence the association of the mentioned genetic variant with spatial abilities in humans. A "protective" effect of the *APOE* ε2 allele on enhanced cognitive functioning is observed only under certain conditions related to childhood rearing.
Key words: *KIBRA*; *APOE*; cognitive abilities; mental rotation; linear regression; gene-environment interactions.

For citation: Kazantseva A.V., Enikeeva R.F., Davydova Yu.D., Mustafin R.N., Takhirova Z.R., Malykh S.B., Lobaskova M.M., Tikhomirova T.N., Khusnutdinova E.K. The role of the *KIBRA* and *APOE* genes in developing spatial abilities in humans. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8):839-846. DOI 10.18699/VJ21.097

Вовлеченность генов *KIBRA* и *APOE* в формирование особенностей пространственного мышления человека

А.В. Казанцева[1, 2] ✉, Р.Ф. Еникеева[1], Ю.Д. Давыдова[1], Р.Н. Мустафин[3], З.Р. Тахирова[4], С.Б. Малых[5, 6], М.М. Лобаскова[5], Т.Н. Тихомирова[5, 6], Э.К. Хуснутдинова[1, 6]

[1] Институт биохимии и генетики – обособленное структурное подразделение Уфимского федерального исследовательского центра Российской академии наук, Уфа, Россия
[2] Уфимский государственный нефтяной технический университет, кафедра молекулярных технологий, Уфа, Россия
[3] Башкирский государственный медицинский университет, кафедра медицинской генетики и фундаментальной медицины, Уфа, Россия
[4] Башкирский государственный университет, кафедра генетики и фундаментальной медицины, Уфа, Россия
[5] Психологический институт Российской академии образования, Москва, Россия
[6] Московский государственный университет им. М.В. Ломоносова, факультет психологии, Москва, Россия
✉ Kazantsa@mail.ru

Аннотация. В современном высокотехнологичном обществе пространственные способности являются предиктором успешности в жизни и профессиональной деятельности, особенно в STEM дисциплинах (*от англ.* Science, Technology, Engineering, and Mathematics). Согласно нейробиологическим гипотезам, существование индивидуальных различий в когнитивных способностях может быть обусловлено особенностями функцио-

нирования генов, участвующих в регуляции нейрогенеза и синаптической пластичности. С другой стороны, полногеномный анализ ассоциаций идентифицировал rs17070145 в гене *KIBRA*, ассоциированный с индивидуальными различиями в эпизодической памяти. Учитывая важную роль генетической и средовой компоненты в манифестации когнитивных функций, целью настоящего исследования являлись оценка основного эффекта полиморфных вариантов генов *NGF* (rs6330), *NRXN1* (rs1045881, rs4971648), *KIBRA* (rs17070145), *NRG1* (rs6994992), *BDNF* (rs6265), *GRIN2B* (rs3764030), *APOE* (rs7412, rs429358), *SNAP25* (rs363050) и оценка ген-средовых взаимодействий в формировании индивидуальных особенностей пространственного мышления у индивидов без когнитивных нарушений 18–25 лет (*N* = 1011, 80 % женщин). Измерение уровня пространственных способностей осуществлялось с помощью батареи тестовых заданий на вращение 3D фигур (shape rotation). Множественный линейный регрессионный анализ, проведенный в общей выборке с включением половой, этнической принадлежности и наличия «рискового» аллеля *APOE* ε4 в качестве ковариат, продемонстрировал ассоциацию аллеля rs17070145*T в гене *KIBRA* с более высоким уровнем пространственного мышления (β = 1.32; p_{FDR} = 0.037) по сравнению с носителями генотипа rs17070145*CC. Анализ ген-средовых взаимодействий выявил, что табакокурение (β = 3.74; *p* = 0.010) и место воспитания в детстве (β = –6.94; *p* = 0.0002) модулируют ассоциацию полиморфных вариантов в гене *KIBRA* (rs17070145) и гене *APOE* (rs7412, rs429358) с индивидуальными различиями в пространственных способностях соответственно. Полученные результаты подтверждают связь аллеля rs17070145*T в гене *KIBRA* с улучшением когнитивных функций и впервые свидетельствуют об ассоциации данного генетического варианта с особенностями пространственного мышления. «Протективный» эффект аллеля *APOE* ε2 на улучшение когнитивного функционирования наблюдается только при сочетании определенных особенностей воспитания в детстве.

Ключевые слова: *KIBRA*; *APOE*; когнитивные способности; мысленное вращение предметов; линейная регрессия; ген-средовые взаимодействия.

Introduction

The study of the productivity of cognitive functions as an integral part of individual potential is becoming increasingly relevant today since the level of cognitive functioning is the basis of individual life success and self-actualization. In particular, in a modern high-tech society, spatial abilities (i. e., ability for 3D mental rotation) predict success in life and professional activity, especially in STEM (Science, Technology, Engineering, and Mathematics) disciplines (Nagy-Kondor, 2017). The hypotheses of an individual trajectory of spatial ability existing to date suggest a significant role of genetic, epigenetic and environmental factors (Mustafin et al., 2020; Takhirova et al., 2021). According to twin research, the impact of the genetic component on individual variance in this trait varies within 64–84 %, depending on the type of examined spatial ability (Malanchini et al., 2020).

According to neurobiological hypotheses, individual differences in cognitive abilities may be due to the specificity of gene functioning involved in the regulation of neurogenesis and synaptic plasticity in such brain regions as prefrontal cortex and hippocampus (Mustafin et al., 2020). The latter process represents the development of neuronal connections as a response to novel experiences. An important role in the regulation of this process belongs to neurotrophic factors (*BDNF*, *NGF*), neurexins (*NRXN1*), neuregulin (*NRG1*), synaptosomal-associated protein (*SNAP25*), glutamatergic receptor (*GRIN2B*) (Enikeeva et al., 2017; Mustafin et al., 2020). One of the most significant and replicating results obtained in the studies of cognitive functioning is the association of the *APOE* ε4 allelic variant with an increased risk of developing Alzheimer's disease and higher rate of cognitive decline (Porter et al., 2018; Li X. et al., 2019). Previous attempts were made to evaluate the effect of different variants of the genes involved in neurogenesis (*APOE*, *TOMM40*, *BDNF*, *SORL1*, and *CLSTN2*) on cognitive changes in non-demented individuals above 65–70 years (Laukka et al., 2020). Considering that about 60 % of variance in age-related cognitive changes correlates with different cognitive domains (episodic and semantic memory, information processing speed, nonverbal intelligence, spatial ability, etc.) (Tucker-Drob et al., 2019), it can be assumed that allelic variants of genes, which encode neurogenesis-involved proteins, can also attribute to differences in spatial abilities.

Together with the candidate gene approach, a significant contribution in the study of complex traits is related to such methodological approach as genome-wide association analysis (GWAS), which made it possible to identify genetic variants involved in the regulation of cognitive functioning. The rs17070145 located in intron 9 of the *KIBRA* (KIdney and BRAin expressed protein) gene, which was initially identified in the GWAS of episodic memory in cohorts from Sweden and America (Papassotiropoulos et al., 2006), represents one of the loci associated with cognitive functioning. Subsequent studies confirmed the association of the minor T-allele with improved episodic memory (Porter et al., 2018) and spatial learning (Schuck et al., 2013). A recent meta-analysis based on 20 case-control studies confirmed the association of the rs17070145 C-allele with an increased risk of developing Alzheimer's disease and cognitive decline among aged individuals (Ling et al., 2018). It is known that the *KIBRA* gene (also known as *WWC1*, WW domain-containing protein 1) encodes a signal transduction protein, which is widely expressed in the kidney and brain regions related to memory regulation (hippocampus, prefrontal cortex, cerebellum and hypothalamus). It is involved in multiple cellular functions, including cell migration, vesicular transport, transcription, synaptogenesis, neuronal signaling, and has a neuroprotective effect, thus inhibiting Aβ-induced apoptosis (Heitz et al., 2016). Moreover, from a functional point of view, reduced Kibra level was shown to mediate memory and synaptic plasticity decrease (Heitz et al., 2016). It should be noted that cognitive functioning can be mediated by an additive and epistatic interaction

of proteins encoded by the *APOE* and *KIBRA* genes (Wang et al., 2019), which indicates the requirement of simultaneous analysis of both genes.

To date, it remains unknown whether rs17070145 in the *KIBRA* gene is involved in the regulation of other cognitive abilities (including spatial ability) in individuals of younger age. Therefore, considering that the T-allele in the *KIBRA* gene was associated with improved memory and executive functions in the majority of studies in individuals without cognitive deficit and was related to better functioning of prefrontal cortex and hippocampus (Papassotiropoulos et al., 2006; Zhang et al., 2009), we suggest that a similar relationship may be observed with enhanced spatial ability in mentally healthy individuals.

Together with a genetic component, individual variance in spatial abilities may be attributed to a specific micro- and macro-environment in ontogenetic development, including sex (Lauer et al., 2019). In this regard, the present study aimed to: (1) estimate the main effect of polymorphic variants of genes involved in neurogenesis and synaptic plasticity; (2) estimate the effect of gene-by-environment interactions on individual differences in spatial ability in individuals without cognitive impairments.

Materials and methods

The study sample included 1011 mentally healthy young adults (80 % women) – students at Universities of the Republic of Bashkortostan and the Udmurt Republic (mean age 19.79 ± 1.69 years), consisted of Russians – 535, Tatars – 231, Udmurts – 160, and individuals of mixed ethnicity – 85. All volunteers had no individual and familial history of psyhopathologies.

The assessment of spatial abilities was conducted in 2017–2019 via the battery of cognitive tests, which estimated the number of correct answers on the items related to "shape rotation" and were implemented online in the psychodiagnostic platform designed by the Russian Academy of Education. All enrolled individuals filled the inventory consisting of questions on ethnicity by three generations together with several social parameters such as the specificity of child-parent relationship (a style of parental rearing, maltreatment in childhood, rearing in a full/incomplete family), family income level, maternal and paternal age at individual's birth, place of residence in childhood, sibship size and birth order, bilingual rearing, the presence of chronic disorders and smoking. Place of residence in childhood (urban/rural residency) was determined on the basis of its population size according to (Kazantseva et al., 2020a): demographic locations with population size under 50,000 individuals were determined as a rural residency. All the volunteers filled informed consents for the participation in the study. The study was approved by the Bioethical Committee at the Institute of Biochemistry and Genetics UFRC RAS (Ufa, Russia).

Collection of biological material was performed within 2017–2019 followed by DNA isolation from the peripheral tissue leukocytes. Genotyping of 10 SNPs in the *NGF* (rs6330), *NRXN1* (rs1045881, rs4971648), *KIBRA* (rs17070145), *NRG1* (rs6994992), *BDNF* (rs6265), *GRIN2B* (rs3764030), *APOE* (rs7412, rs429358), and *SNAP25* (rs363050) genes was carried out via real-time PCR using KASP kits (LGC Genomics, UK) with CFX96 DNA Analyzer (BioRad, USA) and end-point fluorescence analysis. The *APOE* genotypes were grouped based on the presence of ε2, ε3, ε4 allelic variants.

Quantitative data were tested for the correspondence to the Gaussian distribution via Shapiro–Wilk *W*-test ($p > 0.05$). The main effect assessment was performed via multiple linear regression analysis. Various statistically significant models (additive, dominant, recessive) were analyzed with PLINK v.1.09, while sex, ethnicity, presence/absence of *APOE* ε4 allele were included as independent variables (covariates) together with the genotypes (formula (1)). In addition, to analyze gene-by-environment interactions examined socio-demographic parameters and genotypes were included in linear regression models as independent variables according to formula (2):

$$Y_{(G)} = k + \beta_1 COV_1 + \beta_2 COV_2 + \beta_3 COV_3 + \beta_4 x_4, \quad (1)$$

$$Y_{(G\times E)} = k + \beta_1 COV_1 + \beta_2 COV_2 + \beta_3 COV_3 + \beta_4 x_4 + \beta_5 COV_5 + \beta_6 x_4 COV_5, \quad (2)$$

where Y – spatial ability score; k – intercept; $\beta_{1, \ldots, 6}$ – regression coefficients; COV_1 – sex; COV_2 – ethnicity; COV_3 – presence/absence of *APOE* ε4 variant; x_4 – the presence of minor allele of examined SNP in dominant model (the number of copies of minor allele for the additive model); COV_5 – environmental predictor; $x_4 COV_5$ – allele-by-environment interaction.

For statistically significant gene-by-environment interaction model, a stratification analysis between the groups split by either environmental predictor or genetic component was conducted (SPSS 23.0). Correction for multiple comparisons was carried out via FDR procedure (PLINK v.1.09).

Results

In the present study, allele and genotype frequencies distribution corresponded to the Hardy–Weinberg equilibrium (see the Table). Subsequent multiple regression analysis, which was performed in the total sample controlling for sex, ethnicity and the presence of "risky" *APOE* ε4 allele, demonstrated the association of *KIBRA* rs17070145 T-allele with enhanced spatial ability ($\beta = 1.32$; $\beta_{ST} = 0.10$; $p = 0.003$; $p_{FDR} = 0.037$; $r^2 = 0.007$) compared to carriers of rs17070145 CC-genotype (in additive model, see the Table).

Liner regression models, which were applied separately to men, women, *APOE* ε4 allele carriers/non-carriers, failed to observe a statistically significant effect of the examined loci after correction for multiple comparisons ($p_{FDR} > 0.05$, see the Table). Mean spatial ability scores depending on the *KIBRA* rs17070145 genotype in the total sample, as well as in men, women, *APOE* ε4 allele carriers/non-carriers are shown in Fig. 1.

As a result of gene-by-environment interactions analysis, which considered both the effect of genetic variants and various social parameters, smoking was revealed to modulate association of *KIBRA* rs17070145 with individual differences in spatial ability ($\beta = 3.74$; $\beta_{ST} = 0.14$; $p = 0.010$). To clarify the effect of smoking on cognitive abilities we conducted strati-

Examined SNPs, the Hardy–Weinberg equilibrium test and the results of linear regression analysis of SNPs association with spatial ability (under additive model) in the total sample and in subgroups

Gene	SNP	Alleles[a]	MAF	p_{HWE}	Total sample		Women		Men		*APOE* ε4+		*APOE* ε4–	
					β_{ST}	*p*	β_{ST}	*p*	β_{ST}	*p*	β_{ST}	*p*	β_{ST}	*p*
NGF	rs6330	A/G	0.400	0.056	–0.01	0.582	0.01	0.677	–0.14	0.056	0.01	0.846	–0.01	0.835
NRXN1	rs1045881	T/C	0.176	0.446	–0.04	0.216	–0.05	0.200	–0.01	0.868	–0.08	0.213	–0.01	0.851
	rs4971648	C/T	0.223	0.053	0.01	0.813	0.02	0.648	–0.03	0.711	–0.05	0.467	0.01	0.715
NRG1	rs6994992	T/C	0.428	0.697	–0.05	0.168	–0.04	0.288	–0.08	0.327	–0.15	0.027[f]	–0.02	0.604
BDNF	rs6265	A/G	0.146	0.248	0.01	0.890	0.01	0.857	0.01	0.949	0.07	0.325	–0.01	0.790
GRIN2B	rs3764030	T/C	0.224	0.359	0.08	0.019[b]	0.07	0.079	0.13	0.083	0.13	0.049	0.06	0.137
APOE	rs7412	ε4+	0.134	0.825	–0.12	0.337	–0.32	0.036[d]	0.29	0.211	–	–	–	–
	rs429358	ε2+	0.079	0.359	–0.02	0.468	–0.02	0.608	–0.07	0.399	–	–	–	–
SNAP25	rs363050	G/A	0.442	0.897	0.01	0.801	0.02	0.679	–0.02	0.789	0.01	0.914	0.01	0.875
KIBRA	rs17070145	T/C	0.431	0.055	0.10	**0.003**[c]	0.09	0.022[e]	0.15	0.057	0.15	0.033[g]	0.08	0.036[h]

Note. MAF – minor allele frequency; p_{HWE} – *p*-value for the Hardy–Weinberg test; β_{ST} – standardized regression coefficient; *p* – *p*-value for the Wald test. Statistically significant differences (after FDR-correction) are shown in bold. [a] minor/major alleles; [b] $p_{FDR} = 0.098$; [c] $p_{FDR} = 0.037$; [d] $p_{FDR} = 0.183$; [e] $p_{FDR} = 0.183$; [f] $p_{FDR} = 0.164$; [g] $p_{FDR} = 0.164$; [h] $p_{FDR} = 0.368$.

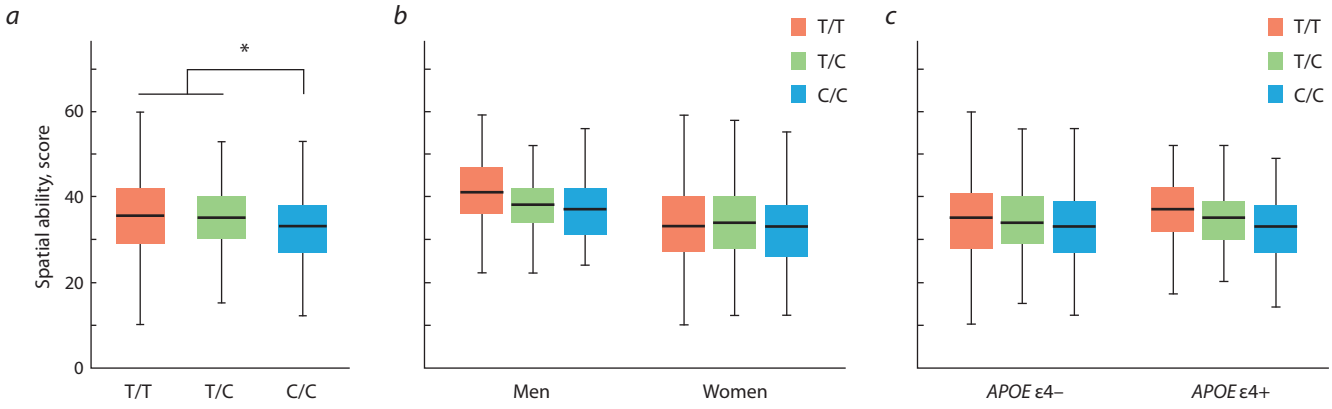


Fig. 1. Mean spatial ability scores depending on the *KIBRA* rs17070145 genotype in the total sample (*a*), in men, women (*b*) and the carriers/non-carriers of *APOE* ε4 variant (*c*).

Statistically significant differences in the mean values of spatial ability between the groups are marked with brackets, $^{*}p_{FDR} < 0.05$.

fication analysis, which demonstrated that enhanced spatial ability was characteristic for ever-smoking rs17070145 T-allele carriers compared to non-smoking individuals ($\beta = 4.59$; $\beta_{ST} = 0.22$; $r^2 = 0.003$; $p_{FDR} = 0.004$) (Fig. 2, *a*). Moreover, the model with inclusion of *APOE* variants and place of residence in childhood was also statistically significant ($\beta = –6.94$; $\beta_{ST} = –0.23$; $p = 0.0002$). In addition, higher spatial ability was demonstrated in the carriers of "favorable" *APOE* ε2 allele, who indicated their rural residency, compared to those from the urban regions ($\beta = –6.04$; $\beta_{ST} = –0.25$; $r^2 = 0.06$; $p_{FDR} = 0.021$) (see Fig. 2, *b*).

Discussion

Since previous research indicated the necessity to control for the well-known "risk" factor of developing cognitive deficit (*APOE* ε4 variant) in the statistical models (Porter et al., 2018; Li X. et al., 2019), the hypothesis suggested in the present study was examined in both total sample and in the groups split by the presence of *APOE* ε4 allele. Previous findings evidence that individuals without cognitive decline carrying *APOE* ε4 variant (related to the accumulation of amyloid beta (Aβ)) demonstrated an increased decline in the verbal episodic memory and hippocampal hypotrophy in the presence of *KIBRA* rs17070145 CC-genotype compared to minor T-allele carriers (Porter et al., 2018). In the present study, the analysis of individuals aged 18–25 years without cognitive impairments failed to detect significant effect of *APOE* ε4 allele on the association of *KIBRA* gene variants with the level of spatial abilities. Nevertheless, we included the mentioned "risk" allele in the *APOE* gene as an independent variable

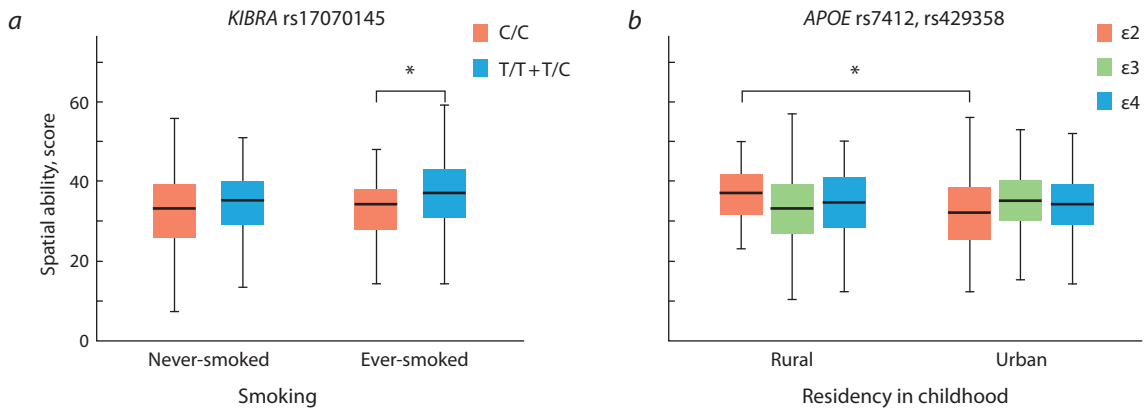


Fig. 2. The results of gene-by-environment interaction analysis, which demonstrated a modulating effect of (*a*) smoking on association of *KIBRA* rs17070145 variants on spatial ability; (*b*) place of residence in childhood on association of *APOE* variants with spatial ability.

Statistically significant differences in the mean values of spatial ability between the groups are marked with brackets, * $p_{FDR} < 0.05$.

in multiple regression models. As a result of these analyses, for the first time we demonstrated a positive effect of *KIBRA* rs17070145 T-allele on higher spatial ability in individuals without cognitive deficit, which at some extent is congruent to findings obtained by other research groups in non-demented healthy individuals (Schuck et al., 2013; Porter et al., 2018).

The effect observed is confirmed by functional studies, which reported that rs17070145 in the *KIBRA* gene was related to grey matter volume in the prefrontal cortex and parahippocampal gyrus in elderly individuals (Li R. et al., 2020). In particular, as a result of voxel-oriented morphometry grey matter volume was diminished in carriers of rs17070145 C-allele compared to individuals with rs17070145 TT-genotype in both elderly participants (Li R. et al., 2020), and younger-age volunteers (Wang et al., 2013), which, in turn, reflects an improved cognitive functioning in carriers of minor T-allele. Interestingly, young-age individuals with C-allele, which is associated with reduced grey matter volume, demonstrated a compensatory effect via enhanced synchronization between the brain regions involved in the regulation of executive control (Wang et al., 2013). The results obtained by our group can be explained by an increased level of long-term potentiation (LTP) in hippocampus and associated enlarged cognitive functioning related to an increased expression of the *KIBRA* gene (Heitz et al., 2016), which can be due to the presence of rs17070145 T-allele. From the other side, rs17070145 may be in a linkage disequilibrium with other functional variants (resulting in missence mutations such as rs3822660G/T or M734I, rs3822659T/G or S735A), which are located in the exon 15 of the *KIBRA* gene and mediate differential Ca^{2+}-dependent binding of protein C2-domain with phosphatidylinositol, therefore regulating cellular pathways (Duning et al., 2013).

The researchers suggest that controversy of the impact of *KIBRA* rs17070145 in published findings may be related to the cognitive status of the examined sample, as well as to demographic parameters including age (Zhang et al., 2009; Li X. et al., 2019). Therefore, we analyzed different regression models, which consisted of various environmental predictors. One of the interesting findings of the present study is the effect of smoking, which was shown to modulate the association of allelic variant in the *KIBRA* gene with spatial ability. Another research group succeeded to identify that reduced number of constant mistakes in cognitive tests was observed in smoking individuals from European populations compared to never-smoking individuals, but this association was prominent only in carriers of rs17070145 T-allele in the *KIBRA* gene (Zhang et al., 2009). Notably, in the present study the inclusion of interaction terms (*KIBRA* rs17070145 genotype and smoking) in the linear regression model also revealed the association of minor T-allele with higher shape rotation ability, which was characteristic for ever-smoking individuals compared to never-smoking ones. Therefore, it was suggested that nicotine might positively affect cognitive abilities (including executive functions and attention) in individuals with T-allele (Zhang et al., 2009). According to our previous research, nicotine may have a modulating effect on genetic association with individual cognitive and psychological domains in mentally healthy individuals (Davydova et al., 2020), which may be caused by nicotine-related changes in epigenetic profile in the examined genes.

Although multiple studies evidence the association of *APOE* "risky" ε4 allele with cognitive decline and Alzheimer's disease, diminished grey matter volume in hippocampus (Porter et al., 2018; Li X. et al., 2019), and lower spatial abilities (Laczó et al., 2020), we failed to identify the main effect of *APOE* gene variants on individual variance in spatial ability in mentally healthy individuals without cognitive decline. Previously, the attempts to estimate a combined effect of *APOE* ε4 allele and environmental predictors (smoking, physical activity, overweight, education level) on cognitive domains in individuals aged 40–79 years have been performed, which failed to detect statistically significant models of gene-by-environment interactions (Rodriguez et al., 2018). Nevertheless, the analysis of gene-by-environment interactions conducted by our group made it possible to observe the involvement of *APOE* gene variants in mental rotation ability depending on the place of residence in childhood (rural/urban). The highest

spatial ability level was characteristic for individuals bearing "favorable" *APOE* ε2 allele, who indicated rural residency, compared to urban-residency participants. Accordingly, based on the data obtained it can be assumed that unfavorable effect of urban residency in childhood is even observed in the case of presence of "favorable" *APOE* ε2 allele related to enhanced neuronal activity (Davis et al., 2020).

Published data from other research groups also indicated a correlation between absent cognitive decline in elderly individuals and the presence of "green" territorial neighborhood in childhood; moreover, this effect was characteristic for individuals without *APOE* ε4 allele (Cherrie et al., 2018). Interestingly, the presence of available "green" neighborhood positively affected memory and attention in school-aged children even during one year (Dadvand et al., 2015), while long-lasting accommodation in the "green" neighborhood correlated with an enhanced grey matter volume in the prefrontal cortex, thus, explaining improved cognitive functioning (Dadvand et al., 2018). This observation can be explained by several items. First, urban residency is related to higher level of ecopollutants and xenobiotics, which results in impaired regulation of various neurotransmitter systems in the brain (Dadvand et al., 2018). Second, urban/rural residency results in the differences in the lateralization of functions, specificity of language development and visual-spatial processes (Polyakov, 2008). In particular, rural residency is characterized by forced development of right-hemispheric brain structures, whereas urban school-aged children demonstrated predominant development of left-hemispheric functional systems, which reflects the specificity of their cognitive ability. Third, a positive effect of rural rearing on cognitive domains may be related to the nutrition specificity, although only in the absence of "unfavorable" *APOE* ε4 allele, which was published previously (Martínez-Lapiscina et al., 2014). From the one side, the consumption of eco-friendly available farmer nutrition products by children of rural residency and higher level of their physical activity (including their help to older family members in the gardens) on the other hand may provide the "manifestation" of a positive effect of *APOE* ε2 allele on spatial ability. Fourth, as a consequence of rural residency, children may obtain a "favorable" gut microbiota content and diversity, which directly affects brain development via gut-brain axis based on the recent research findings (Mancabelli et al., 2017).

On the other hand, population urbanization is accompanied by development of mental and cognitive impairments due to a diminished exposure to macro- and microorganisms, thus resulting in disturbed immunoregulation. In turn, it may result in increased inflammatory response of the body on psychological stressors related to residency in a high-tech society compared to small territorial units (Rook et al., 2013). In addition, the impact of place of residency on cognitive functioning may be attributed to enhanced stress level characteristic for urban residency and correlating with cortisol level. One of the studies demonstrated the interaction between cortisol level and the presence of "risky" *APOE* ε4 allele, which caused a decline in cognitive functioning (Lee et al., 2008).

Therefore, the demonstrated effect of gene-by-environment interactions, which is related to the spatial ability development, observed in the present study allows us to suggest that a positive effect of the *APOE* ε2 allele on cognitive abilities may be identified only under rural residency, which is presumably related to the favorable impact on neuronal processes.

Despite the possible epistatic or additive effect of interactions of the genes involved in neurogenesis and synaptic plasticity (Wang et al., 2019), we failed to observe either the main or epistatic effect of the *BDNF*, *NGF*, *NRXN1*, *NRG1*, *SNAP25*, and *GRIN2B* genes on interindividual differences in spatial ability. Other studies also revealed no association of neurogenesis-related gene variants (*APOE*, *SORL1*, *BDNF*, *TOMM40*, *KIBRA*, and *COMT*) with spatial abilities in individuals aged 40–60 years (Korthauer et al., 2018). According to our previous research, several genes involved in synaptic plasticity regulation such as the *SNAP25*, *NRXN1*, and *NRG1* are responsible for individual differences in such cognitive domains as mathematical abilities (Kazantseva et al., 2020b) and working memory volume (Enikeeva et al., 2017). Despite the suggestion proposed in the present study on the association of allelic variants in neurogenesis-involved genes in spatial ability, we failed to confirm such association.

Conclusion

As a result of the present cross-sectional study, the main effect of the *KIBRA* gene on the development of the spatial ability was observed in individuals without cognitive deficit; moreover, respondents' smoking positively affected the examined cognitive domain in carriers of rs17070145 minor T-allele. It should be noted that we confirmed a "protective" effect of *APOE* ε2 allele on improved cognitive functioning, which manifested only in the presence of such favorable factor as rural residency in childhood. The data obtained are congruent with previously claimed suggestions on the association of rs17070145 minor T-allele in the *KIBRA* gene with improved cognitive functioning and primarily evidence the involvement of this genetic variant in individual differences in spatial ability.

The present study has several advantages including a large sample size of similar-age individuals, control for sex, ethnicity and "risky" allele in the *APOE* gene in regression models (i. e. inclusion of mentioned predictors in multiple linear regression models). The examined sample was collected prior to COVID-19 pandemic, which allowed us to avoid the possible effect of SARS-CoV-2 on nervous system and cognitive functions, which has been repeatedly demonstrated (Fotuhi et al., 2020).

References

Cherrie M.P.C., Shortt N.K., Mitchell R.J., Taylor A.M., Redmond P., Thompson C.W., Starr J.M., Deary I.J., Pearce J.R. Green space and cognitive ageing: a retrospective life course analysis in the Lothian Birth Cohort 1936. *Soc. Sci. Med*. 2018;196:56-65. DOI 10.1016/j.socscimed.2017.10.038.

Dadvand P., Nieuwenhuijsen M.J., Esnaola M., Forns J., Basagaña X., Alvarez-Pedrerol M., Rivas I., López-Vicente M., De Castro Pascual M., Su J., Jerrett M., Querol X., Sunyer J. Green spaces and cognitive development in primary schoolchildren. *Proc. Natl. Acad. Sci. USA*. 2015;112(26):7937-7942. DOI 10.1073/pnas.1503402112.

Dadvand P., Pujol J., Macià D., Martínez-Vilavella G., Blanco-Hinojo L., Mortamais M., Alvarez-Pedrerol M., Fenoll R., Esnaola M., Dalmau-Bueno A., López-Vicente M., Basagaña X., Jerrett M., Nieuwenhuijsen M.J., Sunyer J. The association between lifelong greenspace exposure and 3-dimensional brain magnetic resonance imaging in Barcelona schoolchildren. *Environ. Health Perspect*. 2018;126(2):027012. DOI 10.1289/EHP1876.

Davis A.A., Inman C.E., Wargel Z.M., Dube U., Freeberg B.M., Galluppi A., Haines J.N., Dhavale D.D., Miller R., Choudhury F.A., Sullivan P.M., Cruchaga C., Perlmutter J.S., Ulrich J.D., Benitez B.A., Kotzbauer P.T., Holtzman D.M. APOE genotype regulates pathology and disease progression in synucleinopathy. *Sci. Transl. Med*. 2020;12(529):eaay3069. DOI 10.1126/scitranslmed.aay3069.

Davydova Yu.D., Kazantseva A.V., Enikeeva R.F., Mustafin R.N., Enikeeva R.F., Lobaskova M.M., Malykh S.B., Gilyazova I.R., Khusnutdinova E.K. The role of oxytocin receptor (*OXTR*) gene polymorphisms in the development of aggressive behavior in healthy individuals. *Russ. J. Genet.* 2020;56(9):1129-1138. DOI 10.1134/S1022795420090057.

Duning K., Wennmann D.O., Bokemeyer A., Reissner C., Wersching H., Thomas C., Buschert J., Guske K., Franzke V., Flöel A., Lohmann H., Knecht S., Brand S.M., Pöter M., Rescher U., Missler M., Seelheim P., Pröpper C., Boeckers T.M., Makuch L., Huganir R., Weide T., Brand E., Pavenstädt H., Kremerskothen J. Common exonic missense variants in the C2 domain of the human KIBRA protein modify lipid binding and cognitive performance. *Transl. Psychiatry*. 2013;3(6):e272. DOI 10.1038/tp.2013.49.

Enikeeva R.F., Lobaskova M.M., Kazantseva A.V., Romanova A.R., Karunas A.A., Malykh S.B., Khusnutdinova E.K., Tikhomirova T.N. The role of synaptic plasticity genes in individual differences in working memory. *Teoreticheskaya i Experimental'naya Psikhologiya = Theoretical and Experimental Psychology*. 2017;10(4):6-15. (in Russian)

Fotuhi M., Mian A., Meysami S., Raji C.A. Neurobiology of COVID-19. *J. Alzheimers Dis*. 2020;76(1):3-19. DOI 10.3233/JAD-200581.

Heitz F.D., Farinelli M., Mohanna S., Kahn M., Duning K., Frey M.C., Pavenstädt H., Mansuy I.M. The memory gene KIBRA is a bidirectional regulator of synaptic and structural plasticity in the adult brain. *Neurobiol. Learn. Mem*. 2016;135:100-114. DOI 10.1016/j.nlm.2016.07.028.

Kazantseva A.V., Davydova Yu.D., Enikeeva R.F., Lobaskova M.M., Mustafin R.N., Malykh S.B., Takhirova Z.R., Khusnutdinova E.K. AVPR1A main effect and OXTR-by-environment interplay in individual differences in depression level. *Heliyon*. 2020a;6(10):e05240. DOI 10.1016/j.heliyon.2020.e05240.

Kazantseva A.V., Enikeeva R.F., Romanova A.R., Malykh S.B., Galyautdinova S.I., Khusnutdinova E.K. Stress-associated cognitive functioning is controlled by variations in synaptic plasticity genes. *Russ. J. Genet*. 2020b;56(1):88-95. DOI 10.1134/S1022795420010068.

Korthauer L.E., Awe E., Frahmand M., Driscoll I. Genetic risk for age-related cognitive impairment does not predict cognitive performance in middle age. *J. Alzheimers Dis*. 2018;64(2):459-471. DOI 10.3233/JAD-171043.

Laczó J., Cechova K., Parizkova M., Lerch O., Andel R., Matoska V., Kaplan V., Matuskova V., Nedelska Z., Vyhnalek M., Hort J. The combined effect of *APOE* and *BDNF* Val66Met polymorphisms on spatial navigation in older adults. *J. Alzheimers Dis*. 2020;78(4):1473-1492. DOI 10.3233/JAD-200615.

Lauer J.E., Yhang E., Lourenco S.F. The development of gender differences in spatial reasoning: a meta-analytic review. *Psychol. Bull*. 2019;145(6):537-565. DOI 10.1037/bul0000191.

Laukka E.J., Köhncke Y., Papenberg G., Fratiglioni L., Bäckman L. Combined genetic influences on episodic memory decline in older adults without dementia. *Neuropsychology*. 2020;34(6):654-666. DOI 10.1037/neu0000637.

Lee B.K., Glass T.A., Wand G.S., McAtee M.J., Bandeen-Roche K., Bolla K.I., Schwartz B.S. Apolipoprotein E genotype, cortisol, and cognitive function in community-dwelling older adults. *Am. J. Psychiatry*. 2008;165(11):1456-1464. DOI 10.1176/appi.ajp.2008.07091532.

Li R., Wan W., Li J. KIBRA polymorphism modulates gray matter volume to influence cognitive ability in the elderly. *Brain Imaging Behav*. 2020;14(5):1388-1394. DOI 10.1007/s11682-019-00047-1.

Li X., Hildebrandt A., Sommer W., Wilhelm O., Reuter M., Montag C., Zhou C. Cognitive performance in young APOE ε4 carriers: a latent variable approach for assessing the genotype-phenotype relationship. *Behav. Genet*. 2019;49(5):455-468. DOI 10.1007/s10519-019-09961-y.

Ling J., Huang Y., Zhang L., Wei D., Cheng W. Association of KIBRA polymorphism with risk of Alzheimer's disease: evidence based on 20 case-control studies. *Neurosci. Lett.* 2018;662:77-83. DOI 10.1016/j.neulet.2017.08.057.

Malanchini M., Rimfeld K., Shakeshaft N.G., McMillan A., Schofield K.L., Rodic M., Rossi V., Kovas Y., Dale P.S., Tucker-Drob E.M., Plomin R. Evidence for a unitary structure of spatial cognition beyond general intelligence. *NPJ Sci. Learn*. 2020;5:9. DOI 10.1038/s41539-020-0067-8.

Mancabelli L., Milani C., Lugli G.A., Turroni F., Ferrario C., van Sinderen D., Ventura M. Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environ. Microbiol*. 2017;19(4):1379-1390. DOI 10.1111/1462-2920.13692.

Martínez-Lapiscina E.H., Galbete C., Corella D., Toledo E., Buil-Cosiales P., Salas-Salvado J., Ros E., Martinez-Gonzalez M.A. Genotype patterns at *CLU*, *CR1*, *PICALM* and *APOE*, cognition and Mediterranean diet: the PREDIMED-NAVARRA trial. *Genes Nutr*. 2014;9(3):393. DOI 10.1007/s12263-014-0393-7.

Mustafin R.N., Kazantseva A.V., Enikeeva R.F., Malykh S.B., Khusnutdinova E.K. Longitudinal genetic studies of cognitive characteristics. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(1):87-95. DOI 10.18699/VJ20.599.

Nagy-Kondor R. Spatial ability: measurement and development. In: Khine M.S. (Ed.) Visual-Spatial Ability in STEM Education: Transforming research into practice. Cham: Springer Int. Publ., 2017; 35-58. DOI 10.1007/978-3-319-44385-0_3.

Papassotiropoulos A., Stephan D.A., Huentelman M.J., Hoerndli F.J., Craig D.W., Pearson J.V., Huynh K.-D., Brunner F., Corneveaux J., Osborne D., Wollmer M.A., Aerni A., Coluccia D., Hänggi J., Mondadori C.R.A., Buchmann A., Reiman E.M., Caselli R.J., Henke K., de Quervain D.J.-F. Common *Kibra* alleles are associated with human memory performance. *Science*. 2006;314:475-478. DOI 10.1126/science.1129837.

Polyakov V.M. Development of higher mental processes in urban and rural children. *Kul'turno-istoritcheskaya Psikhologiya = Cultural-Historical Psychology*. 2008;4(1):9-16. (in Russian)

Porter T., Burnham S.C., Doré V., Savage G., Bourgeat P., Begemann K., Milicic L., Ames D., Bush A.I., Maruff P., Masters C.L., Rowe C.C., Rainey-Smith S., Martins R.N., Groth D., Verdile G., Villemagne V.L., Laws S.M. KIBRA is associated with accelerated cognitive decline and hippocampal atrophy in APOE ε4-positive cognitively normal adults with high Aβ-amyloid burden. *Sci. Rep*. 2018;8:2034. DOI 10.1038/s41598-018-20513-y.

Rodriguez F.S., Schroeter M.L., Arélin K., Witte A.V., Baber R., Burkhardt R., Engel C., Löffler M., Thiery J., Villringer A., Luck T., Riedel-Heller S.G. APOE e4-genotype and lifestyle interaction on cognitive performance: results of the LIFE-Adult-study. *Health Psychol*. 2018;37(2):194-205. DOI 10.1037/hea0000515.

Rook G.A., Lowry C.A., Raison C.L. Microbial 'Old Friends', immunoregulation and stress resilience. *Evol. Med. Public Health*. 2013; 2013(1):46-64. DOI 10.1093/emph/eot004.

Schuck N.W., Doeller C.F., Schjeide B.M., Schröder J., Frensch P.A., Bertram L., Li S.C. Aging and KIBRA/WWC1 genotype affect spa-

tial memory processes in a virtual navigation task. *Hippocampus*. 2013;23(10):919-930. DOI 10.1002/hipo.22148.

Takhirova Z.R., Kazantseva A.V., Enikeeva R.F., Vartanyan G.A., Soldatova E.L., Zavyalova I.Yu., Malykh A.S., Davydova Yu.D., Valiev R.R., Nurgalieva A.Kh., Khameta Ya.A., Khusnutdinova E.K., Malykh S.B. Psychogenetics of human spatial abilities. *Rossiyskiy Psikhologicheskiy Zhurnal = Russian Psychological Journal*. 2021; 18(2):67-93. DOI 10.21702/rpj.2021.2.5. (in Russian)

Tucker-Drob E.M., Brandmaier A.M., Lindenberger U. Coupled cognitive changes in adulthood: a meta-analysis. *Psychol. Bull*. 2019; 145(3):273-301. DOI 10.1037/bul0000179.

Wang D., Hu L., Xu X., Ma X., Li Y., Liu Y., Wang Q., Zhuo C. KIBRA and APOE gene variants affect brain functional network connectivity in healthy older people. *J. Gerontol. A Biol. Sci. Med. Sci*. 2019;74(11):1725-1733. DOI 10.1093/gerona/glz004.

Wang D., Liu B., Qin W., Wang J., Zhang Y., Jiang T., Yu C. KIBRA gene variants are associated with synchronization within the default-mode and executive control networks. *Neuroimage*. 2013;69:213-222. DOI 10.1016/j.neuroimage.2012.12.022.

Zhang H., Kranzler H.R., Poling J., Gruen J.R., Gelernter J. Cognitive flexibility is associated with *KIBRA* variant and modulated by recent tobacco use. *Neuropsychopharmacology*. 2009;34(12):2508-2516. DOI 10.1038/npp.2009.80.

ORCID ID

A.V. Kazantseva orcid.org/0000-0002-3744-8058
R.F. Enikeeva orcid.org/0000-0002-4301-5283
Yu.D. Davydova orcid.org/0000-0003-3508-4710
R.N. Mustafin orcid.org/0000-0002-4091-382X
Z.R. Takhirova orcid.org/0000-0003-3609-1971
S.B. Malykh orcid.org/0000-0002-3786-7447
M.M. Lobaskova orcid.org/0000-0003-0318-6480
T.N. Tikhomirova orcid.org/0000-0002-6748-763X
E.K. Khusnutdinova orcid.org/0000-0003-2987-3334

Acknowledgements. The study was partially supported by the Russian Science Foundation (project no.17-78-30028) in the part of psychological assessment and collection of biological materials, by the Ministry of Science and Higher Education of Russian Federation (project no. 075-15-2021-595) in the part of statistical and bioinformatics analysis and by the mega-grant of the Ministry of Science and Higher Education of the Republic of Bashkortostan in the part of genetic analysis.

Conflict of interest. The authors declare no conflict of interest.

Received May 14, 2021. Revised August 25, 2021. Accepted August 27, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):847-854
DOI 10.18699/VJ21.098

Original Russian text www.bionet.nsc.ru/vogis/

Prevalence of the polymorphic H-ficolin (*FCN3*) genes and mannose-binding lectin-associated serine protease-2 (*MASP2*) in indigenous populations from the Russian Arctic regions

M.V. Smolnikova ✉, S.Yu. Tereshchenko

Scientific Research Institute of Medical Problems of the North – a separate division of the Federal Research Center "Krasnoyarsk Science Center" of the Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk, Russia
✉ smarinv@yandex.ru

Abstract. Lectins, being the main proteins of the lectin pathway activating the complement system, are encoded by polymorphic genes, wherein point mutations cause the protein conformation and expression to change, which turns out to have an effect on the functionality and ability to respond to the pathogen. In the current study, large-scale data on the population genotype distribution of the genes for H-ficolin *FCN3* rs28357092 and mannose-binding lectin-associated serine protease *MASP2* rs72550870 among the indigenous peoples of the Russian Arctic regions (Nenets, Dolgans and Nganasans, a mixed population and Russians: a total sample was about 1000 newborns) have been obtained for the first time. Genotyping was carried out using RT-PCR. The frequency of the homozygous variant del/del *FCN3* rs28357092 associated with the total absence of the most powerful activator of the lectin complement pathway, N-ficolin, was revealed; 0 % in the Nenets, 0.8 % in the Dolgans and Nganasans, and 3.5 % among the Russians ($p < 0.01$). Analysis of the prevalence of the *MASP2* genotypes has shown the predominance of the homozygous variant AA in all studied populations, which agrees with the available world data. The heterozygous genotype AG rs72550870 associated with a reduced level of protease was found to occur rarely in the Nenets, Dolgans and Nganasans compared to newborns of Caucasoid origin from Krasnoyarsk: 0.5 % versus 3.3 %, respectively. Moreover, among 323 examined Nenets, one AG carrier was identified, whereas in Russians, 16 out of 242 examined newborns were found to be AG carriers ($p < 0.001$). A homozygous variant (GG) in total absence of protease with impaired binding of both MBL and ficolins was not detected in any of the 980 examined newborns. An additional analysis of infectious morbidity in Arctic populations allows one to find phenotypic characteristics related to a high functional activity of the lectin pathway of complement activation as an most important factor for the first-line of anti-infectious defense, including such new viral diseases as COVID-19.
Key words: *FCN3*; *MASP2*; gene polymorphism; newborns; Russia; Arctic populations.

For citation: Smolnikova M.V., Tereshchenko S.Yu. Prevalence of the polymorphic H-ficolin (*FCN3*) genes and mannose-binding lectin-associated serine protease-2 (*MASP2*) in indigenous populations from the Russian Arctic regions. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8):847-854. DOI 10.18699/VJ21.098

Распространенность полиморфных вариантов генов H-фиколина (*FCN3*) и маннозосвязывающей лектин-ассоциированной сериновой протеазы-2 (*MASP2*) у коренных популяций российских Арктических территорий

М.В. Смольникова ✉, С.Ю. Терещенко

Научно-исследовательский институт медицинских проблем Севера – обособленное подразделение Федерального исследовательского центра «Красноярский научный центр Сибирского отделения Российской академии наук», Красноярск, Россия
✉ smarinv@yandex.ru

Аннотация. Лектины – основные протеины лектинового пути активации системы комплемента, кодируются полиморфными генами, точечные мутации в которых приводят к изменению конформации и экспрессии белка, что в свою очередь отражается на функциональности и способности отвечать на патоген. В настоящем исследовании впервые получены масштабные данные о популяционном распределении частот аллелей генов H-фиколина *FCN3* rs28357092 и маннозосвязывающей лектин-ассоциированной сериновой протеазы-2 *MASP2* rs72550870 среди коренных народностей российских Арктических территорий (ненцы, долганы, нганасаны, смешанная популяция и русские: общая выборка составила около 1000 новорожденных). Генотипирование осуществлено с использованием ПЦР-РВ. Нами выявлена частота гомозиготного варианта del/del *FCN3* rs28357092, ассоциированного с полным отсутствием наиболее мощного активатора лектинового пути

комплемента H-фиколина: у ненцев 0 %, у долган-нганасан 0.8 %, в то время как среди европеоидов 3.5 % ($p < 0.01$). Анализ распространенности генотипов *MASP2* показал преобладание гомозиготного варианта AA во всех исследованных популяциях, что согласуется с доступными мировыми данными. Гетерозиготный генотип AG rs72550870, ассоциированный со сниженным уровнем протеазы, встречается в единичных случаях у ненцев, долган и нганасан по сравнению с новорожденными европеоидного происхождения г. Красноярска: 0.5 и 3.3 % соответственно. Причем у ненцев был выявлен один носитель AG из 323 обследованных, тогда как у европеоидов – 16 из 242 обследованных новорожденных ($p < 0.001$). Гомозиготный вариант GG, которому сопутствует полное отсутствие протеазы с нарушением связывания MBL и фиколинов, не обнаружен ни у одного из 980 обследованных новорожденных. Дополнительный анализ инфекционной заболеваемости в арктических популяциях позволит выявить фенотипические характеристики, сопряженные с высокой функциональной активностью лектинового пути активации комплемента в роли важнейшего фактора первой линии противоинфекционной защиты, в том числе в отношении новых вирусных заболеваний, таких как COVID-19.
Ключевые слова: *FCN3*; *MASP2*; полиморфизм генов; новорожденные; Россия; арктические популяции.

Introduction

The innate immune system provides an immediate, non-specific first line of defense through humoral, cellular and mechanical processes, playing a vital role in protection against pathogenic effects (Dunkelberger, Song, 2010). In the world literature, considerable attention has recently been paid to the study of birth defects of the complement system (CS) in the pathogenesis of various infectious, autoimmune and cardio-metabolic diseases. Thus, in the document from European Society for Immunodeficiencies (ESID) 2020, specifically devoted to summarizing the current state of the problem of various complement component deficiencies, such birth defects were established to account for at least 5 % of the total number of primary immunodeficiencies, with their prevalence and pathogenesis being unexplored (Brodszki et al., 2020).

Plasma proteins of CS interact with each other in three known ways, i. e. lectin (the most phylogenetically ancient), alternative, and classical. All three complement pathways are initiated by many independent stimuli, and subsequently proteolytic cascades are reduced to the main component C3 activation, which leads to the assembly of the membrane-attacking complex (Blom et al., 2004). The lectin pathway (LP) can be activated in the absence of immune complexes and initiated by the binding of molecules of the pattern recognition receptor superfamily (lectins), such as mannose-binding lectin (MBL), collectin 11 (CL-K1), or ficolins, to carbohydrates or acetylated residues found on the pathogen surface or host apoptotic/cancer cells (Ali et al., 2012). Circulating MBL, CL-K1, and ficolins form complexes with specific serine proteases (mannose-binding lectin-associated serine protease, MASP).

In addition to complement activation, lectins reduce the risk of infection by stimulating the secretion of interferon-gamma (IFN-γ), IL-17, IL-6, tumor necrosis factor-alpha (TNF-α) by macrophages (Ren et al., 2014). Three types of ficolins have been described for humans: M-ficolin encoded by the *FCN1* gene, L-ficolin (*FCN2*), and H-ficolin (*FCN3*). M-ficolin is expressed in lungs, monocytes and spleen, L-ficolin is produced in liver and circulates in blood, H-ficolin is expressed in liver and lungs. H-ficolin has been shown to be produced to the greatest extent in lungs, and its complement-activating ability exceeds that of MBL. Ficolin-3 is the most abundant recognition molecule of the lectin pathway, and since it is highly expressed in liver and lung tissues, this indicates its importance both for activating the lectin pathway and for protecting the host lung (Akaiwa et al., 1999; Hummelshoj et al., 2008). In addition, the first evidence of antimicrobial activity of ficolin-3 against the intestinal commensal and opportunistic intestinal bacteria *Hafnia alvei* has recently been obtained (Michalski et al., 2015). It is noteworthy that ficolin-3 is resistant to collagenases (whereas other ficolins and collagens are not), and this may affect its antimicrobial activity, in particular in gastrointestinal tract (Hummelshoj et al., 2008).

Various polymorphic variants in the promoter and structural regions of ficolin genes were given. The *FCN3* gene is located on chromosome 1p36.11 and is highly conserved in humans. Five point mutations responsible for amino acid substitutions have been described, with allele frequencies being below 5 %: p.Leu12Val, p.Leu117fs (known as +1637delC), p.Thr125Ala, p.Glu166Asp and p.Val287Ala (Hummelshoj et al., 2008). The high conservatism of the gene reveals that ficolin-3 may play a crucial role in the immune response. As a matter of fact, there have been very few cases of ficolin-3 deficiency (Thiel, 2007).

H-ficolin (ficolin-3) is the most potent of the known lectin complement pathway activators, and its serum concentrations are significantly higher than those of L-ficolin and MBL (Sallenbach et al., 2011). The rs28357092 (+1637delC) mutation in exon 5 of the *FCN3* gene is a frame-shift mutation leading to truncation of the C-terminal end of the ficolin-3 protein; it accounts for a decrease in plasma levels of H-ficolin by the gene–effect scheme, i. e. homozygotes with such deletion demonstrate a total absence of H-ficolin plasma levels, and heterozygotes do moderate protein levels (Michalski et al., 2011). Homozygosity for +1637delC happens very rarely (1–2 %): only 6 cases have been described in the literature available (all of them suffered from severe infections in early childhood). Data on the population frequency of heterozygous carriage are also scarce, i. e. 15 heterozygotes out of 483 examined individuals were identified in the Icelandic cohort of healthy donors (the frequency was 1.5 %) (Bjarnadottir et al., 2016).

In addition to MBL and ficolins, one of the key participants in the lectin pathway of complement activation is the

family of mannose-binding lectin-associated serine proteases; three proteases (MASP-1, MASP-2, MASP-3) and two related nonenzymatic proteins, MAp19 (sMAP) and MAp44 (MAP-1), were identified in the MASP family (Ricklin et al., 2010). MASP-1 and MASP-2 are of crucial importance in the lectin pathway activation. MASP-1 have been shown to be automatically activated and leads to the MASP-2 activation (Degn et al., 2012). MASP-2 can also become automatically activated, however, it is MASP-1 that is the main activator of MASP-2 under physiological conditions (Héja et al., 2012). MASP-2 is a protease that cleaves complement factors C2 and C4, leading to the complement cascade activation with the formation of inflammatory mediators (C3a and C5a), membrane attack complex (MAC) assembly and opsonization. On the other hand, MASP-3 seems to reduce the activity of the lectin pathway due to competition for MASP binding sites on the recognition molecules (Degn et al., 2010). Moreover, MASP-3 predominantly forms a complex with ficolin-3 and is believed to have an inhibitory effect on complement activation mediated by ficolin-3 (Skjoedt et al., 2010). Levels of the three MASPs have been demonstrated to be predictors of infection and prolonged dependence on life support in critically ill children (Ingels et al., 2014). The most studied among the specific enzymes capable of activating both MBL and ficolins is the type 2 prosthesis, i. e. MASP-2. Serum MASP-2 levels ranged from 125 to 1150 ng/ml, with an average of 416 ng/ml (Sallenbach et al., 2011). An analysis of plasma MASP-2 levels in people of various ethnic groups revealed that the lowest one was found in Africans, then the Hong Kong Chinese, Indians and Caucasian Danes (Thiel et al., 2007).

The polymorphic gene *MASP2* is located on the 1p36.23-31chromosome, has 12 exons, and encodes MASP-2 and MAp19 proteins. The most significant *MASP2* mutation is rs72550870 (p.D120G), which leads to the substitution of aspartic acid for glycine, thereby the protein loses its ability to activate the complement due to inability to form lectin complexes, in particular with MBL and ficolins. Congenital MASP-2 deficiency is caused by the rs72550870 mutation in the homozygous state (GG), characterized by a total absence of serum protease activity (Thiel et al., 2009). A total of thirteen cases of homozygous GG rs72550870 have been described in the literature since the first case was detected in 2003 (Stengaard-Pedersen et al., 2003). Clinical manifestations of decreased activity/inactivity of MASP-2 can range from full health to severe infections and predisposition to cancer (Bjarnadottir et al., 2016). Since three healthy adults with MASP-2 deficiency, homozygous GG in *MASP2*, were reported (Garcia-Laorden et al., 2008), clinical penetrance of this deficiency has become doubtful. Thus, the association of MASP-2 deficiency (GG rs72550870) with clinical manifestations is currently uncertain. Unidentified molecules and functions are likely to be involved in the LP, which could explain why MASP-2 deficiency is relatively common in apparently healthy people (Bjarnadottir et al., 2016). The lectin pathway of the complement activation has been suggested to be unnecessary or also excessive (for example, in severe COVID) for an immune response in healthiest individuals to be formed, and its deficiency is clinically significant only in certain situations, for example, in premature infants (Matricardi et al., 2020).

There are some pronounced population differences in the genotype frequency distribution for polymorphic genes of lectin pathway proteins of the CS. The results of our earlier studies (Tereshchenko, Smolnikova, 2020) have demonstrated that the frequency of the high-producing haplotype (*HYPA*) of the *MBL2* gene is 35.4 % in Russian newborns of Eastern Siberia, corresponding to that of European populations (Holland – 27 %, Denmark – 30 %, Czech Republic – 33 %), as well as Brazilian Caucasians (28–34 %). However, the *HYPA* haplotype frequency in newborns of the Taimyr Dolgan-Nenets region of the Krasnoyarsk Territory was statistically significantly higher than in the Russians and was 64 % for the Nenets and 56 % for the Dolgan-Nganasans, being close to the distribution frequency identified for the Eskimos (81 %) and the North American Indians (64 %). In the aboriginal populations of both the Nenets and Dolgans/Nganasans of the Taimyr Dolgan-Nenets region of the Krasnoyarsk Territory, our research group found a decrease in the prevalence of the *FCN2* rs7851696 genotype, associated with the low binding capacity of L-ficolin to carbohydrates, as compared with Caucasians of Eastern Siberia. The study results (Smolnikova et al., 2017) showed that the Nenets population has a number of important features compared to the Dolgans and Nganasans, i. e. a much lower prevalence of the T allele for the rs17549193 polymorphism and a higher prevalence of the T allele for the *FCN2* rs7851696 polymorphism. We suppose that this genotype can be a genetic marker of the high functional ability of L-ficolin for the Nenets population. In other words, a high prevalence of genotypes associated with high L-ficolin activity in the Arctic populations of the Nenets and Dolgans/Nganasans, compared to the Caucasians of Eastern Siberia has been shown.

As mentioned above, data on the population frequency of the rs28357092 polymorphic variants of the *FCN3* gene are scarce, with much more works on the population frequencies for the rs72550870 polymorphisms of the *MASP2* given. The frequency of the rare G allele in the Danish cohort was 3.9 %; the same frequency was found in the Icelandic sample of adult donors (Bjarnadottir et al., 2016). Interestingly, the G allele was not detected at all in the populations of Hong Kong Chinese, African Zambians, and Brazilian Native Americans (Fumagalli et al., 2017).

The results of the above studies underlie the hypothesis that human evolution has moved on to the accumulation of genotypes with low lectin pathway activity of complement activation due to the prevalence of some intracellular infections, such as tuberculosis and leprosy, where low MBL and L-ficolin activity can cause a protective effect (Verdu et al., 2006; Dunkelberger, Song, 2010). It was suggested that the isolated Arctic populations of the Taimyr Dolgan-Nenets region of the Krasnoyarsk Territory fought these infections historically later and, as a result, retained the high lectin

pathway activity of complement activation formed at the early stages of human evolution.

According to the analysis of the literature data available, to date, the population frequencies of mutations associated with congenital deficiency of H-ficolin (rs28357092) and MASP-2 (rs72550870) in Russian populations and populations of indigenous peoples of the Russian Arctic regions have not been studied. The relevance of obtaining such data for the Russian Arctic populations is increasing, given the accumulating evidence that the lectin pathway of complement activation plays an important role in relation to viral infections. For example, MBL is assumed to be essential in relation to respiratory viral infections, including new coronavirus ones, i. e. SARS and COVID-19 (Matricardi et al., 2020). The role of congenital deficiencies of LP proteins, including H-ficolin and MASP-2, has not been studied in such clinical situations at all. Given that infections are major contributors to infant mortality and lectins are critical for anti-infectious defense, lectin deficiency is likely to contribute to early childhood mortality.

The aim of the work was to identify ethnic differences in the allelic distribution of the lectin pathway component genes of complement activation among indigenous newborns of the Taimyr Dolgan-Nenets region of the Krasnoyarsk Territory (Nenets, Dolgans and Nganasans) compared to the Caucasians of Krasnoyarsk.

Materials and methods

To study single nucleotide polymorphisms rs28357092 *FCN3* and rs72550870 *MASP2*, 980 samples of dried blood stains from newborns from the Taimyr Dolgan-Nenets region of the Krasnoyarsk Territory and the city of Krasnoyarsk, obtained earlier in the Krasnoyarsk Regional Medical Genetic Center, were used.

The newborns were divided into four groups for the ethnic specificity of gene polymorphisms of the lectin pathway of the complement system to be studied: (1) 323 individuals from villages with a predominantly Nenets population (the Nenets make up 85 % of the population); (2) 138 from villages with predominantly Dolgan and Nganasan populations (the Dolgans and Nganasans make up 91 % of the population); (3) 217 from villages with varying combinations of indigenous and mixed populations; (4) 302 newborns of European origin from the city of Krasnoyarsk (the Russians).

Written informed consent of the subjects under the Helsinki Declaration of the World Association "Ethical Principles for Conducting Scientific Medical Research with Human Participation" as revised in 2000 and "The Rules of Clinical Practice in the Russian Federation", adopted by Order of the Russian Ministry of Healthcare No. 266 dated June 19, 2003 was obtained for the investigation to be carried out. The study was approved by the Ethics Committee of Scientific Research Institute of Medical Problems of the North No. 9 dated September 9, 2014.

To isolate DNA from newborn blood stains, a DIAtom DNAPrep reagent kit (Isogen, Russia) was applied. Single-nucleotide polymorphism genotyping of the lectin pathway component genes of complement activation (*FCN3*, *MASP2*) was carried out under the manufacturer's protocol, using the real-time polymerase chain reaction (RT-PCR) method with specific oligonucleotide primers and fluorescently labeled probes (TaqMan) (DNA-synthesis, Russia). Nucleotide sequences of allele-specific probes for genotyping polymorphisms are as follows: for rs28357092 *FCN3* F – CCT CGGTGTCCATGTCAC, R – CCACCTTGAGCGGCTGG (fluorophore/allele – VIC/del, FAM/G); for rs72550870 *MASP2* F – GCAAGGACACTTTCTACTCGC, R – TCA CCCTCGGCTGCATAG (fluorophore/allele – VIC/G, FAM/A).

The correspondence of genotype frequencies to the Hardy–Weinberg equilibrium was verified using χ^2. Genotype frequency comparisons were performed using the two-sided Fisher's exact test. Statistically significant differences were accepted at $p < 0.05$ after Bonferroni correction for multiple tests.

Results and discussion

The advantage of our approach to population assessment for the prevalence of immunodeficient genotypes of lectin pathway's mediators of complement activation was to study newborn populations, where unfavorable genetic variations had not been excluded, which is possible at an older age as a result of the clinical realization of genetic predisposition.

The genotype frequencies and the variant allele of the H-ficolin gene *FCN3* rs28357092 are given in Table 1. The prevalence analysis of the *FCN3* genotypes revealed the predominance of the homozygous GG variant in all populations studied in the work, which is consistent with the world data available.

The variant deletion allele (del) *FCN3* rs28357092 in a heterozygous state was not found in any newborns of the three indigenous populations of the Taimyr Dolgano-Nenets region, except for one Russian individual from the city of Krasnoyarsk. Although the literature describes underrepresentation of the homozygous genotype for this deletion, it was found in our cohort of the studied samples in 10 Russian newborns (3.3 %), in 4 newborns from a mixed population (2.0 %) and in one of the Dolgan and Nganasan group (0.8 %). In the Nenets, neither homozygotes nor heterozygotes by the mutant deletion *FCN3* rs28357092 were identified. Thus, in the total sample of 926 newborns, del/del homozygotes were detected in 15 individuals, being 1.6 %. According to the Internet source http://www.ensembl.org, the variant allele frequency in world populations is 1–3 %, with it being zero in Asian populations. As mentioned above, the rs28357092 (+1637delC) mutation in the *FCN3* gene leads to a plasma level decrease of H-ficolin, i. e. the deletion homozygotes, being rare, have a complete absence of plasma H-ficolin, and heterozygotes have medium protein levels (Michalski et al., 2011; Bjarnadottir et al., 2016). It is likely that in other studies, homozygotes were not detected with adult populations being examined, demonstrating again the advantage of our approach for identifying the true genotype frequencies with a cohort of newborns, where the

Table 1. *FCN3* rs28357092 genotype frequencies among newborns from different ethnic populations of the Taymyr Dolgan-Nenets region of the Krasnoyarsk Territory and the city of Krasnoyarsk, *n* (%)

Genotype	Nenets (*n* = 292)	Dolgans and Nganasans (*n* = 129)	Mixed Arctic populations (*n* = 203)	Russians (*n* = 302)	*p**
	1	2	3	4	
GG	292 (100.0)	128 (99.2)	199 (98.0)	291 (96.4)	1–3 = 0.02 1–4 < 0.001
G/del	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.3)	–
del/del	0 (0.0)	1 (0.8)	4 (2.0)	10 (3.3)	1–3 = 0.02 1–4 = 0.002
del*	0 (0.0)	0 (0.8)	8 (2.0)	21 (3.5)	1–3 < 0.001 1–4 < 0.001 2–3 = 0.02 2–4 = 0.003

* Only *p*-values ≤ 0.05 are given.

Table 2. *MASP2* rs72550870 genotype frequencies among newborns from different ethnic populations of the Taymyr Dolgan-Nenets region of the Krasnoyarsk Territory and the city of Krasnoyarsk, *n* (%)

Genotype	Nenets (*n* = 323)	Dolgans and Nganasans (*n* = 138)	Mixed Arctic populations (*n* = 217)	Russians (*n* = 242)	*p**
	1	2	3	4	
AA	322 (99.7)	136 (98.6)	213 (98.2)	226 (93.4)	1–4 < 0.001 2–4 = 0.02 3–4 = 0.01
AG	1 (0.3)	2 (1.4)	4 (1.8)	16 (6.6)	1–4 < 0.001 2–4 = 0.02 3–4 = 0.01
GG	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	–
G*	1 (0.2)	2 (0.7)	4 (0.9)	16 (3.3)	1–4 < 0.001 2–4 = 0.03 3–4 = 0.01

* Only *p*-values ≤ 0.05 are given.

deficiency variants have not been eliminated due to infections and decease.

The genotype frequencies and variant allele of the serine protease gene *MASP2* rs72550870 are presented in Table 2. Analysis of the *MASP2* genotype prevalence has shown the predominance of the homozygous AA variant among all populations studied in the work, which is consistent with the data available. The heterozygous genotype AG rs72550870 was found in some isolated incidents in the Nenets and Dolgan-Nganasans as compared with the Caucasian newborns in the city of Krasnoyarsk. The frequency of the AG genotype in the Russians (6.6 %) is statistically significantly higher than in the Arctic populations (the Nenets being 0.3 %, $p < 0.001$; the Dolgan-Nganasans being 1.4 %, $p = 0.02$; the mixed populations being 1.8 %, $p = 0.01$). Thus, the heterozygous AG variant occurred in 16 out of 242 Russian newborns, whereas in 323 Nenets it occurred in only one individual. None of the population groups were found to have homozygotes for the minor G allele associated with the absence of serum protease activity.

The allelic variant G of the *MASP2* rs72550870 has zero or extremely low frequencies in the world populations. According to the Internet source http://www.ensembl.org, the frequency in Caucasoid populations is 4.0 %, in the general American population – 2.0 %, among Asian and African populations – zero. In the course of the study, we obtained data on the prevalence of the mutant allele G rs72550870 in the Russian Arctic populations: 0.5 % among newborns in the Taimyr Dolgan-Nenets region of the Krasnoyarsk Territory ($n = 678$) and 3.3 % among Russians in Krasnoyarsk ($n = 242$).

Data on nine *MASP2* gene mutations based on the two most informative studies carried out by the S. Triel group in 2007 and 2009 were obtained, that is, the prevalence of mutant allelic variants for almost all polymorphisms was too low. There was a change in the MASP-2 protein structure

that resulted from the rs72550870 mutation, which led to impaired binding into the MBL complex, resulting in the inability to activate the complement system. In addition, the authors note that it is the Caucasian population that has the variant G allele (rs72550870) as the main reason for the lower MASP-2 levels. Population analysis reported the absence of the homozygous genotype GG rs72550870 among the adult Chinese, Africans, Caucasians, Greenland Intuits, and Brazilians (Thiel et al., 2007, 2009). The heterozygous variant prevailed in the Caucasians from Denmark (3.9 %) and Inuits of western Greenland (where the European admixture is high, as reported by the authors) (3.7 %), however, that was not found to occur in other studied populations ($p < 0.0001$).

Moreover, the authors (Thiel et al., 2009) provided the frequencies of a rare allelic variant obtained by other researchers in different populations among healthy individuals and patients with various diseases. Thus, 14 heterozygotes were found among 112 patients with cystic fibrosis (the frequency was 6.3 %) and five heterozygotes were found among 200 healthy people (the frequency – 1.3 %) in the Swedish population. In a study of psoriasis patients and their families, 894 individuals were tested for *MASP2* rs72550870 and a total of 62 heterozygotes and one homozygote were found, resulting in a gene frequency of 3.6 % (the allele was not associated with psoriasis). Homozygosity was recorded in one person in a group of 293 Polish children with respiratory infections and in one child with cystic fibrosis. Two homozygotes were found among 2,008 individuals (including 967 pneumonia patients, 130 SLE patients, 43 children with recurrent respiratory infections, and 868 healthy people) in a recent study of the Spanish population, no association of the desease with the variant allele being found. The absence of the allelic variant G of the *MASP2* rs72550870 in China was documented by a report examining the influence of both *MBL2* and *MASP2* genotypes on susceptibility to severe acute respiratory syndrome (SARS). No G allele was found in all 1,757 Asians tested. Thus, MBL-deficiency, as well as deficiencies of other complement components (including *MASP2*) should not be concluded to cause the disease or susceptibility to infections, but rather are clinical modifiers impairing other elements of the activation cascade.

Studying the role of congenital defects of the complement system in the pathogenesis of various diseases is a hot issue, as the congenital complement component deficiencies account for at least 5 % of the total number of primary immunodeficiencies, whereas their prevalence and pathogenesis remain unstudied. Large-scale data on the population genotype distribution of genes for H-ficolin *FCN3* rs28357092 and mannose-binding lectin-associated serine protease *MASP2* rs72550870 among the indigenous peoples of the Russian Arctic regions (the total sample of the newborns studied was 980) were first obtained in the work. As mentioned above, currently, the population frequencies of mutations associated with congenital deficiency of H-ficolin and MASP-2 in Russian populations as a whole, and in populations of indigenous peoples of the Russian Arctic regions in particular, have not been previously studied. Moreover, the previously identified features of the genetic regulation of lectin pathway proteins of complement activation in newborns of the Taimyr Dolgan-Nenets region of the Krasnoyarsk Territory have shown the indigenous populations of the Arctic to be genetically characterized by greater activity of at least two different lectin pathway components of complement activation, i. e. MBL and L-ficolin, indicating a high tone of the lectin pathway of complement activation in general (Smolnikova et al., 2017; Tereshchenko, Smolnikova, 2020).

Currently, there are two competing hypotheses accounting for the high genotype population diversity of the lectin compliment pathway (Eisen, Osthoff, 2014). The first of them proves to be a protective role of low-producing genotypes against some intracellular pathogens, i. e. tuberculosis and leprosy, visceral leishmaniasis, atypical pneumonia. A high level of lectin-mediated phagocytosis may predispose to better penetration of intracellular pathogens into the cytoplasm of host cells, screening the pathogens from factors of adaptive immunity and, consequently, a greater risk of active infection process. Thus, the actual data available suggest a "double pathophysiological role" of the lectin pathway of complement activation, i. e. protective against extracellular pathogens, especially in young children, and provocative against some intracellular pathogens and atherosclerosis. The population genetic consequences of such a "double role" may be the root cause of the ethnic diversity for the corresponding genotypes, being the main point for the first hypothesis mentioned above, based on the assumption for the selection benefit of the components deficiency of the lectin pathway of complement activation for some populations (Seyfarth et al., 2005; Eisen, Osthoff, 2014). The second hypothesis denies the existence of any selection pressure on the genotypes of the lectin complement pathway, accounting for genetic diversity specifically by migration processes and gene drift. However, the authors of the studies stipulate that "It is possible that stochastic evolutionary factors erased much of the ancient imprint left by natural selection and more powerful tests in greater population samples would be necessary to confirm the data" (Verdu et al., 2006; Boldt et al., 2010).

Conclusion

Therefore, lower prevalence of the genetic markers of the H-ficolin and MASP-2 deficiencies in the indigenous populations of the Arctic regions of the Krasnoyarsk Territory as compared with the Caucasians of Krasnoyarsk city, associated with a genetic predisposition to a high functional activity of L-ficolin compared to Caucasian population, has been expected to be revealed in the given work.

The study of the ethnically associated non-specific antiinfectious protection among the indigenous population of the Taimyr Dolgan-Nenets region of the Krasnoyarsk Territory can be used to formulate plans for practical health authorities towards the infection prevention and in order to efficiently attract labor resources for working in conditions

with a high infectious load. An additional analysis of infectious morbidity among the Arctic populations will allow one to reveal phenotypic characteristics associated with a high functional activity of the lectin pathway of complement activation as the most important factor for the first line of anti-infectious defense, including such new viral diseases as COVID-19. Such clinical and genetic comparisons are extremely important for elucidating the physiological role of MBL, ficolins, and MASP-2, and we identified genetic features of the ethnically isolated indigenous Arctic populations of the Krasnoyarsk Territory as a unique material to be studied.

References

Akaiwa M., Yae Y., Sugimoto R., Suzuki S.O., Iwaki T., Izuhara K., Hamasaki N. Hakata antigen, a new member of the ficolin/opsonin p35 family, is a novel human lectin secreted into bronchus/alveolus and bile. *J. Histochem. Cytochem.* 1999;47:777-786. DOI 10.1177/002215549904700607.

Ali Y.M., Lynch N.J., Haleem K.S., Fujita T., Endo Y., Hansen S., Holmskov U., Takahashi K., Stahl G.L., Dudler T., Girija U.V., Wallis R., Kadioglu A., Stover C.M., Andrew P.W., Schwaeble W.J. The lectin pathway of complement activation is a critical component of the innate immune response to pneumococcal infection. *PLoS Pathog.* 2012;8(7):e1002793. DOI 10.1371/journal.ppat.1002793.

Bjarnadottir H., Arnardottir M., Ludviksson B.R. Frequency and distribution of *FCN2* and *FCN3* functional variants among *MBL2* genotypes. *Immunogenetics.* 2016;68(5):315-325. DOI 10.1007/s00251-016-0903-4.

Blom A.M., Villoutreix B.O., Dahlbäck B. Complement inhibitor C4b-binding protein-friend or foe in the innate immune system? *Mol. Immunol.* 2004;40:1333-1346. DOI 10.1016/j.molimm.2003.12.002.

Boldt A.B., Messias-Reason I.J., Meyer D., Schrago C.G., Lang F., Lell B., Dietz K., Kremsner P.G., Petzl-Erler M., Kun J.F. Phylogenetic nomenclature and evolution of mannose-binding lectin (MBL2) haplotypes. *BMC Genetics.* 2010;11(1):38. DOI 10.1186/1471-2156-11-38.

Brodszki N., Frazer-Abel A., Grumach A.S., Kirschfink M., Litzman J., Perez E., Seppänen M.R.J., Sullivan K.E., Jolles S. European Society for Immunodeficiencies (ESID) and European Reference Network on Rare Primary Immunodeficiency, Autoinflammatory and Autoimmune Diseases (ERN RITA) complement guideline: deficiencies, diagnosis, and management. *J. Clin. Immunol.* 2020;40(4):576-591. DOI 10.1007/s10875-020-00754-1.

Degn S.E., Jensen L., Gál P., Dobó J., Holmvad S.H., Jensenius J.C., Thiel S. Biological variations of MASP-3 and MAp44, two splice products of the *MASP1* gene involved in regulation of the complement system. *J. Immunol. Methods.* 2010;361:37-50. DOI 10.1016/j.jim.2010.07.006.

Degn S.E., Jensen L., Hansen A.G., Duman D., Tekin M., Jensenius J.C., Thiel S. Mannan-binding lectin-associated serine protease (MASP)-1 is crucial for lectin pathway activation in human serum, whereas neither MASP-1 nor MASP-3 is required for alternative pathway function. *J. Immunol.* 2012;189:3957-3969. DOI 10.4049/jimmunol.1201736.

Dunkelberger J.R., Song W.-C. Complement and its role in innate and adaptive immune responses. *Cell Res.* 2010;20:34-50. DOI 10.1038/cr.2009.139.

Eisen D.P., Osthoff M. If there is an evolutionary selection pressure for the high frequency of *MBL2* polymorphisms, what is it? *Clin. Exp. Immunol*. 2014;176(2):165-171. DOI 10.1111/cei.12241.

Fumagalli S., Perego C., Zangari R., De Blasio D., Oggioni M., De Nigris F., Snider F., Garred P., Ferrante A.M., De Simoni M.G. Lectin pathway of complement activation is associated with vulnerability of atherosclerotic plaques. *Front. Immunol.* 2017;8:288. DOI 10.3389/fimmu.2017.00288.

Garcia-Laorden M.I., Sole-Violan J., Rodriguez de Castro F., Aspa J., Briones M.L., Garcia-Saavedra A., Rajas O., Blanquer J., Caballero-Hidalgo A., Marcos-Ramos J.A., Hernandez-Lopez J., Rodriguez-Gallego C. Mannose-binding lectin and mannose-binding lectin-associated serine protease 2 in susceptibility, severity, and outcome of pneumonia in adults. *J. Allergy Clin. Immunol.* 2008;122(2):368-374. DOI 10.1016/j.jaci.2008.05.037.

Héja D., Harmat V., Fodor K., Wilmanns M., Dobó J., Kékesi K.A. Monospecific inhibitors show that both mannan-binding lectin-associated serine protease-1 (MASP-1) and -2 are essential for lectin pathway activation and reveal structural plasticity of MASP-2. *J. Biol. Chem.* 2012;287:20290-20300. DOI 10.1074/jbc.M112.354332.

Hummelshoj T., Fog L.M., Madsen H.O., Sim R.B., Garred P. Comparative study of the human ficolins reveals unique features of ficolin-3 (Hakata antigen). *Mol. Immunol.* 2008;45:1623-1632. DOI 10.1016/j.molimm.2007.10.006.

Ingels C., Vanhorebeek I., Steffensen R., Derese I., Jensen L., Wouters P.J., Hermans G., Thiel S., Berghe G.V. Lectin pathway of complement activation and relation with clinical complications in critically ill children. *Pediatr. Res.* 2014;75:99-108. DOI 10.1038/pr.2013.180.

Matricardi P.M., Negro R.W.D., Nisin R. The first, holistic immunological model of COVID-19: implications for prevention, diagnosis, and public health measures. *Pediatr. Allergy Immunol.* 2020;31(5):454-470. DOI 10.1111/pai.13271.

Michalski M., Swierzko A.St., Lukasiewicz J., Man-Kupisinska A., Karwaciak I., Przygodzka P., Cedzynski M. Ficolin-3 activity towards the opportunistic pathogen, *Hafnia alvei*. *Immunobiology.* 2015;220:117-123. DOI 10.1016/j.imbio.2014.08.012.

Michalski M., Szala A., Swierzko A.St., Lukasiewicz J., Maciejewska A., Kilpatrick D.C., Matsushita M., Domzalska-Popadiuk I., Borkowska-Klos M., Sokolowska A., Szczapa J., Lugowski C., Cedzynski M. H-ficolin (ficolin-3) concentrations and *FCN3* gene polymorphism in neonates. *Immunobiology.* 2011;217:730-737. DOI 10.1016/j.imbio.2011.12.004.

Ren Y., Ding Q., Zhang X. Ficolins and infectious diseases. *Virol. Sin.* 2014;29:25-32. DOI 10.1007/s12250-014-3421-2.

Ricklin D., Hajishengallis G., Yang K., Lambris J.D. Complement: a key system for immune surveillance and homeostasis. *Nat. Immunol.* 2010;11:785-797. DOI 10.1038/ni.1923.

Sallenbach S., Thiel S., Aebi C., Otth M., Bigler S., Jensenius J.C., Schlapbach L.J., Ammann R.A. Serum concentrations of lectin-pathway components in healthy neonates, children and adults: mannan-binding lectin (MBL), M-, L-, and H-ficolin, and MBL-associated serine protease-2 (MASP-2). *Pediatr. Allergy Immunol.* 2011;22:424-430. DOI 10.1111/j.1399-3038.2010.01104.x.

Seyfarth J., Garred P., Madsen H.O. The 'involution' of mannose-binding lectin. *Hum. Mol. Genet*. 2005;14(19):2859-2869. DOI 10.1093/hmg/ddi318.

Skjoedt M.-O., Hummelshoj T., Palarasah Y., Honore C., Koch C., Skjodt K., Garred P. A novel mannose-binding lectin/ficolin-associated protein is highly expressed in heart and skeletal muscle tissues and inhibits complement activation. *J. Biol. Chem.* 2010;285:8234-8243. DOI 10.1074/jbc.M109.065805.

Smolnikova M.V., Freidin M.B., Tereshchenko S.Y. The prevalence of the variants of the L-ficolin gene (*FCN2*) in the arctic populations of East Siberia. *Immunogenetics.* 2017;69(6):409-413. DOI 10.1007/s00251-017-0984-8.

Stengaard-Pedersen K., Thiel S., Gadjeva M., Møller-Kristensen M., Sørensen R., Jensen L.T., Sjøholm A.G., Fugger L., Jensenius J.C. Inherited deficiency of mannan-binding lectin-associated serine protease 2. *N. Engl. J. Med.* 2003;349(6):554-560. DOI 10.1056/NEJMoa022836.

Tereshchenko S.Yu., Smolnikova M.V. The polymorphism of the gene of mannose-binding lectin in indigenous populations of the territories of the Arctic zone of the Russian Federation. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):868-875. DOI 10.18699/VJ20.685. (in Russian)

Thiel S. Complement activating soluble pattern recognition molecules with collagen-like regions, mannan-binding lectin, ficolins and associated proteins. *Mol. Immunol.* 2007;44:3875-3888. DOI 10.1016/j.molimm.2007.06.005.

Thiel S., Kolev M., Degn S., Steffensen R., Hansen A.G., Ruseva M., Jensenius J.C. Polymorphisms in mannan-binding lectin (MBL)-associated serine protease 2 affect stability, binding to MBL, and enzymatic activity. *J. Immunol.* 2009;182:2939-2947. DOI 10.4049/jimmunol.0802053.

Thiel S., Steffensen R., Christensen I.J., Ip W.K., Lau Y.L., Reason I.J., Eiberg H., Gadjeva M., Ruseva M., Jensenius J.C. Deficiency of mannan-binding lectin associated serine protease-2 due to missense polymorphisms. *Genes Immun.* 2007;8:154-163. DOI 10.1038/sj.gene.6364373.

Verdu P., Barreiro L.B., Patin E., Gessain A., Cassar O., Kidd J.R., Kidd K.K., Behar D.M., Froment A., Heyer E., Sica L., Casanova J.L., Abel L., Quintana-Murci L. Evolutionary insights into the high worldwide prevalence of *MBL2* deciency alleles. *Hum. Mol. Genet.* 2006;15(17):2650-2658. DOI 10.1093/hmg/ddl193.

ORCID ID

M.V. Smolnikova orcid.org/0000-0001-9984-2029
S.Yu. Tereshchenko orcid.org/0000-0002-1605-7859

Acknowledgements. The authors are grateful to Maxim Freydin for statistical evaluation advice.

Conflict of interest. The authors declare no conflict of interest.

Received June 15, 2021. Revised August 26, 2021. Accepted August 26, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):855-863
DOI 10.18699/VJ21.099

Original Russian text www.bionet.nsc.ru/vogis/

Putative regulatory functions of SNPs associated with bronchial asthma, arterial hypertension and their comorbid phenotype

I.A. Goncharova ✉, E.Yu. Bragina, I.Zh. Zhalsanova, M.B. Freidin, M.S. Nazarenko

Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia
✉ irina.goncharova@medgenetics.ru

Abstract. Linkage disequilibrium (LD) of single nucleotide polymorphisms (SNPs) of *TLR4/AL160272.2* (rs1927914, rs1928298, rs7038716, rs7026297, rs7025144) was estimated in the Slavs of West Siberia. We further investigated an association of SNPs in *TLR4/AL160272.2* (rs1927914, rs7038716, rs7025144), *SERPINA1* (rs1980616), *ATXN2/BRAP* (rs11065987), *IL2RB* (rs2284033), *NT5C2* (rs11191582), *CARD8* (rs11669386), *ANG/RNASE4* (rs1010461), and *ABTB2/CAT* (rs2022318) genes with bronchial asthma (BA), arterial hypertension (AH) and their comorbidity. Then, the disease-associated SNPs were annotated *in silico* in relation to their potential regulatory functions. Strong LD was detected between rs1928298 and rs1927914, as well as rs7026297 and rs7038716 in the Slavs of West Siberia. It was found that the rs1927914 G allele of the *TLR4* gene and the rs1980616 C allele of the *SERPINA1* gene are associated with the predisposition to BA. These SNPs can affect binding affinity of transcription factors of the Pou and Klf4 families, as well as the expression levels of the *TLR4* and *SERPINA1* genes. The rs11065987 allele A of the *ATXN2/BRAP* genes, the rs11669386 A allele of the *CARD8* gene, the rs2284033 allele G of the *IL2RB* gene, and the rs11191582 allele G of the *NT5C2* gene were associated with the risk of AH. These variants can alter binding affinity of the Hoxa9, Irf, RORalpha1 and HMG-IY transcription factors, as well as the expression levels of the *ALDH2*, *CARD8*, *NT5C2*, *ARL3*, and *SFXN2* genes in blood cells/vessels/heart, respectively. The risk of developing a comorbid phenotype of AD and AH is associated with the A allele of rs7038716 and the T allele of rs7025144 of the *TLR4/AL160272.2* genes, the A allele of rs1010461 of the *ANG* gene and the C allele of rs2022318 of the *ABTB2/CAT* genes. Variants rs7038716 and rs7025144 can change the expression levels of the *TLR4* gene in blood cells, while rs1010461 and rs2022318 influence the expression levels of the *ANG* and *RNASE4* genes as well as the *CAT* and *ABTB2* genes in blood cells, lungs/vessels/heart.

Key words: bronchial asthma; arterial hypertension; comorbidity; SNP; regulatory functions.

For citation: Goncharova I.A., Bragina E.Yu., Zhalsanova I.Zh., Freidin M.B., Nazarenko M.S. Putative regulatory functions of SNPs associated with bronchial asthma, arterial hypertension and their comorbid phenotype. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8):855-863. DOI 10.18699/VJ21.099

Регуляторный потенциал SNP-маркеров генов, ассоциированных с бронхиальной астмой, артериальной гипертензией и их коморбидным фенотипом

И.А. Гончарова ✉, Е.Ю. Брагина, И.Ж. Жалсанова, М.Б. Фрейдин, М.С. Назаренко

Научно-исследовательский институт медицинской генетики Томского национального исследовательского медицинского центра Российской академии наук, Томск, Россия
✉ irina.goncharova@medgenetics.ru

Аннотация. В работе оценено неравновесие по сцеплению однонуклеотидных вариантов (SNP) генов *TLR4/AL160272.2* (rs1927914, rs1928298, rs7038716, rs7026297, rs7025144) у славян-жителей Западной Сибири. Проведен анализ ассоциаций SNP в области генов *TLR4/AL160272.2* (rs1927914, rs7038716, rs7025144), *SERPINA1* (rs1980616), *ATXN2/BRAP* (rs11065987), *IL2RB* (rs2284033), *NT5C2* (rs11191582), *CARD8* (rs11669386), *ANG/RNASE4* (rs1010461) и *ABTB2/CAT* (rs2022318) при бронхиальной астме (БА), артериальной гипертензии (АГ) и их сочетании. Выполнено *in silico* аннотирование SNP, ассоциированных с данными заболеваниями, в отношении их регуляторного потенциала. В результате у славян-жителей Западной Сибири выявлено сильное неравновесие по сцеплению rs1928298 и rs1927914, а также rs7026297 и rs7038716. Установлено, что к развитию БА предрасполагают аллель G rs1927914 гена *TLR4* и аллель C rs1980616 гена *SERPINA1*. Данные SNP влияют на изменение аффинности транскрипционных факторов семейств Pou и Klf4, а также на экспрессию генов *TLR4* и *SERPINA1* в клетках крови соответственно. Аллель A rs11065987 генов *ATXN2/BRAP*, аллель A rs11669386 гена

CARD8, аллель G rs2284033 гена *IL2RB* и аллель G rs11191582 гена *NT5C2* ассоциированы с риском развития артериальной гипертензии. Данные варианты изменяют аффинность транскрипционных факторов Hoxa9, Irf, RORalpha1 и HMG-IY, а также экспрессию генов *ALDH2*, *CARD8*, *NT5C2*, *ARL3* и *SFXN2* в клетках крови, сосудах и сердце. Риск развития коморбидного фенотипа БА и АГ ассоциирован с аллелем A rs7038716 и аллелем T rs7025144 генов *TLR4/AL160272.2*, аллелем A – rs1010461 гена *ANG* и аллелем C – rs2022318 генов *ABTB2/CAT*. Варианты rs7038716 и rs7025144 изменяют экспрессию гена *TLR4* в клетках крови, а rs1010461 и rs2022318 – генов *ANG* и *RNASE4*, *CAT* и *ABTB2* в клетках крови, легких, сосудов и сердца.
Ключевые слова: бронхиальная астма; артериальная гипертензия; коморбидность; SNP; регуляторный потенциал.

Introduction

To date, association studies have identified a large number of genetic variants associated with the development of the risk of multifactorial diseases (https://www.ebi.ac.uk/gwas/). However, moving from establishing an association to understanding the mechanisms underlying diseases is a more difficult task. Linkage disequilibrium (LD) that differs in populations of different ethnic origin makes it difficult to identify causal genetic variants of a disease or a complex trait.

On the other hand, to understand the mechanisms of multifactorial diseases, it is important to assess the genetic variants' functional significance, including annotation of the polymorphisms' regulatory potential in relation to changes in the genes' functional activity. Genetic variants affecting quantitative changes in gene expression profile (eQTL), or single nucleotide polymorphic variants (SNP), which have a regulatory status (rSNP) and are located in actively transcribed DNA regions, cause deviations from the optimal program of gene functioning in tissue cells and organs, which leads to an increased risk of diseases (van Arensbergen et al., 2019).

Bronchial asthma (BA) is a widespread heterogeneous disease characterized by chronic airway inflammation. Various comorbidities are common among BA patients, including allergic conditions: allergic rhinitis, dermatitis, and food allergy (Weatherburn et al., 2017), as well as some non-allergic pathological conditions: arterial hypertension (AH), obesity, type 2 diabetes mellitus, and other metabolic and endocrine disorders (Su et al., 2016). These diseases have common pathogenesis with BA and can modify clinical symptoms and the pathological process course in patients. For example, BA patients with concomitant AH, as a rule, have a number of phenotypic features, including older age, late-onset BA, high body mass index, and are characterized by a predominantly neutrophilic type of inflammation, which is not implemented through Th-2 lymphocytes (Moore et al., 2014).

Previously, we selected the genes most significant for the development of BA and AH comorbidity by analyzing the associative gene network structure and prioritization methods (Saik et al., 2018). As a result of the SNPs analysis of the selected priority genes it was shown that BA is associated with rs1928298 and rs1927914, localized at a distance of 19.6 and 1.7 Kb, respectively, from the 5′-end of *TLR4* gene, as well as the intron variant rs1980616 of *SERPINA1* gene. AH was associated with rs11065987 localized between *ATXN2* and *BRAP* genes, as well as intron variants of *IL2RB* gene rs2284033, rs11191582 of *NT5C2* gene, and rs11669386 of *CARD8* gene. Rs7038716, rs7026297, and rs7025144 localized at the 1 intron of the *AL160272.2* gene and at a distance of 35.6, 33.1, and 9.4 Kb, respectively, from the *TLR4* gene 3′-end; rs1010461 (*ANG/RNASE4*), and intergenic rs2022318 localized between *ABTB2* and *CAT* genes were associated with comorbid BA+AH phenotype (Bragina et al., 2018). The *CAT* and *TLR4* genes' haplotypes associations with the development of BA and the comorbid BA+AH phenotype have been established (Bragina et al., 2019). However, it has not previously been assessed which alleles of these SNPs are associated with diseases, what is the possible molecular mechanism explaining the obtained associations, and whether genetic variants have a regulatory potential for the functional activity of genes, which was the purpose of this study.

Materials and methods

The study examined three groups of patients: with BA ($n = 145$, 73.1 % women, 25.9 % men, age 44.89 ± 8.86 years); with AH without BA in history ($n = 144$, 32.6 % women, 67.4 % men, age 51.27 ± 6.05 years); with BA and AH ($n = 146$, 72.6 % women, 27.4 % men, age 56.32 ± 10.47 years). The control group included individuals with normal blood pressure and no clinical asthma manifestations ($n = 152$, 73.7 % women, 26.3 % men, mean age in the group 47.75 ± 9.92 years). The diagnosis of "bronchial asthma" and "arterial hypertension" was established on the basis of patients' clinical examination, according to generally accepted criteria. All individuals were Eastern Europeans (predominantly, Slavs).

Detailed methods for selecting and prioritizing SNPs, as well as the procedure for genotyping patients' DNA samples using mass spectrometry on a Sequenom MassARRAY® device (USA), are given elsewhere (Bragina et al., 2018, 2019; Saik et al., 2018).

In this work, we evaluated linkage disequilibrium (LD) between single nucleotide polymorphisms (SNPs) of *TLR4/AL160272.2* genes (rs1927914, rs1928298, rs7038716, rs7026297, rs7025144) in Slavs living in West Siberia. Lewontin's LD coefficient (D′) and Pearson's correlation coefficient (r^2) were estimated using HaploView v. 4.2.

Odds ratio (OR) and 95 % confidence intervals (CI) in relation to the risk of bronchial asthma (BA), arterial hypertension (AH), and their combination were calculated for SNPs in the region of *TLR4/AL160272.2* (rs1927914, rs7038716, rs7025144), *SERPINA1* (rs1980616), *ATXN2/BRAP* (rs11065987), *IL2RB* (rs2284033), *NT5C2* (rs11191582), *CARD8* (rs11669386), *ANG/RNASE4* (rs1010461) and *ABTB2/CAT* (rs2022318) genes using logistic regression. Sex and age were used as covariates in the regression analysis. Odds ratios were calculated by exponentiating corresponding regression coefficients. For statistical analysis, Stats and Genetics packages were used in R program environment (The R Foundation).

To assess the SNPs regulatory potential associated with these diseases, the following databases were used: rSNPBase v. 3.1 (http://rsnp3.psych.ac.cn), HaploReg v. 4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php), RegulomeDB v. 2.0 (https://www.regulomedb.org/regulome), GTEx Portal (https://gtexportal.org/home), Blood eQTL browser (https://genenetwork.nl/bloodeqtlbrowser). The search for data on the relationship of the studied genetic variants with diseases was carried out using the DisGeNET resource (https://www.disgenet.org/).

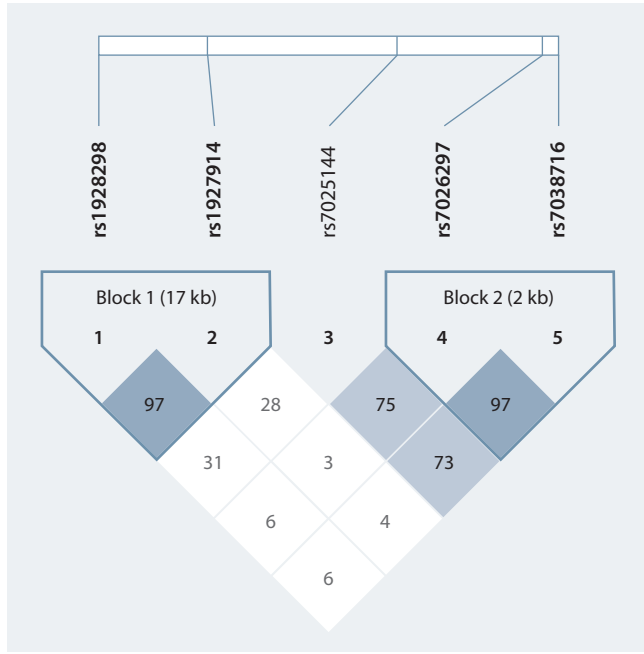
The study was carried out using the Biobank of the Population of Northern Eurasia on the basis of the Core Facilities Center of Scientific Research Equipment and Experimental Biological Material "Medical Genomics" of the Research Institute of Medical Genetics of the Tomsk National Research Medical Center of the Russian Academy of Sciences. The study protocol was approved by the Ethics Committee of the Research Institute of Medical Genetics (Protocol No. 2 dated 05/30/2016). Informed consent was obtained for all participants.

Results and discussion

Linkage disequilibrium analysis of genetic variants located at locus 9q33.1 (*TLR4/AL160272.2*)

Since five of the associated variants – rs1927914, rs1928298, rs7038716, rs7026297, and rs7025144 – are located in the 9q33.1 locus (*TLR4/AL160272.2*), at the first stage we used all available individuals (n = 587) to analyze LD between the SNPs, which showed the presence of two blocks, one of which included rs1928298 and rs1927914 (D′ = 0.974; r^2 = 0.949) (see the Figure).

The haplotype structure of the 9q33.1 locus (*TLR4/AL160272.2*) in the Slavs of West Siberia did not differ from that characteristic in the USA Caucasians (http://www.ensembl.org/Homo_sapiens/Variation/HighLD?db=core;r=9:117684048-117685048;v=rs1928298;vdb=variation;vf=729411740#373514_tablePanel/.



Haplotype structure of locus 9q33.1 (*TLR4/AL160272.2*), including rs1927914, rs1928298, rs7038716, rs7026297, and rs7025144 in the Slavs of West Siberia.

Association analysis of SNPs with bronchial asthma, arterial hypertension, and their comorbid phenotype

We chose variants in linkage equilibrium (rs1927914, rs7038716) and a SNP that showed weak LD with the second block (rs7025144) to carry out association analysis. As a result, we revealed an association between BA and the G allele of rs1927914 of *TLR4* gene and allele C of rs1980616 of *SERPINA1* gene; between AH and allele A of rs11065987 (intergenic region *ATXN/BRAP*), allele A of rs11669386 of *CARD8* gene, allele G of rs2284033 of *IL2RB* gene, and allele G of rs11191582 of *NT5C2* gene; and between combined pathology (BA+AH) and allele A of rs1010461 of *ANG/RNASE4* genes, allele T of rs7038716 and allele T of rs7025144 of *AL160272.2* gene, and allele C of rs2022318 (intergenic region *ABTB2/CAT*) (Table 1).

Assessment of the regulatory potential of SNPs associated with bronchial asthma

Regulatory regions of the genome are characterized by the presence of modified histones (methylated H3K4me1, H3K4me3 and acetylated H3K27ac, H3K9ac), which are the "labels" of active promoters and enhancers in various cells, including blood cells and cells in target organs for BA and AH – lungs, blood vessels, heart, and brain (Table 2). Thus, the variants rs1927914 and rs1980616 associated with BA are located in the promoter region of *TLR4* gene and the enhancer of *SERPINA1* gene in blood cells (monocytes). In addition, histone modifications in the rs1980616 (*SERPINA1*) region are recorded in other cells and tissues, including the lungs (see Table 2).

Genetic variants can alter the transcription factors (TF) affinity and regulated genes expression. It was shown that the G allele of rs1927914, located in *TLR4* gene promoter, has a multidirectional effect on the binding of TFs of the Pou family (see Table 2). It is known that the Pou proteins, Oct-1 and Oct-2, activate the expression of interleukin-5 gene in the PER-117 cell line, by binding to the CLE0 promoter element (Thomas et al., 1999), which, in turn, causes airways inflammation and is considered as a target for BA therapy (Busse et al., 2019). In another study (Aneas et al., 2020), rs1888909 variant associated with BA and located at a distance of 2.3 Kb from the 5′-end of *IL33* gene was found to be associated with the expression of this gene in epithelial cells in respiratory tract and the protein level in blood plasma, through the differential affinity of Oct-1 (Pou2f1) with a risk allele for this disease. It is possible that a similar mechanism is responsible for the association between the G allele of rs1927914 located 1.7 Kb upstream of *TLR4* gene and BA.

The C allele of rs1980616 of *SERPINA1* gene is associated with an increase in the affinity of TF Klf4 involved in inflammatory reactions, airway remodeling, and control of Th2-type cellular response (Tussiwand et al., 2015). In mice, in suppressor cells of myeloid origin and in epithelial cells of the respiratory tract, after exposure to the allergen,

Table 1. Genetic variants associated with the development of BA, AH and their combination

Disease	Gene; SNP	Allele associated with the disease: frequency in patients/healthy people, %	OR (95 % CI); *p*
Bronchial asthma	*TLR4*; rs1927914	G: 35.7/29.1	1.49 (1.02–2.18); 0.0397
	SERPINA1; rs1980616	C: 83.0/75.7	1.54 (1.01–2.34); 0.0204
Arterial hypertension	Intergenic region *ATXN2/BRAP*; rs11065987	A: 63.7/43.6	1.67 (1.15–2.45); 0.0076
	CARD8; rs11669386	A: 56.4/48.6	1.50 (1.01–2.21); 0.0433
	IL2RB; rs2284033	G: 50.0/46.0	1.49 (1.03–2.17); 0,0359
	NT5C2; rs11191582	G: 93.2/88.9	2.07 (1.03–4.16); 0.0419
Bronchial asthma + arterial hypertension (comorbid phenotype)	*ANG*, *RNASE4*; rs1010461	A: 62.1/50.0	1.71 (1.21–2.42); 0.0024
	AL160272.2; rs7038716	T: 28.9/20.4	1.74 (1.13–2.68); 0.0113
	AL160272.2; rs7025144	T: 28.1/21.7	1.60 (1.04–2.48); 0.0343
	Intergenic region *ABTB2/CAT*; rs2022318	C: 50.0/39.2	1.43 (1.01–2.04); 0.0444

Klf4 gene expression increased in the direct proportion to the severity of allergic asthma (Nimpong et al., 2017). Among the other variants studied in this study, rs1980616 has the highest regulatory potential with the rank of 2b and the scale of 0.83 (see Table 2).

According to the GTEx Portal and Blood eQTL browser, all studied SNPs are eQTLs and affect nearby genes expression (cis-eQTL) as well as the expression of those located at a significant distance (trans-eQTL) in different tissues. According to the up-to-date information, rs1927914 and rs1980616 associated with BA are connected with changes in the expression of *TLR4* and *SERPINA1* genes, respectively. The GG genotype of rs1927914 is associated with an increase in the expression of *TLR4* gene in whole blood, arteries, and left ventricular myocardium and with a decrease in the expression of this gene in brain tissues (cerebellum), and the CC genotype of rs1980616 is associated with an increase in the expression of *SERPINA1* gene in whole blood (see Table 2).

The rs1927914 (*TLR4*) variant associated with BA is associated in this study with the severity of BA in Chinese population (Zhang et al., 2011). This variant is included in an LD of 13 rSNPs, of which rs2737190 is associated with chronic obstructive disease development and pulmonary tuberculosis with a multidirectional effect (https://www.disgenet.org/browser/2/1/1/rs2737190/).

In the present work, the association of rs1980616 of *SERPINA1* gene with the development of BA was established for the first time. Previous studies have shown that chromosomal region 14q23-q32 is associated with blood IgE levels, bronchial hyperreactivity, and other signs of allergic inflammation in different populations (Malerba et al., 2001). The role of the proteins coded by the serpin family genes in the pathogenesis of BA is unknown; however, they can exhibit a protective effect by inhibiting endogenous proteases associated with inflammatory response, which merits further investigation.

Assessment of the regulatory potential of SNPs associated with arterial hypertension

Histone proteins modifications for the rs2284033 variant localized in an intron of *IL2RB* gene are recorded in peripheral blood leukocytes (T and NK cells) and for rs11191582 localized in an intron of *NT5C2* gene in a wide range of cells and tissues, including blood cells, lungs, vessels, heart, brain (see Table 2).

Single nucleotide variants associated with AH affect the change in TF affinity. Allele A of rs11065987 localized in the intergenic region *ATXN2*/*BRAP* is associated with a decrease in the TF Mrg1 and Hoxa9 binding affinity (see Table 2). Previously, in patients with hypertension, peripheral blood CD34+ hematopoietic stem cells showed a decrease in *HOXA9* gene expression, which may be associated with a decrease in circulating endothelial progenitor cells and impaired neovascularization and vascular damage repair (Pirro et al., 2007).

Allele A of the intron variant rs11669386 of the *CARD8* gene affects the increase in the affinity of several TFs, including factors of the Irf and Pou3f2 families (see Table 2). In experimental studies of model animals, it was shown that the regulatory factor interferon 1 (Irf) plays a leading role in the regulation of cardiac remodeling during pressure overload (Jiang et al., 2014).

Allele G of rs2284033 of *IL2RB* gene affects a decrease in affinity for TF RORalpha1 and an increase in RXR/LXR (see Table 2). No association of RORalpha1 and RXR/LXR with hypertension was found in the available scientific literature; however, it was shown that adenovirus-mediated overexpression of RORalpha1 suppresses TNF-alpha-induced expression of adhesion molecules VCAM-1 and ICAM-1 in umbilical vein endothelial cells (Migita et al., 2004). LXR/RXR family transcription factors regulate inflammation, cholesterol homeostasis, lipid, and glucose metabolism. By modulating the components of the renin-angiotensin-aldosterone system

Table 2. Regulatory potential of SNPs associated with BA, AH and their combination

SNP ID/gene localization/ predisposing allele (diseases)	rSNPBase	HaploReg		RegulomeDB	GTEx Portal, Blood eQTL browser	DisGeNET
	Intersection of SNPs with regulatory elements	Regulation type (number of tissues)	TF binding (allele, direction of TF affinity change)	Rank*/scale**	Genotype, direction of expression change, gene (tissue)	Number of associations with diseases total/BA, AH
rs1927914/1.7 Kb from the 5'-end (promoter) *TLR4*/**G** (BA)	–	HM(1), TF(6)	**G**↑ Arid5b; Pou1f1; Pou2f2_known5; Pou3f2; Pou3f3. **G**↓ Pou2f2_known6,9; Pou6f1	6/0.0	**GG** ↑ *TLR4* (TA, LV, WB) and ↓ *TLR4* (BR)	14/BA
rs1980616/intron *SERPINA1*/**C** (BA)	–	HM(5), DNAse (2), TF(4), BP(1)	**C**↑ Klf4, Pbx3, SP2; **C**↓ p300	2b/0.83	**CC**↑ *SERPINA1* (WB)	–
rs11065987/ intergenic region *ATXN2* and *BRAP*/**A** (AH)	Open chromatin regions	HM(4), DNAse (2), TF(2)	**A**↓ Mrg1, Hoxa9	3a/0.60	**AA**↓ *ALDH2* (A); **AA**↑ *ADAM1B* (ATR), *NAA25* (BR), *ALDH2* (WB)	17/–
rs2284033/intron *IL2RB*/**G** (AH)	TF	HM(1), DNAse (2), TF(3)	**G**↓ RORalpha1 **G**↑ RXR, LXR	4/0.61	**GG**↓ *ELFN2* (L)	7/BA
rs11191582/intron *NT5C2*/**G** (AH)	Circular RNA	HM(31), DNAse (30), TF(2), BP(1)	**G**↓ HMG-IY 2, Nanog	1f/0.55	**GG**↓ *CALHM2* (A), *MARCKSL1P1* (A, L, WB, CA), *NT5C2* (WB); **GG**↑ *ARL3* (L), *SFXN2* (L)	2/–
rs11669386/intron *CARD8*/**A** (AH)	Circular RNA	TF(4)	**A**↑ Irf, Nkx6-1, Pax-6, Pou3f2	1f/0.22	**AA**↑ *CCDC114* (TA, WB, L), *CARD8* (TA, BR)	–
rs7038716/intron *AL160272.2*/**T** (BA+AH)	Long non-coding RNA	TF(2)	**T**↓ Dobox4	7/0.18	**TT**↓ *TLR4* (WB)	–
rs7025144/intron *AL160272.2*/**T** (BA+AH)	–	HM(9), DNAse (5)	–	4/0.61	**TT**↓ *TLR4* (WB)	–
rs1010461/intron *ANG*, *RNASE4*/**A** (BA+AH)	Circular RNA	DNAse (4)	–	1f/0.55	**AA**↓ *ANG* (TA, A), **AA**↑ *RNASE4* (LV, L, WB, BR, TA), *ANG* (WB)	–
rs2022318/ intergenic region *ABTB2* and *CAT*/**C** (BA+AH)	–	HM(2), TF(4)	**C**↑ HEN1, Zfp691; **C**↓ ZNF263, p53	1f/0.22	**CC**↓ *CAT* (A, CA, TA, LV, L, WB), *ABTB2* (WB)	–

Note. A – aorta; CA – coronary arteries; TA – tibial artery; LV – left ventricle; WB – whole blood; Br – brain; Atr – atrium; L – lungs; HM – histone modifications; DNAse – DNAse I-hypersensitive sites; TF – transcription factors; BP – bound proteins. * Ranking of the regulatory potential: 1f – eQTL + binding TF/DNAse I-hypersensitive site; 2b – eQTL + binding of TF + other regulatory elements + DNAse I-hypersensitive region; 3a – binding of TF + any TF motif + DNAse I-hypersensitive site; 4 – binding of TF + DNAse I-hypersensitive site; 6 – any TF motive; 7 – others. ** The closer the scale value is to one, the higher the SNP regulatory potential.

LXR/RXR, they reduce peripheral vascular resistance and blood pressure (Cannon et al., 2016).

Allele G of rs11191582 of *NT5C2* gene is associated with a decrease in affinity for the transcription factors HMG-IY and Nanog (see Table 2). Previously, it was shown that HMG-IY is involved in the chromatin structure formation and regulation of transcription of many genes. Changes in HMG-IY levels have been shown to affect promoter activity and *IL2* gene expression in human cell lines (Himes et al., 2000). An important role of T-cells has been shown in the pathogenesis of hypertension, namely the activation of Th1 lymphocytes which are producers of IL2. Increased production of IL2 along with the cytokines IL1β, IL6, TNFα, and IFNγ promotes vascular inflammation and hypertension development (Schiffrin, 2014).

The genetic variants associated with hypertension are eOTLs. It was shown that rs11065987 (*ATXN2*/*BRAP*) is a cis- and trans-eQTL and affects the *ALDH2*, *ADAM1B*, and *NAA25* genes activity in various cells and tissues (see Table 2). The AA genotype of rs11065987 (*ATXN2*/*BRAP*) is associated with a decrease in *ALDH2* gene expression in aorta and an increase in its expression in whole blood leukocytes, as well as with an increase in *ADAM1B* gene expression in heart atrium and *NAA25* gene in basal ganglia. Previously, a link with the cardiovascular disease and hypertension for some of these genes and their protein products was established. Thus, *BRAP* gene variants were found to be associated with the risk of carotid atherosclerosis (Liao et al., 2011); protein products of genes *ATXN2* and *SH2B3* were involved in AH development (Siedlinski et al., 2020); rs671 of *ALDH2* gene was associated with a decrease in the risk of AH development (Mei et al., 2020). For the variant rs11065987, an association with many pathological conditions was registered, including the level of lipids in blood (Willer et al., 2013), body mass index (Locke et al., 2015), ischemic stroke, ischemic heart disease (Dichgans et al., 2014) and AH (Levy et al., 2009).

The rs 11669386 variant is associated with *CARD8* gene functional activity, its SNPs are significant for the formation of susceptibility to hypertension, aortic aneurysm, and stroke (Zhao et al., 2016). The AA genotype of rs11669386 is associated with an increase in *CARD8* gene expression in arteries and brain, as well as with an increase in *CCDC114* gene expression in tibial artery, lungs, and whole blood (see Table 2). Proteins CARD8 and NLRP3 together control the activity of inflammatory caspase-1 and are involved in the regulation of caspase-1-mediated activation of *IL1B* in mouse macrophages, dendritic cells, and macrophages of human peripheral blood (Abdelaziz et al., 2015). As shown in a Kawasaki disease murine model, caspase-1 and IL-1B are important inflammatory cytokines in coronary artery disease development (Lee et al., 2012).

The rs2284033 variant has a pronounced effect on *IL2RB* gene expression in peripheral blood leukocytes (Westra et al., 2013), and the GG genotype is associated with a decrease in *ELFN2* gene expression in lung tissues (see Table 2). Earlier, the association of rs2284033 with BA in total was established, which, however, diminishes when considering individual phenotypes associated with the disease age of onset (Moffatt et al., 2010). In addition, this SNP is in LD with 10 rSNPs, one of which (rs228953) has been shown to be associated with eosinophil counts in blood and the risk of BA (Han et al., 2020).

In our study, the G allele of rs2284033 (*IL2RB*) is associated with AH (see Table 1). As mentioned above, the role of immune system and inflammation in AH pathogenesis is currently well established. Increased expression of *IL2RB* gene was detected in leukocytes in patients with AH (Huan et al., 2015). In this regard, *IL2RB* gene polymorphism may predispose, in combination with other factors, to the development of AH.

The GG genotype of rs11191582 (*NT5C2*) is associated with a decrease in *CALHM2* and *MARCKSL1P1* genes expression in aorta, *NT5C2* and *MARCKSL1P1* genes in whole blood cells, and an increase in *ARL3* and *SFXN2* genes expression in the lung tissue (see Table 2). The protein product of *CALHM2* gene modulates calcium homeostasis, the maintenance of which is necessary for the performance of cellular functions, such as contraction, proliferation, migration, and growth, the disruption of which is associated with the development of cardiovascular diseases. It was previously shown that rs11191582 is localized in the same LD block with rSNPs (rs11191548, rs11191559, rs11191580, rs11191593, rs12413409, and rs943037), for which a relationship with blood pressure regulation and coronary artery disease development was established (Matsunaga et al., 2020). In the nucleotide sequence of *ARL3* gene, SNPs were identified that are located in different LD blocks and were associated with blood pressure level. This suggested that in the 10q24.32 locus, where *ARL3* and *SFXN2* genes are located, numerous "causal" genetic variants of susceptibility to AH may be situated (Li et al., 2017).

Assessment of the regulatory potential of SNPs associated with a comorbid phenotype – BA and AH

Genetic variants associated with BA+AH phenotype are located in DNAse I hypersensitive sites (rs1010461, rs7025144), in the enhancer region (rs7025144, rs2022318) and affect TF affinity (rs7038716, rs2022318) (see Table 2).

The rs1010461 variant localized in intron region of *ANG* and *RNASE4* genes does not change the TF affinity, but its relationships with the functional activity of these genes in blood cells, arteries, heart, lungs, and brain have been shown (see Table 2). It was demonstrated that the AA rs1010461 genotype is associated with a decrease in *ANG* gene expression in tibial artery and aorta, but an increase in the expression level of this gene in whole blood cells. The AA genotype is associated with an increase in the expression level of ribonuclease 4 (*RNASE4*) gene in whole blood leukocytes, tissues of the left ventricle, lungs, brain, and tibial artery.

RNASE4 gene has the same promoter and is co-expressed with angiogenin gene (*ANG*). However, the role of *RNASE4* in the development of cardiovascular and bronchopulmonary diseases is unknown. In turn, angiogenin is a potent inducer of blood vessel formation and, possibly, can be used as a serum marker for the development of cardiovascular diseases (Yu et al., 2018). During the period of an asthmatic attack, an increase in vascular endothelial growth factor and angiogenin is recorded in the patients' sputum, which is significantly reduced after corticosteroid therapy (Abdel-Rahman et al., 2006). It is possible that the treatment of asthma with corticosteroids can act as an unfavorable factor for the development of subsequent arterial hypertension in individuals predisposed to this pathology.

Allele T of rs7038716 (*AL160272.2*) is associated with a decrease in affinity for TF Dobox4. Genotypes TT rs7038716 and TT rs7025144 are associated with a decrease in *TLR4* gene functional activity in whole blood leukocytes (see Table 2). In the present study, BA is associated with alleles and genotypes associated with an increase in functional activity of *TLR4* gene, while BA and AH comorbidity is associated with genotypes characterized by a decreased *TLR4* gene expression in blood cells. This result requires further research and explanation.

Allele C rs2022318 (*ABTB2*/*CAT*) is associate p53. P53 expression is increased with an increase in affinity for TF HEN1 and Zfp691, but with a decrease in affinity for ZNF263 and

ased in response to DNA damage, hypoxia, and oxidative stress, which provides cellular protection. New evidence is emerging to support the protective effect of p53 in inflammatory processes in the lungs. In particular, it has been shown that p53 deficiency in vascular lung endothelial cells is associated with severe respiratory disorders (Uddin, Barabutis, 2020). For ZNF263, an association with atherosclerosis was shown through the regulation of *TGFB1* gene expression, the protein product of which is actively involved in various diseases' pathogenesis (Dhaouadi et al., 2014).

The CC genotype rs2022318 is associated with a decrease in the expression of *ABTB2* gene in whole blood cells and catalase gene (*CAT*) in tissues of aorta, left ventricle, lungs, coronary arteries, and whole blood cells (see Table 2). Catalase is an important antioxidant enzyme the decreased activity of which is seen in many diseases associated with oxidative stress. It has been shown that SNPs localized in the catalase gene promoter are associated with the development of cardiometabolic diseases (Doğan et al., 2019) and bronchial asthma (Taniguchi et al., 2014). There are no data on the relationship of *ABTB2* gene polymorphism, its functional activity, or protein product with AH and BA in the available scientific literature.

Conclusion

Thus, SNPs associated with the studied phenotypes – BA, AH, and BA+AH, are regulatory (rSNP), localized in actively transcribed regions of the genome, are eQTLs, and affect functional activity of various genes in blood cells and target organs of the diseases – lungs, blood vessels, heart, and brain.

The relationship of rs1927914 and rs1980616 with BA can be explained by: the effect of the G allele rs1927914 on the change in the affinity of transcription factors of the Pou family and the relationship of the GG genotype with an increase in the level of *TLR4* gene expression in blood cells; the relationship of the C allele rs1980616 with an increase in the affinity of the transcription factor Klf4, which is involved in inflammation, remodeling of the airways and control of Th2-type cell response, as well as the association of the CC genotype with increased expression of the *SERPINA1* gene in blood cells.

The association of rs11065987, rs2284033, rs11191582, and rs11669386 with AH may be based on the following: the effect of the A allele rs11065987 on a decrease in the affinity of the Hoxa9 transcription factor and a decrease in the pool of endothelial progenitor cells, which disrupts neovascularization and changes in *ALDH2* expression of vascular damage, and in blood cells and blood vessels; the relationship of the A allele rs11669386 with an increase in the affinity of the transcription factor Irf, which plays a leading role in the regulation of cardiac remodeling under pressure overload, and the Pou3f2 factor associated with the development of left ventricular remodeling in hypertension, the AA genotype is associated with an increase in *CARD8* gene expression in blood vessels, the protein product of which is included in the composition of the inflamosome and is involved in the regulation of inflammatory reactions in various diseases, including cardiovascular ones; the effect of the G rs2284033 allele on a decrease in affinity for TF RORalpha1, which suppresses TNF-alpha-induced expression of adhesion molecules VCAM-1 and ICAM-1 in endothelial cells and the relationship of the GG rs11191582 genotype with a change in the level of expression of *NT5C2*, *ARL3*, and *SFXN2* genes in blood/vascular cells and heart.

The association of rs7038716, rs7025144, rs1010461, and rs2022318 with the comorbid phenotype of BA and AH may be due to: the relationship of the TT and TT genotypes rs7038716 and rs7025144 with a decrease in *TLR4* gene functional activity in blood cells; the relationship of the AA rs1010461 genotype with a change in *ANG* gene expression in blood and vascular cells, the protein product of which plays an important role in the pathogenesis of both AH and BA; the influence of the C allele rs2022318 on the increase in the affinity of the transcription factor p53, the expression of which increases during inflammatory processes in the lungs and the connection of the CC genotype with a decrease in *CAT* gene expression in blood cells, lungs, blood vessels and heart, the decrease in the activity of which is observed in diseases associated with oxidative stress.

References

Abdel-Rahman A.M., El-Sahrigy S.A., Bakr S.I. A comparative study of two angiogenic factors: vascular endothelial growth factor and angiogenin in induced sputum from asthmatic children in acute attack. *Chest.* 2006;129(2):266-271. DOI 10.1378/chest.129.2.266.

Abdelaziz D.H.A., Khalil H., Cormet-Boyaka E., Amer A.O. The cooperation between the autophagy machinery and the inflammasome to implement an appropriate innate immune response: do they regulate each other? *Immunol. Rev*. 2015;265(1):194-204. DOI 10.1111/imr.12288.

Aneas I., Decker D.C., Howard C.L., Sobreira D.R., Sakabe N.J., Blaine K.M., Stein M.M., … Gomez-Skarmeta J.L., Schoetler N., Ober C., Sperling A.I., Nobrega M.A. Asthma-associated variants induce IL33 differential expression through a novel regulatory region. *bioRxiv.* 2020.09.09.290098. DOI 10.1101/2020.09.09.290098.

Bragina E.Y., Goncharova I.A., Freidin M.B., Zhalsanova I.Z., Gomboeva D.E., Nemerov E.V., Puzyrev V.P. Analysis of haplotypes of *CAT*, *TLR4*, and *IL10* genes in bronchial asthma patients comorbid with arterial hypertension. *Sibirskiy Nauchnyy Meditsinskiy Zhurnal = Siberian Scientific Medical Journal*. 2019;39(6):55-64. DOI 10.15372/SSMJ20190607. (in Russian)

Bragina E.Y., Goncharova I.A., Garaeva A.F., Zhalsanova I.Z., Gomboeva D.E., Freidin M.B., Nemerov E.V., Babovskaya A.A., Karpov A.B., Semenova Y.V., Saik O.V., Ivanisenko V.A., Zolotareva O.I., Hofestaedt R., Dosenko V.E. Molecular relationships between bronchial asthma and hypertension as comorbid diseases. *J. Integr. Bioinform.* 2018;15(4):20180052. DOI 10.1515/jib-2018-0052.

Busse W.W. Biological treatments for severe asthma: A major advance in asthma care. *Allergol. Int.* 2019;68(2):158-166. DOI 10.1016/j.alit.2019.01.004.

Cannon M.V., van Gilst W.H., de Boer R.A. Emerging role of liver X receptors in cardiac pathophysiology and heart failure. *Basic Res. Cardiol*. 2016;111(1):3. DOI 10.1007/s00395-015-0520-7.

Dhaouadi N., Li G.-Y., Feugier P., Gustin M.-P., Dab H., Kacem K., Bricca G., Cerutti C. Computational identification of potential transcriptional regulators of TGF-ß1 in human atherosclerotic arteries. *Genomics*. 2014;103(5-6):357-370. DOI 10.1016/j.ygeno.2014.05.001.

Dichgans M., Malik R., König I.R., Rosand J., Clarke R., Gretarsdottir S., Thorleifsson G., … Roberts R., Markus H.S., Samani N.J., Farrall M., Schunkert H., METASTROKE Consortium; CARDIoGRAM Consortium; C4D Consortium; International Stroke Genetics Consortium. Shared genetic susceptibility to ischemic stroke and

coronary artery disease: a genome-wide analysis of common variants. *Stroke.* 2014;45(1):24-36. DOI 10.1161/STROKEAHA.113.002707.

Doğan A., Özşensoy Y., Türker F.S. MnSOD, CAT and GPx-3 genetic polymorphisms in coronary artery disease. *Mol. Biol. Rep*. 2019; 46(1):841-845. DOI 10.1007/s11033-018-4539-3.

Han Y., Jia Q., Jahani P.S., Hurrell B.P., Pan C., Huang P., Gukasyan J., Woodward N.C., Eskin E., Gilliland F.D., Akbari O., Hartiala J.A., Allayee H. Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nat. Commun*. 2020;15;11(1):1776. DOI 10.1038/s41467-020-15649-3.

Himes S.R., Reeves R., Attema J., Nissen M., Li Y., Shannon M.F. The role of high-mobility group I(Y) proteins in expression of IL-2 and T cell proliferation. *J. Immunol*. 2000;164(6):3157-3168. DOI 10.4049/jimmunol.164.6.3157.

Huan T., Esko T., Peters M.J., Pilling L.C., Schramm K., Schurmann C., Chen B.H., … Prokisch H., Völker U., van Meurs J.B., Ferrucci L., Levy D. A meta-analysis of gene expression signatures of blood pressure and hypertension. *PLoS Genet*. 2015;11(3):e1005035. DOI 10.1371/journal.pgen.1005035.

Jiang D.S., Li L., Huang L., Gong J., Xia H., Liu X., Wan N., Wei X., Zhu X., Chen Y., Chen X., Zhang X.D., Li H. Interferon regulatory factor 1 is required for cardiac remodeling in response to pressure overload. *Hypertension.* 2014;64(1):77-86. DOI 10.1161/HYPERTENSIONAHA.114.03229.

Lee Y., Schulte D.J., Shimada K., Chen S., Crother T.R., Chiba N., Fishbein M.C., Lehman T.J., Arditi M. Interleukin-1β is crucial for the induction of coronary artery inflammation in a mouse model of Kawasaki disease. *Circulation.* 2012;125(12):1542-1550. DOI 10.1161/CIRCULATIONAHA.111.072769.

Levy D., Ehret G.B., Rice K., Verwoert G.C., Launer L.J., Dehghan A., Glazer N.L., ... Gudnason V., Larson M.G., Chakravarti A., Psaty B.M., van Duijn C.M. Genome-wide association study of blood pressure and hypertension. *Nat. Genet*. 2009;41(6):677-687. DOI 10.1038/ng.384.

Li C., Kim Y.K., Dorajoo R., Li H., Lee I.T., Cheng C.Y., He M., ... Wu J.Y., Lin X., Tai E.S., Kim B.J., Kelly T.N. Genome-wide association study meta-analysis of long-term average blood pressure in East Asians. *Circ. Cardiovasc. Genet*. 2017;2:e001527. DOI 10.1161/CIRCGENETICS.116.001527.

Liao Y.C., Wang Y.S., Guo Y.C., Ozaki K., Tanaka T., Lin H.-F., Chang M.-H., Chen K.-C., Yu M.-L., Sheu S.-H., Juo S.-H. *BRAP* activates inflammatory cascades and increases the risk for carotid atherosclerosis. *Mol. Med*. 2011;17(9-10):1065-1074. DOI 10.2119/molmed.2011.00043.

Locke A.E., Kahali B., Berndt S.I., Justice A.E., Pers T.H., Day F.R., Powell C., ... North K.E., Ingelsson E., Hirschhorn J.N., Loos R.J.F., Speliotes E.K. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518(7538):197-206. DOI 10.1038/nature14177.

Malerba G., Patuzzo C., Trabetti E., Lauciello M.C., Galavotti R., Pescollderungg L., Whalen M.B., Zanoni G., Martinati L.C., Boner A.L., Pignatti P.F. Chromosome 14 linkage analysis and mutation study of 2 serpin genes in allergic asthmatic families. *J. Allergy Clin. Immunol.* 2001;107(4):654-658. DOI 10.1067/mai.2001.113865.

Matsunaga H., Ito K., Akiyama M., Takahashi A., Koyama S., Nomura S., Ieki H., ... Murakami Y., Akazawa H., Kubo M., Kamatani Y., Komuro I. Transethnic meta-analysis of genome-wide association studies identifies three new loci and characterizes population-specific differences for coronary artery disease. *Circ. Genom. Precis. Med*. 2020;13(3):e002670. DOI 10.1161/CIRCGEN.119.002670.

Mei X.F., Hu S.D., Liu P.F., Li F., Zhou X.Y., Zhou Y.F., Chen T. *ALDH2* gene rs671 polymorphism may decrease the risk of essential hypertension. *Int. Heart J*. 2020;61(3):562-570. DOI 10.1536/ihj.19-259.

Migita H., Satozawa N., Lin J.-H., Morser J., Kawai K. RORalpha1 and RORalpha4 suppress TNF-alpha-induced VCAM-1 and ICAM-1 expression in human endothelial cells. *FEBS Lett*. 2004;557(1-3): 269-274. DOI 10.1016/s0014-5793(03)01502-3.

Moffatt M.F., Gut I.G., Demenais F., Strachan D.P., Bouzigon E., Heath S., von Mutius E., Farrall M., Lathrop M., Cookson W., GABRIEL Consortium. A large-scale, consortium-based genome-wide association study of asthma. *N. Engl. J. Med*. 2010;363(13): 1211-1221. DOI 10.1056/NEJMoa0906312.

Moore W.C., Hastie A.T., Li X., Li H., Busse W.W., Jarjour N.N., Wenzel S.E., Peters S.P., Meyers D.A., Bleecker E.R. National Heart, Lung, and Blood Institute's Severe Asthma Research Program. Sputum neutrophil counts are associated with more severe asthma phenotypes using cluster analysis. *J. Allergy Clin. Immunol.* 2014;133(6):1557-63.e5. DOI 10.1016/j.jaci.2013.10.011. Epub 2013 Dec 9.

Nimpong J.A., Gebregziabher W., Singh U.P., Nagarkatti P., Nagarkatti M., Hodge J., Liu C., Fan D., Ai W. Deficiency of *KLF4* compromises the lung function in an acute mouse model of allergic asthma. *Biochem. Biophys. Res. Commun*. 2017;493(1):598-603. DOI 10.1016/j.bbrc.2017.08.146.

Pirro M., Schillaci G., Menecali C., Bagaglia F., Paltriccia R., Vaudo G., Mannarino M.R., Mannarino E. Reduced number of circulating endothelial progenitors and HOXA9 expression in CD34+ cells of hypertensive patients. *J. Hypertens.* 2007;25(10):2093-2099. DOI 10.1097/HJH.0b013e32828e506d.

Saik O.V., Demenkov P.S., Ivanisenko T.V., Bragina E.Y., Freidin M.B., Goncharova I.A., Dosenko V.E., Zolotareva O.I., Hofestaedt R., Lavrik I.N., Rogaev E.I., Ivanisenko V.A. Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks. *BMC Med. Genomics*. 2018;11(Suppl. 1): 15. DOI 10.1186/s12920-018-0331-4.

Schiffrin E.L. Immune mechanisms in hypertension and vascular injury. *Clin. Sci. (Lond.).* 2014;126(4):267-274. DOI 10.1042/CS20130407.

Siedlinski M., Jozefczuk E., Xu X., Teumer A., Evangelou E., Schnabel R.B., Welsh P., Maffia P., Erdmann J., Tomaszewski M., Caulfield M.J., Sattar N., Holmes M.V., Guzik T.J. White blood cells and blood pressure: A Mendelian randomization study. *Circulation.* 2020;141(16):1307-1317. DOI 10.1161/CIRCULATIONAHA.119.045102.

Su X., Ren Y., Li M., Zhao X., Kong L., Kang J. Prevalence of comorbidities in asthma and nonasthma patients: a meta-analysis. *Medicine (Baltimore)*. 2016;95(22):e3459. DOI 10.1097/MD.0000000000003459.

Taniguchi N., Konno S., Isada A., Hattori T., Kimura H., Shimizu K., Maeda Y., Makita H., Hizawa N., Nishimura M. Association of the CAT-262C>T polymorphism with asthma in smokers and the non-emphysematous phenotype of chronic obstructive pulmonary disease. *Ann. Allergy Asthma Immunol.* 2014;113(1):31-36.e2. DOI 10.1016/j.anai.2014.04.012.

Thomas M.A., Mordvinov V.A., Sanderson C.J. The activity of the human interleukin-5 conserved lymphokine element 0 is regulated by octamer factors in human cells. *Eur. J. Biochem*. 1999;265(1):300-307. DOI 10.1046/j.1432-1327.1999.00732.x.

Tussiwand R., Everts B., Grajales-Reyes G.E., Kretzer N.M., Iwata A., Bagaitkar J., Wu X., Wong R., Anderson D.A., Murphy T.L., Pearce E.J., Murphy K.M. Klf4 expression in conventional dendritic cells is required for T helper 2 cell responses. *Immunity.* 2015; 42(5):916-928. DOI 10.1016/j.immuni.2015.04.017.

Uddin M.A., Barabutis N. P53 in the impaired lungs. *DNA Repair (Amst.)*. 2020;95:102952. DOI 10.1016/j.dnarep.2020.102952.

van Arensbergen J., Pagie L., FitzPatrick V., de Haas M., Baltissen M., Comoglio F., van der Weide R., Teunissen H., Võsa U., Franke L., de Wit E., Vermeulen M., Bussemaker H., van Steensel B. Systematic identification of human SNPs affecting regulatory element activity. *bioRxiv.* 2018. DOI 10.1101/460402.

Weatherburn C.J., Guthrie B., Mercer S.W., Morales D.R. Comorbidities in adults with asthma: Population-based cross-sectional

analysis of 1.4 million adults in Scotland. *Clin. Exp. Allergy*. 2017; 47(10):1246-1252. DOI 10.1111/cea.12971.

Westra H.J., Peters M.J., Esko T., Yaghootkar H., Schurmann C., Kettunen J., Christiansen M.W., Fairfax B.P., … Teumer A., Frayling T.M., Metspalu A., van Meurs J.B.J., Franke L. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet*. 2013;45(10):1238-1243. DOI 10.1038/ng.2756.

Willer C.J., Schmidt E.M., Sengupta S., Peloso G.M., Gustafsson S., Kanoni S., Ganna A., ... Deloukas P., Kathiresan S., Mohlke K.L., Ingelsson E., Abecasis G.R. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* 2013;45(11):1274-1283. DOI 10.1038/ng.2797.

Yu D., Cai Y., Zhou W., Sheng J., Xu Z. The potential of angiogenin as a serum biomarker for diseases: systematic review and meta-analysis. *Dis. Markers.* 2018;2018:1984718. DOI 10.1155/2018/1984718.

Zhang Q., Qian F.H., Zhou L.F., Wei G.Z., Jin G.F., Bai J.L., Yin K.S. Polymorphisms in toll-like receptor 4 gene are associated with asthma severity but not susceptibility in a Chinese Han population. *J. Investig. Allergol. Clin. Immunol*. 2011;21(5):370-377.

Zhao X., Gu C., Yan C., Zhang X., Li Y., Wang L., Ren L., Zhang Y., Peng J., Zhu Z., Han Y. NALP3-inflammasome-related gene polymorphisms in patients with prehypertension and coronary atherosclerosis. *Biomed. Res. Int*. 2016;2016:7395627. DOI 10.1155/2016/7395627.

ORCID ID

I.A. Goncharova orcid.org/0000-0002-9527-7015
E.Yu. Bragina orcid.org/0000-0002-1103-3073
I.Zh. Zhalsanova orcid.org/0000-0001-6848-7749
M.B. Freidin orcid.org/0000-0002-1439-6259
M.S. Nazarenko orcid.org/0000-0002-0673-4094

Acknowledgements. The work was carried out within the framework of the State Assignment of the Ministry of Science and Higher Education of the Russian Federation No. 075-00603-19-00.

Conflict of interest. The authors declare no conflict of interest.

Received May 5, 2021. Revised July 7, 2021. Accepted July 7, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):864-873
DOI 10.18699/VJ21.100

Таксономический состав и биоразнообразие кишечного микробиома пациентов с синдромом раздраженного кишечника, язвенным колитом и бронхиальной астмой

А.Ю. Тикунов[1], А.Н. Швалов[2], В.В. Морозов[1], И.В. Бабкин[1], Г.В. Селедцова[3], И.О. Волошина[1], И.П. Иванова[3], А.В. Бардашева[1], В.В. Морозова[1], В.В. Власов[1], Н.В. Тикунова[1] ✉

[1] Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора Российской Федерации, р.п. Кольцово, Новосибирская область, Россия
[3] Научно-исследовательский институт фундаментальной и клинической иммунологии, Новосибирск, Россия
✉ tikunova@niboch.nsc.ru; arttik@ngs.ru

Аннотация. К настоящему времени показана ассоциация дисбаланса кишечной микробиоты с различными заболеваниями человека, включая не только патологии желудочно-кишечного тракта, но и нарушения иммунной системы. Однако, несмотря на значительный объем накопленных данных, многие ключевые вопросы до сих пор остаются без ответа. Описаны различия в микробных сообществах кишечника в зависимости от возраста больных, типа питания и региона проживания. Учитывая ограниченность данных о составе микробиоты кишечника при язвенном колите (ЯК) и синдроме раздраженного кишечника (СРК) у пациентов из регионов Сибири, а также отсутствие сведений о кишечной микробиоте больных бронхиальной астмой (БА), цель исследования – оценка биоразнообразия кишечного микробиома пациентов с СРК, ЯК и БА в сравнении с таковым здоровых добровольцев (ЗД). Проведена сравнительная оценка биоразнообразия и таксономической структуры микробиома содержимого кишечника пациентов с СРК, ЯК, БА и ЗД, определенных на основании 16S рРНК-последовательностей бактериальных генов. В четырех выборках доминировали последовательности типов Firmicutes и Bacteroidetes. Третьи по встречаемости во всех группах – последовательности типа Proteobacteria, членами которого являются патогенные и условно-патогенные бактерии. Последовательности типа Actinobacteria были в среднем четвертыми по встречаемости. Результаты показали наличие дисбиоза в образцах пациентов по сравнению со здоровыми участниками. Соотношение Firmicutes/Bacteroidetes в выборках СРК и ЯК уменьшилось относительно ЗД, а в группе БА – увеличилось. В образцах пациентов с заболеваниями кишечника (СРК и ЯК) обнаружено увеличение доли последовательностей типа Bacteroidetes и уменьшение доли последовательностей класса Clostridia, а также семейства Ruminococcaceae, но не Erysipelotrichaceae. В выборках СРК, ЯК и БА отмечено достоверно больше в сравнении с ЗД последовательностей Proteobacteria, включая *Methylobacterium*, *Sphingomonas*, *Parasutterella*, *Halomonas*, *Vibrio*, а также последовательности *Escherichia* и *Shigella*. В кишечном микробиоме взрослых пациентов с БА выявлено уменьшение доли последовательностей *Roseburia*, *Lachnospira*, *Veillonella*, однако доля последовательностей *Faecalibacterium* и *Lactobacillus* была такой же, как и у здоровых участников. Впервые получены данные о существенном увеличении доли последовательностей *Halomonas* и *Vibrio* в кишечном микробиоме пациентов с бронхиальной астмой.
Ключевые слова: микробиом; 16S рРНК-последовательности; язвенный колит; синдром раздраженного кишечника; бронхиальная астма.

Для цитирования: Тикунов А.Ю., Швалов А.Н., Морозов В.В., Бабкин И.В., Селедцова Г.В., Волошина И.О., Иванова И.П., Бардашева А.В., Морозова В.В., Власов В.В., Тикунова Н.В. Таксономический состав и биоразнообразие кишечного микробиома пациентов с синдромом раздраженного кишечника, язвенным колитом и бронхиальной астмой. *Вавиловский журнал генетики и селекции*. 2021;25(8):864-873. DOI 10.18699/VJ21.100

Taxonomic composition and biodiversity of the gut microbiome from patients with irritable bowel syndrome, ulcerative colitis, and asthma

A.Y. Tikunov[1], A.N. Shvalov[2], V.V. Morozov[1], I.V. Babkin[1], G.V. Seledtsova[3], I.O. Voloshina[1], I.P. Ivanova[3], A.V. Bardasheva[1], V.V. Morozova[1], V.V. Vlasov[1], N.V. Tikunova[1] ✉

[1] Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] State Research Center of Virology and Biotechnology "Vector", Rospotrebnadzor, Koltsovo, Novosibirsk region, Russia
[3] Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia
✉ tikunova@niboch.nsc.ru; arttik@ngs.ru

Abstract. To date, the association of an imbalance of the intestinal microbiota with various human diseases, including both diseases of the gastrointestinal tract and disorders of the immune system, has been shown. However, despite the huge amount of accumulated data, many key questions still remain unanswered. Given limited data on the composition of the

gut microbiota in patients with ulcerative colitis (UC) and irritable bowel syndrome (IBS) from different parts of Siberia, as well as the lack of data on the gut microbiota of patients with bronchial asthma (BA), the aim of the study was to assess the biodiversity of the gut microbiota of patients with IBS, UC and BA in comparison with those of healthy volunteers (HV). In this study, a comparative assessment of the biodiversity and taxonomic structure of gut microbiome was conducted based on the sequencing of 16S rRNA genes obtained from fecal samples of patients with IBS, UC, BA and volunteers. Sequences of the Firmicutes and Bacteroidetes types dominated in all samples studied. The third most common in all samples were sequences of the Proteobacteria type, which contains pathogenic and opportunistic bacteria. Sequences of the Actinobacteria type were, on average, the fourth most common. The results showed the presence of dysbiosis in the samples from patients compared to the sample from HVs. The ratio of Firmicutes/Bacteroidetes was lower in the IBS and UC samples than in HV and higher the BA samples. In the samples from patients with intestinal diseases (IBS and UC), an increase in the proportion of sequences of the Bacteroidetes type and a decrease in the proportion of sequences of the Clostridia class, as well as the Ruminococcaceae, but not Erysipelotrichaceae family, were found. The IBS, UC, and BA samples had significantly more Proteobacteria sequences, including *Methylobacterium*, *Sphingomonas*, *Parasutterella*, *Halomonas*, *Vibrio*, as well as *Escherichia* spp. and *Shigella* spp. In the gut microbiota of adults with BA, a decrease in the proportion of *Roseburia*, *Lachnospira*, *Veillonella* sequences was detected, but the share of *Faecalibacterium* and *Lactobacillus* sequences was the same as in healthy individuals. A significant increase in the proportion of *Halomonas* and *Vibrio* sequences in the gut microbiota in patients with BA has been described for the first time.
Key words: microbiome; 16S rRNA sequences; ulcerative colitis; irritable bowel syndrome; bronchial asthma.

For citation: Tikunov A.Y., Shvalov A.N., Morozov V.V., Babkin I.V., Seledtsova G.V., Voloshina I.O., Ivanova I.P., Bardasheva A.V., Morozova V.V., Vlasov V.V., Tikunova N.V. Taxonomic composition and biodiversity of the gut microbiome from patients with irritable bowel syndrome, ulcerative colitis, and asthma. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8):864-873. DOI 10.18699/VJ21.100. (in Russian)

Введение

К настоящему времени накоплено достаточно данных о микробных сообществах кишечника человека и показана ассоциация дисбаланса кишечной микробиоты с различными патологическими состояниями, в том числе не только заболеваниями желудочно-кишечного тракта, но и нарушениями иммунной системы (O'Hara, Shanahan, 2006). Однако, несмотря на значительный объем информации, многие ключевые вопросы остаются без ответа. Так, до сих пор неизвестно, являются ли такие заболевания кишечника, как язвенный колит (ЯК) и синдром раздраженного кишечника (СРК), результатом нарушенного иммунного ответа на нормальную микробиоту или служат проявлением нормального иммунного ответа на нарушения в микрофлоре кишечника (Cheng, Fisher, 2017).

Патогенез этих заболеваний также не вполне ясен: вероятно, механизм развития имеет сложную природу и опосредован нарушениями кишечной микробиоты, генетической предрасположенностью и экологическими факторами (Shen et al., 2018). Известно лишь, что при ЯК и СРК снижено биоразнообразие микробиоты кишечника (Machiels et al., 2014; Dubinsky, Braun, 2015). При ЯК отмечено уменьшение количества представителей типов Bacteroidetes и Firmicutes; для последнего зарегистрировано уменьшение встречаемости последовательностей *Roseburia hominis* и *Faecalibacterium prausnitzii* (отряд Clostridia, сем. Lachnospiraceae и Ruminococcaceae соответственно) (Machiels et al., 2014; Тикунов и др., 2020). Особенно заметно снижение *Akkermansia muciniphila* (тип Verrucomicrobia), в норме достигающей 1–5 % всего бактериального сообщества (Manichanh et al., 2012; Bajer et al., 2017). Одновременно в микробиоте пациентов с ЯК увеличено количество представителей Actinomycetes и Proteobacteria, а среди протеобактерий часто выявляют *Helicobacter* spp., *Salmonella* spp., *Yersinia* spp. и энтероинвазивные *Escherichia coli* (Saebo et al., 2005; Gradel et al., 2009; Sonnenberg, Genta, 2012; Shen et al., 2018; Тикунов и др., 2020). В случае СРК, как и при ЯК, в микробиоте кишечника количество представителей типа Proteobacteria увеличено, а число представителей Actinomycetes, наоборот, уменьшено (Bennet et al., 2015; Su et al., 2018).

Соотношение основных представителей микробного сообщества кишечника человека, Firmicutes/Bacteroidetes, при СРК может и увеличиваться, и уменьшаться (Tana et al., 2010; Rajilić-Stojanović et al., 2011; Jeffery et al., 2012; Jalanka-Tuovinen et al., 2014; Pozuelo et al., 2015; Tap et al., 2017). Такое несовпадение данных может быть связано как с динамичностью кишечного микробного сообщества у отдельного индивидуума, так и высоким биоразнообразием кишечной микробиоты у людей не только при различном состоянии здоровья, но и в зависимости от возраста, региона проживания и особенностей питания (Fujimura et al., 2010; Qin et al., 2010; Donaldson et al., 2016). В связи с этим изучение ассоциации особенностей микробиоты с различными заболеваниями человека – одно из наиболее актуальных направлений современных биомедицинских исследований.

Учитывая ограниченность данных о составе микробиоты кишечника при ЯК и СРК у пациентов из регионов Сибири, а также отсутствие сведений о кишечной микробиоте больных бронхиальной астмой (БА), цель исследования – оценка 16S рРНК-профилей пациентов с СРК, ЯК и БА в сравнении с таковыми здоровых доноров.

Материалы и методы

В работе использовали образцы фекалий, полученные от восьми пациентов с БА, восьми больных с СРК и 18 пациентов с ЯК. Среди исследуемых женщины составили 47.4 %, мужчины – 52.6 %. Диагнозы «синдром раздраженного кишечника» и «язвенный колит» подтверждали на основании результатов исследования уровня фекального кальпротектина, данных фиброколоноскопии и гистологического исследования биоптатов, взятых из разных отделов толстой и подвздошной кишок. Образцы пациен-

тов, в которых рутинными методами были обнаружены *Clostridium difficile,* не исследовали. Диагноз «атопическая бронхиальная астма» подтверждали на основании общего, биохимического и иммунологического анализов крови и данных сенсибилизации. Также в исследовании использованы восемь образцов здоровых добровольцев без хронических заболеваний и не болевших последние три месяца. Все пациенты и добровольцы отрицали систематическое употребление алкоголя, только три больных мужского пола с СРК употребляли табачные изделия. Все участники исследования в течение как минимум двух недель до забора образцов не принимали антибактериальные препараты. Все пациенты и добровольцы предоставили информированное согласие с проводимым исследованием и анонимной обработкой данных. Исследование одобрено локальным этическим комитетом Автономной некоммерческой организации «Центр новых медицинских технологий в Академгородке».

Выделение суммарной ДНК проводили, как описано ранее (Тикунов и др., 2020). Основные этапы включали осветление 50 мг из каждого образца с последующим применением набора для выделения ДНК из клеток тканей и крови (ООО «БиоЛабМикс», Россия) с добавлением лизоцима для повышения эффективности получения ДНК из грамположительных бактерий. Амплификацию фрагмента гена 16S рРНК, содержащего вариабельные участки V3 и V4, проводили методом полимеразной цепной реакции с использованием в качестве матрицы полученной ДНК фьюжн-праймеров (NEB-FF 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTACGGGAGGCAGCAG-3′, NEB-FR 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGACTACCGGGGTATCT-3′) и высокоточной полимеразы Q5 (New England Biolabs, США). Продукты амплификации очищали электрофоретически в геле из легкоплавкой SeaKem GTG-агарозы (Lonza, США).

Конструирование библиотек выполняли, как описано ранее (Тикунов и др., 2020). Основные этапы включали обогащение полученных ампликонов, введение баркодов и служебных последовательностей с использованием полимеразы Q5 и набора олигонуклеотидов Dual index set (New England Biolabs, США) с последующей очисткой полученных библиотек на магнитных частицах AMPure XP (Beckman Coulter, США). Концентрацию ДНК в библиотеках измеряли с помощью набора Qubit dsDNA HS (Life Technologies, США). По результатам измерений библиотеки объединяли в пул таким образом, чтобы соотношение ДНК-библиотек в пуле было эквимолярным Секвенирование вели на платформе MiSeq с использованием набора реагентов MiSeq reagent kit v2 2 × 250-cycles (Illumina, США).

Методы анализа данных секвенирования описаны ранее (Тикунов и др., 2020). Предварительно из последовательностей ридов удаляли последовательности адаптеров и проводили фильтрацию ридов по качеству. Полученные риды анализировали с помощью генерации операционных таксономических единиц (OTU) с последующим картированием последовательностей на полученные OTU в пакете программ Usearch-9.2 и использованием классификации ридов алгоритмом Kraken по базе данных известных последовательностей 16S рРНК Silva v.132 (full). В первом случае OTU генерировали алгоритмом unoise2 с отбраковкой химерных последовательностей и учетом ошибок чтения. Таблицы полученных частот встречаемости OTU обработаны в среде R3.3.3. Во втором случае полученные риды картировали на базу данных 16S рРНК Silva с помощью алгоритма seed-kraken с использованием разреженного *k*-мера со специальной решеткой, позволяющей увеличить специфичность классификации. Индекс Шеннона рассчитывали в пакете программ R; достоверность различий между индексами Шеннона определяли с помощью t-теста Хатчесона. Визуализацию результатов анализа библиотек последовательностей методом главных координат PCoA проводили на основе матриц дистанций с применением пакета программ vegan. Для установления достоверности различий между величинами использовали критерий Стьюдента.

Результаты и обсуждение

Оценка биоразнообразия бактериальных сообществ

На основе ДНК, выделенной из образцов фекалий пациентов и доноров, сконструированы 42 библиотеки фрагментов гена 16S рРНК. Фрагменты содержали вариабельные участки V3 и V4 гена 16S рРНК – на основе этой последовательности возможна таксономическая классификация большинства бактерий (Chakravorty et al., 2007; Wang, Qian, 2009). Образцы сгруппированы в четыре выборки: БА – 8 образцов больных атопической БА (средний возраст 38.1 года; от 25 до 56 лет), СРК – 8 образцов пациентов с СКР (средний возраст 44.9 года; от 25 до 64 лет), ЯК – 18 образцов пациентов с ЯК (средний возраст 39.6 года; от 25 до 65 лет), здоровые добровольцы (ЗД) – 8 образцов здоровых добровольцев (средний возраст 27.1 года; от 20 до 39 лет). Средний возраст добровольцев в группе ЗД был достоверно ниже в сравнении с группами пациентов с БА, СРК и ЯК ($p \leq 0.01$ во всех случаях); при этом данный показатель в трех выборках больных статистически значимо не различался.

Результаты секвенирования и классификации полученных ридов представлены в табл. 1. Все библиотеки содержали более 100 тыс. ридов, в среднем более 99.6 % ридов таксономически отнесены к определенному типу бактерий. В двух образцах больных БА и одном образце пациента с СРК обнаружены последовательности архей (тип Euryarchaeota).

Альфа-разнообразие бактериальных сообществ в исследуемых группах оценивали с помощью индекса Шеннона – комплексного показателя, учитывающего количество видов и их выровненность. Самым низким индекс Шеннона был для выборки ЯК, самым высоким – для ЗД (рис. 1), хотя различия не имели статистической значимости. Оценка выборок методом главных компонентов показала, что наиболее компактно расположены точки, характеризующие библиотеки последовательностей из группы ЗД (рис. 2). Расположение точек, характеризующих библиотеки из выборок пациентов, при меньшем значении индекса Шеннона по сравнению с таковым для ЗД свидетельствует об относительной нестабильности микробиомов у таких больных.

Table 1. The taxonomy of reads in 16S rRNA libraries

Parameter		IBS, *n* = 8	UC, *n* = 18	Asthma, *n* = 8	HV, *n* = 8
Number of reads, N ± δ		149863 ± 13477	138479 ± 16887	142314 ± 13724	129435 ± 12429
Reads with identified bacterium phylum, % ± δ		99.57 ± 0.18	99.66 ± 0.12	99.62 ± 0.11	99.64 ± 0.13
Shannon index, I ± δ		3.27 ± 0.85	3.05 ± 0.63	3.28 ± 0.72	3.48 ± 0.39
The overall number of bacterium phyla		11, 7 to 8 per library	11, 5 to 9 per library	11, 6 to 9 per library	11, 5 to 9 per library
Percentage of detected bacterium phyla	Firmicutes	53.2	56.1	65.1	70.4
	Bacteroidetes	38.5	31.3	18.5	26.5
	Proteobacteria	5.7	10.0	13.4	1.4
	Actinobacteria	1.0	1.9	0.7	0.7
	Verrucomicrobia	1.0	0.4	1.9	0.5
	Fusobacteria	< 0.1	< 0.1	< 0.1	< 0.1
	Tenericutes	< 0.1	< 0.1	< 0.1	< 0.1
	Synergistetes	< 0.1	< 0.1	n.d.	< 0.1
	Epsilonbacteraeota	< 0.1	< 0.1	< 0.1	< 0.1
	Cyanobacteria	< 0.1	< 0.1	< 0.1	< 0.1
	Patescibacteria	< 0.1	< 0.1	< 0.1	n.d.
	Fibrobacteres	n.d.	n.d.	< 0.1	< 0.1
The overall number of bacterium classes		17, 13 to 14 per library	15, 11 to 15 per library	16, 12 to 15 per library	17, 10 to 15 per library
Percentage of detected bacterium classes	Bacilli	0.8	1.5	0.6	0.5
	Clostridia	41.8	46.7	55.8	66.0
	Erysipelotrichia	0.2	0.2	0.2	< 0.1
	Negativicutes	10.2	7.5	8.3	3.7
	Bacteroidia	38.5	31.3	18.5	26.5
	Alphaproteobacteria	1.9	5.4	2.0	0.4
	Deltaproteobacteria	0.2	< 0.1	< 0.1	< 0.1
	Gammaproteobacteria	3.6	4.4	11.2	0.9
	Actinobacteria	0.6	0.6	0.5	0.4
	Coriobacteriia	0.4	1.3	0.3	0.3
	Verrucomicrobia	1.0	0.4	1.9	0.5
	Fusobacteriia	< 0.1	< 0.1	< 0.1	< 0.1
	Mollicutes	< 0.1	< 0.1	< 0.1	< 0.1
	Synergistia	< 0.1	< 0.1	n.d.	< 0.1
	Campylobacteria	< 0.1	< 0.1	< 0.1	< 0.1
	Oxyphotobacteria	< 0.1	< 0.1	< 0.1	< 0.1
	Saccharimonadia	< 0.1	< 0.1	< 0.1	n.d.
	Fibrobacteria	n.d.	n.d.	< 0.1	< 0.1

Note. Library designations here and below: IBS, irritated bowel syndrome; UC, ulcerative colitis, HV, healthy volunteers. n.d., not detected.

Всего в библиотеках выявлены последовательности 12 типов бактерий (см. табл. 1); последовательности девяти типов присутствовали во всех четырех выборках. Из них последовательности пяти типов (Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria и Fusobacteria) определены во всех библиотеках, а последовательности, относящиеся к типу Verrucomicrobia, обнаружены во всех библиотеках из групп БА и ЗД, при этом в СРК и ЯК они отсутствовали в одной и четырех библиотеках соответственно. Последовательности представителей типов Tenericutes, Epsilonbacteraeota, Cyanobacteria, Synergistetes, Patescibacteria и Fibrobacteres найдены не во всех библиотеках, а последних трех типов – не во всех выборках. В среднем во всех группах доминировали последовательности Firmicutes (в среднем по выборке > 53 %) и Bacteroidetes (в среднем > 18 %), причем в группе СРК

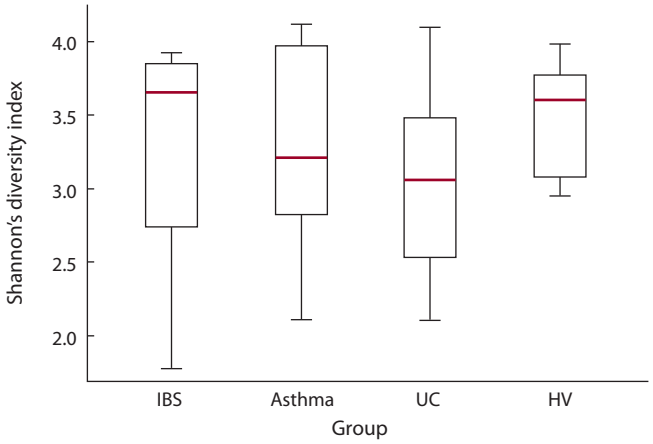


Fig. 1. Shannon index, reflecting the alpha-diversity of the gut microbiome in samples obtained from patients with irritable bowel syndrome (IBS), bronchial asthma, ulcerative colitis (UC), and healthy volunteers (HV). Median values and quartiles are indicated.

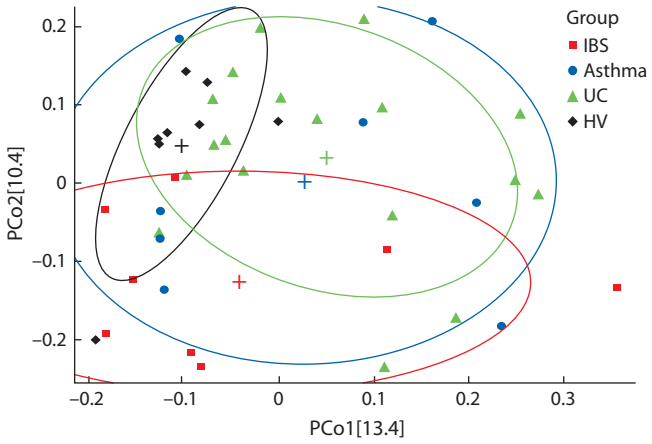


Fig. 2. Visualization of sequence library analysis by the method of principal coordinates PCoA based on matrices of distances. Red squares indicate data for libraries from IBS samples; blue dots, from asthma samples; green triangles, from UC samples; and black diamonds – from HV samples.

Plus signs of the corresponding colors mark the centers of areas occupied by the respective groups of samples. The values of the first and second principal coordinates are presented on the *X* and *Y* axes, respectively.

последовательности Bacteroidetes выявлены статистически значимо чаще, чем в БА (в среднем 38.5 против 18.5 %; $p \leq 0.05$). Третьими по встречаемости во всех выборках были последовательности Proteobacteria. Из табл. 1 видно, что меньше всего таких последовательностей зафиксировано в группе ЗД (1.36 %), в остальных выборках таких последовательностей отмечено существенно больше (СРК – 5.71 %, БА – 13.35 %, ЯК – 10.00 %); статистическая достоверность различий показана даже для групп ЗД и СРК. Последовательности типа Actinobacteria были в среднем четвертыми по встречаемости, однако в выборках ЯК и СРК (заболевания кишечника) их встречаемость в среднем превышала таковую для последовательностей Proteobacteria в группе ЗД.

Девять из двенадцати обнаруженных бактериальных типов представлены последовательностями единственного класса, включая один доминирующий тип Bacteroidetes (класс Bacteroidia). Еще один доминирующий тип, Firmicutes, был представлен последовательностями четырех классов – Bacilli, Clostridia, Erysipelotrichia, Negativicutes, причем во всех выборках доминировали последовательности класса Clostridia. Тип Proteobacteria содержал последовательности трех классов – Alphaproteobacteria, Deltaproteobacteria и Gammaproteobacteria, у пациентов из этого типа преобладали последовательности последнего класса. Тип Actinobacteria включал последовательности двух классов – Actinobacteria и Coriobacteriia. Всего выявлены последовательности 18 классов (см. табл. 1). Встречаемость последовательностей основных порядков показана на рис. 3.

Сравнительный анализ таксономического состава микробных сообществ

Известно, что основными компонентами микробного сообщества кишечника здоровых людей выступают представители типов Firmicutes (~70 %), Bacteroidetes (~30 %), Proteobacteria (<5 %), Actinobacteria (<2 %), Verrucomicrobia (<1 %) и Fusobacteria (<1 %); представители еще 10–12 типов бактерий являются транзиентными или встречаются в зависимости от региона проживания человека, его возраста и типа питания (Belizário et al., 2018). Анализ таксономического состава библиотек из выборки ЗД показал соответствие средних значений приведенным показателям (см. табл. 1). Это позволило оценить сдвиг таксономического состава в кишечном микробиоме пациентов с СРК и ЯК (заболевания кишечника) и БА относительно кишечного микробиома здоровых доноров.

Изменение соотношения количества представителей Firmicutes и Bacteroidetes в микробиоте кишечника может быть одним из индикаторов нарушения состояния микробиоты, при этом сдвиг этого соотношения как в сторону увеличения, так и уменьшения описан при ряде заболеваний (Tana et al., 2010; Rajilić-Stojanović et al., 2011; Jeffery et al., 2012; Jalanka-Tuovinen et al., 2014; Pozuelo et al., 2015; Tap et al., 2017). Соотношения Firmicutes/Bacteroidetes в группе ЗД составило 2.6, в СРК и ЯК это соотношение существенно уменьшилось (1.4 и 1.8 соответственно), а в выборке БА – увеличилось (3.5) (табл. 2). Полученные данные согласуются с описанным ранее уменьшением встречаемости Firmicutes при ЯК (Machiels et al., 2014) и СРК (Jalanka-Tuovinen et al., 2014; Bennet et al., 2015; Pozuelo et al., 2015; Su et al., 2018), однако относительно СРК результаты разнятся и в ряде исследований зарегистрировано увеличение соотношения Firmicutes/Bacteroidetes (Tana et al., 2010; Rajilić-Stojanović et al., 2011; Jeffery et al., 2012; Tap et al., 2017).

Обнаруженное в данном исследовании уменьшение доли последовательностей, принадлежащих типу Firmicutes, в выборках больных относительно группы ЗД наиболее выраженно среди пациентов с заболеваниями кишечника. Наибольший вклад в эти различия внесли последовательности класса Clostridia, уменьшение доли которых в выборках СРК и ЯК было статистически значимым ($p < 0.01$ и $p < 0.05$ соответственно). Одновременно

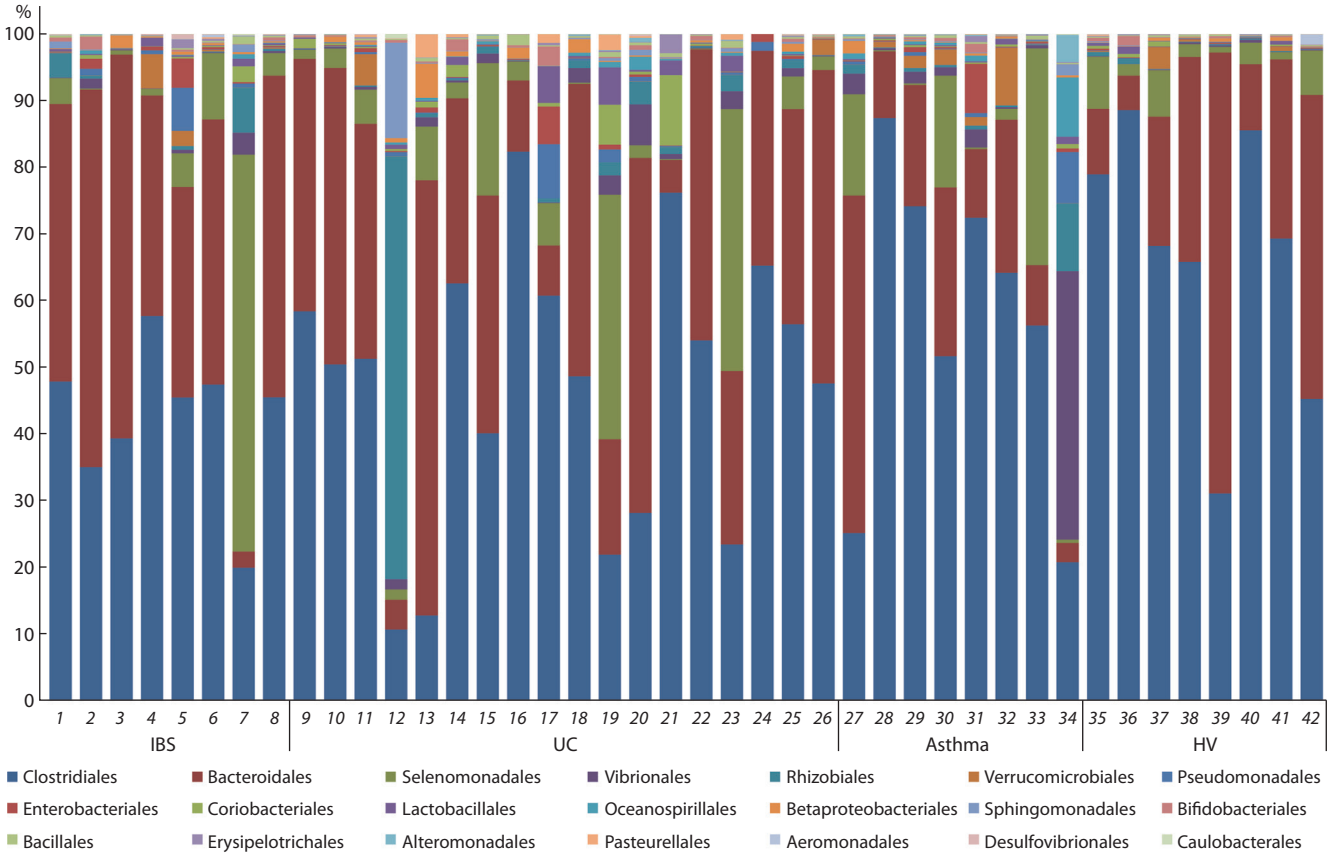


Fig. 3. Taxonomic classification of operational taxonomic units at the order level based on the Silva v.132 (full) database.

Samples *1–8* were obtained from patients with IBS; *9–26* from patients with UC; *27–34* from patients with bronchial asthma; and *35–42* from healthy volunteers (HV). Orders with abundance exceeding 0.1 % are shown.

доля последовательностей класса Negativicutes во всех группах больных, наоборот, увеличилась (см. табл. 1). Ранее отмечено, что при СРК уменьшается доля последовательностей семейств Ruminococcaceae и Erysipelotrichaceae, представители которых участвуют в продукции короткоцепочечных жирных кислот (Załęski et al., 2013; Pozuelo et al., 2015).

В данном исследовании также выявлено снижение доли последовательностей Ruminococcaceae как при СРК, так и при ЯК, однако доля последовательностей Erysipelotrichaceae в выборках пациентов с заболеваниями кишечника и БА существенно не изменилась по сравнению с ЗД (см. табл. 2). Среди последовательностей Ruminococcaceae особенно заметно снижение доли последовательностей *Faecalibacterium* spp. в группе СРК относительно ЗД ($p<0.01$), что соответствует полученным ранее результатам (Rajilić-Stojanović et al., 2011; Rodiño-Janeiro et al., 2018). Кроме этого, следует отметить заметное увеличение доли последовательностей сем. Veillonellaceae среди больных, однако описанное в литературе увеличение доли последовательностей *Veillonella* spp. при СРК (Malinen et al., 2005; Tana et al., 2010; Rigsbee et al., 2012) в нашей работе не определено, хотя при ЯК наблюдалось. Не обнаружено существенных отличий встречаемости во всех исследуемых выборках последовательностей *Staphylococcus* spp., *Enterococcus* spp. и *Streptococcus* spp., среди которых присутствуют патогенные и условно-патогенные виды. Зафиксировано существенное уменьшение доли последовательностей *Roseburia* spp. и *Lachnospira* spp. в группе БА (см. табл. 2). Ранее показано, что *Roseburia* spp. наряду с представителями родов *Bifidobacterium* и *Lactobacillus* является основным продуцентом полиненасыщенных жирных кислот, уменьшение концентрации которых в кишечнике детей раннего возраста ассоциировано с риском развития БА (Chiu et al., 2019; Lee-Sarwar et al., 2020). Снижение числа представителей рода *Lachnospira* связано с увеличением относительного обилия представителей рода *Clostridium*, что, в свою очередь, также ассоциировано с повышением риска развития астмы у детей (Arrieta et al., 2015; Stiemsma et al., 2016; Hufnagl et al., 2020).

В литературе описано уменьшение доли последовательностей, принадлежащих типу Bacteroidetes у пациентов с ЯК (Machiels et al., 2014). В нашем исследовании средняя доля последовательностей этого типа в группах ЯК и СРК, наоборот, несколько увеличилась по сравнению с ЗД (см. табл. 1), причем наиболее значимое увеличение зарегистрировано для последовательностей *Bacteroides* spp. (см. табл. 2). В выборке БА доля последовательностей Bacteroidetes незначительно уменьшилась по сравнению с ЗД, однако встречаемость последовательностей *Bacteroides* spp. в этой группе значимо не изменилась. Следует отметить, что в выборках СРК, ЯК и БА заметно увеличилось соотношение представителей

Table 2. Mean percentages of sequences from different bacterial taxa in 16S rRNA libraries

Taxon	IBS	UC	Asthma	HV
Ratio Firmicutes/Bacteroidetes	1.4	1.8	3.5	2.6
Firmicutes				
fam. Lactobacillaceae	< 0.1	0.1	0.1	< 0.1
Lactobacillus spp.	< 0.1	0.1	< 0.1	< 0.1
fam. Ruminococcaceae	31.5	31.8	43.3	51.4
Faecalibacterium spp.	5.4	18.4	17.0	23.6
fam. Erysipelotrichaceae	0.2	0.3	0.2	< 0.1
fam. Lachnospiraceae	6.3	11.5	7.3	9.9
Roseburia spp.	1.6	2.15	0.5	1.2
Lachnospira spp.	0.1	0.3	< 0.1	0.1
fam. Veillonellaceae	9.2	6.8	8.2	1.4
Veillonella spp.	< 0.1	2.5	< 0.1	0.1
Bacteroidetes				
fam. Prevotellaceae	8.8	4.8	3.2	9.3
fam. Bacteroidaceae	20.7	18.4	10.6	8.9
Bacteroides spp.	20.7	18.4	10.6	8.9
Proteobacteria				
fam. Beijerinckiaceae	0.7	2.0	1.0	0.2
Methylobacterium spp.	0.7	1.9	1.0	0.1
fam. Sphingomonadaceae	0.3	1.0	0.3	< 0.1
Sphingomonas spp.	0.3	0.8	0.2	< 0.1
fam. Burkholderiaceae	0.4	0.8	0.4	0.2
Parasutterella spp.	0.3	0.3	0.3	0.1
fam. Enterobacteriaceae	0.9	0.6	1.2	0.1
Enterobacter spp.	< 0.1	< 0.1	0.2	< 0.1
Escherichia-Shigella	0.6	0.4	0.2	< 0.1
fam. Halomonadaceae	0.2	0.3	1.4	< 0.1
Halomonas spp.	0.2	0.3	1.3	< 0.1
fam. Vibrionaceae	0.7	1.2	6.2	0.2
Vibrio spp.	0.7	1.2	6.2	0.2
Actinobacteria				
fam. Bifidobacteriaceae	0.4	0.5	0.3	0.3
Bifidobacterium spp.	0.4	0.5	0.3	0.3

семейств Bacteroidaceae/Prevotellaceae по сравнению с ЗД (2.4, 3.8, 3.3 против 0.9 соответственно). Обычно доля последовательностей Bacteroidaceae больше у людей, питающихся преимущественно белковой пищей, а доля Prevotellaceae – у людей, отдающих предпочтение растительной пище: жителей средиземноморского региона, Юго-Восточной Азии и у вегетарианцев (Ley, 2016). Доминирующие виды из обоих семейств, обнаруживаемые в кишечной микробиоте человека, способны расщеплять различные растительные полисахариды (Flint et al., 2012), и для объяснения различий в исследуемых выборках необходима видоспецифическая характеризация образцов.

Значительные отличия обнаружены во встречаемости в исследуемых группах последовательностей типа Proteobacteria, ряд представителей которого относятся к патогенным и условно-патогенным бактериям. В выборках СРК, ЯК и БА по сравнению с ЗД вклад таких последовательностей был существенно выше (см. табл. 1), что коррелировало с описанными ранее данными для пациентов с СРК и ЯК (Saebo et al., 2005; Gradel et al., 2009; Sonnenberg, Genta, 2012; Bennet et al., 2015; Shin et al., 2015; Shen et al., 2018). При этом в группах пациентов заметно увеличилась доля последовательностей классов Alphaproteobacteria и Gammaproteobacteria: в частности, обнаружен рост встречаемости последовательностей родов *Methylobacterium* (сем. Beijerinckiaceae), *Sphingomonas* (сем. Sphingomonadaceae), *Parasutterella* (сем. Burkholderiaceae), *Halomonas* и *Vibrio* (сем. Halomonadaceae), а также последовательностей *Escherichia* spp. и *Shigella* spp. В выборке БА обнаружено значимо больше последовательностей рода *Enterobacter* (сем. Enterobacteriaceae). Важно отметить, что некоторые виды вышеперечисленных родов могут вызывать инфекции, особенно у людей со сниженным иммунитетом (Lai et al., 2011; Kovaleva et al., 2014; Baker-Austin et al., 2018; Chen et al., 2018). Более того, доказана ассоциация *Parasutterella*, в частности *P. excrementihominis*, с СРК и воспалительными заболеваниями кишечника (Shin et al., 2015; Chen et al., 2018). У пациентов по сравнению с ЗД не зафиксировано заметного увеличения последовательностей родов протеобактерий, членами которых являются хорошо известные инфекционные агенты: *Klebsiella*, *Proteus*, *Salmonella*, *Serratia* (все сем. Enterobacteriaceae), *Acinetobacter* (сем. Moraxellaceae), *Pseudomonas* (сем. Pseudomonadaceae) и *Stenotrophomonas* (сем. Xanthomonadaceae). Несмотря на отсутствие заметного повышения доли последовательностей вышеперечисленных родов в группах СРК, ЯК и БА относительно ЗД факт увеличения встречаемости последовательностей Proteobacteria, в частности, Enterobacteriaceae, в исследуемых выборках библиотек от пациентов служит признаком дисбиоза кишечной микрофлоры (Shin et al., 2015).

Доля последовательностей типа Actinobacteria в группах ЯК и СРК в среднем превышала таковую для ЗД (см. табл. 1), однако вклад последовательностей *Bifidobacterium* spp. во всех четырех исследуемых выборках практически не различался (см. табл. 2), что не соответствует результатам, описанным ранее для пациентов с СРК (Malinen et al., 2005; Rajilić-Stojanović et al., 2011; Parkes et al., 2012; Zhuang et al., 2017). *Bifidobacterium* spp. наряду с *Lactobacillus* spp. (сем. Bacilli) способны модулировать состав кишечной микробиоты и влиять на иммунную систему человека путем взаимодействия с рецепторами CD209, экспонированными на поверхности дендритных клеток (Pace et al., 2015). Кроме того, отдельные представители этих родов секретируют бактериоцины – соединения, обладающие бактерицидным эффектом в отношении ряда патогенных бактерий (Angelakis et al., 2013; Rodiño-Janeiro et al., 2018). Отличий во встречаемости последовательностей лактобацилл в исследуемых группах также не обнаружено.

Заключение

В данном пилотном исследовании проведена сравнительная оценка биоразнообразия и таксономической структуры 16S-профилей кишечного микробиома пациентов с СРК, ЯК и БА и здоровых добровольцев. Подтверждены различия между кишечными сообществами бактерий у пациентов с воспалительными заболеваниями кишечника и таковыми у здоровых людей. В ряде случаев полученные нами данные согласуются с результатами аналогичных зарубежных исследований: уменьшение индекса Шеннона, соотношения и обилия представителей Firmicutes/Bacteroidetes и увеличение доли последовательностей Proteobacteria в микробиоме пациентов с воспалительными заболеваниями кишечника в отличие от здоровых. Вместе с тем по некоторым показателям наши результаты отличаются от опубликованных ранее. Так, у пациентов с СРК и ЯК не уменьшилась доля последовательностей Erysipelotrichaceae, а даже несколько увеличилась (представители этого семейства продуцируют короткоцепочечные жирные кислоты); не отмечено и уменьшение встречаемости лакто- и бифидобактерий (отвечают за продукцию полиненасыщенных жирных кислот). Также нетипично увеличение доли последовательностей Bacteroidetes, в частности *Bacteroides* spp., у больных ЯК и СРК вместо зарегистрированного ранее уменьшения (Machiels et al., 2014; Dubinsky, Braun, 2015); не обнаружен рост доли последовательностей *Veillonella* spp. при СРК. Подобные отличия могут быть объяснены как многократными попытками лечения пациентов, кишечные сообщества которых изучали, так и региональными характеристиками популяции и особенностями питания.

Следует отметить, что ранее мы описывали похожую выборку больных ЯК из Западной Сибири (Тикунов и др., 2020) и в целом данные по долям последовательностей из семейств, принадлежащих типам Firmicutes, Bacteroidetes и Actinobacteria, в обоих исследованиях существенно не различались; лишь доля последовательностей Proteobacteria в предыдущей работе была несколько меньше. Также в настоящем исследовании, в отличие от предыдущего, у пациентов с ЯК не обнаружены последовательности архей. Отсутствие существенных различий между 16S-профилями кишечных микробиомов двух групп больных ЯК свидетельствует об определенных региональных особенностях микробиоты таких пациентов. Вместе с тем в данное пилотное исследование включены лишь по 8, 18 и 8 образцов пациентов с СРК и ЯК и здоровых участников соответственно. Для вывода о том, являются ли обнаруженные тенденции закономерностями, необходимо продолжение исследований с использованием большего количества анализируемых образцов.

При этом ряд результатов исследования бактериальных сообществ пациентов с БА получены впервые. Выполненные до настоящего времени исследования включали изучение микробиоты лишь младенцев и детей постарше с БА, причем большая часть работ посвящена микробиоте дыхательных путей и ротовой полости. Исследования же содержимого кишечника детей с БА в основном сводились к поиску ассоциаций между содержанием определенных соединений и развитием астмы. В нашей работе оценены бактериальные сообщества кишечника взрослых пациентов с БА. В связи с этим нельзя сделать заключение о том, насколько существенно обнаруженное увеличение соотношения Firmicutes/Bacteroidetes в выборке БА относительно ЗД. Ряд выявленных особенностей кишечной микробиоты взрослых с БА согласуются с описанными ранее показателями таковой у больных детей. Так, в кишечной микробиоте взрослых с БА, как и детей, определено уменьшение доли последовательностей *Roseburia* spp., *Lachnospira* spp., *Veillonella* spp., однако доля последовательностей *Faecalibacterium* spp. и *Lactobacillus* spp. у взрослых пациентов, в отличие от детей, была такой же, как и у здоровых. Возможно, недостаточную представленность некоторых, но не всех бактерий в кишечной микробиоте детей с БА с годами компенсирует диета или прием пробиотиков. Также наблюдаемые различия могут быть связаны не с возрастом пациентов, а региональными особенностями питания. Выявленное в данном исследовании существенное превышение доли последовательностей Enterobacteriaceae в выборке БА по сравнению с ЗД закономерно, так как свидетельствует о дисбиозе кишечной микробиоты. Однако отсутствует достоверное объяснение обнаруженного нами значимого превышения доли последовательностей *Halomonas* и *Vibrio* в кишечной микробиоте пациентов с БА по сравнению не только со здоровыми добровольцами, но и больными СРК и ЯК. Важно, что указанные различия не связаны с особенностями питания пациентов с БА, поскольку доля последовательностей *Halomonas* spp. и *Vibrio* spp. у больных атопическим дерматитом, находившихся одновременно с пациентами с БА в том же отделении медицинского учреждения, не отличалась от таковой в группах ЗД, СРК и ЯК. Вышеперечисленные факты свидетельствуют о том, что дисбаланс кишечной микрофлоры, наблюдаемый при БА, ассоциирован с развитием этого заболевания, однако для понимания роли отдельных компонентов микробиоты кишечника в его патогенезе необходимы дальнейшие исследования с включением более объемных выборок.

Список литературы / References

Тикунов А.Ю., Морозов В.В., Швалов А.Н., Бардашева А.В., Шрайнер Е.В., Максимова О.А., Волошина И.О., Морозова В.В., Власов В.В., Тикунова Н.В. Изменение кишечного микробиома пациентов с язвенным колитом после трансплантации кишечной миробиоты. *Вавиловский журнал генетики и селекции.* 2020; 24(2):168-175. DOI 10.18699/VJ20.610.

[Tikunov A.Y., Morozov V.V., Shvalov A.N., Bardasheva A.V., Shrayner E.V., Maksimova O.A., Voloshina I.O., Morozova V.V., Vlasov V.V., Tikunova N.V. Fecal microbiome change in patients with ulcerative colitis after fecal microbiota transplantation. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2020;24(2):168-175. DOI 10.18699/VJ20.610. (in Russian)]

Angelakis E., Merhej V., Raoult D. Related actions of probiotics and antibiotics on gut microbiota and weight modification. *Lancet Infect. Dis.* 2013;13(10):889-899. DOI 10.1016/S1473-3099(13)70179-8.

Arrieta M.C., Stiemsma L.T., Dimitriu P.A., Thorson L., Russell S., Yurist-Doutsch S., Kuzeljevic B., Gold M.J., Britton H.M., Lefebvre D.L., Subbarao P., Mandhane P., Becker A., McNagny K.M., Sears M.R., Kollmann T. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci. Transl. Med.* 2015; 7(307):307ra152. DOI 10.1126/scitranslmed.aab2271.

Bajer L., Kverka M., Kostovcik M., Macinga P., Dvorak J., Stehlikova Z., Brezina J., Wohl P., Spicak J., Drastich P. Distinct gut microbiota profiles in patients with primary sclerosing cholangitis and

ulcerative colitis. *World J. Gastroenterol.* 2017;23(25):4548-4558. DOI 10.3748/wjg.v23.i25.4548.

Baker-Austin C., Oliver J.D., Alam M., Ali A., Waldor M.K., Qadri F., Martinez-Urtaza J. Vibrio spp. infections. *Nat. Rev. Dis. Primers.* 2018;4(1):8. DOI 10.1038/s41572-018-0005-8.

Bennet S.M., Ohman L., Simren M. Gut microbiota as potential orchestrators of irritable bowel syndrome. *Gut Liver.* 2015;9(3):318-331. DOI 10.5009/gnl14344.

Belizário J.E., Faintuch J., Garay-Malpartida M. Gut microbiome dysbiosis and immunometabolism: new frontiers for treatment of metabolic diseases. *Mediators Inflamm.* 2018;2018:1-12. DOI 10.1155/2018/2037838.

Chakravorty S., Helb D., Burday M., Connell N., Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods.* 2007;69(2):330-339.

Chen Y.J., Wu H., Wu S.D., Lu N., Wang Y.T., Liu H.N., Dong L., Liu T.T., Shen X.Z. Parasutterella, in association with irritable bowel syndrome and intestinal chronic inflammation. *J. Gastroenterol. Hepatol.* 2018;33(11):1844-1852. DOI 10.1111/jgh.14281.

Cheng Y.W., Fischer M. The present status of fecal microbiota transplantation and its value in the elderly. *Curr. Treat. Options Gastroenterol.* 2017;15(3):349-362. DOI 10.1007/s11938-017-0143-1.

Chiu C.Y., Cheng M.L., Chiang M.H., Kuo Y.L., Tsai M.H., Chiu C.C., Lin G. Gut microbial-derived butyrate is inversely associated with IgE responses to allergens in childhood asthma. *Pediatr. Allergy Immunol.* 2019;30(7):689-697. DOI 10.1111/pai.13096.

Donaldson G.P., Lee S.M., Mazmanian S.K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* 2016;14(1):20-32. DOI 10.1038/nrmicro3552.

Dubinsky M., Braun J. Diagnostic and prognostic microbial biomarkers in inflammatory bowel diseases. *Gastroenterology.* 2015;149:1265-1274e3. DOI 10.1053/j.gastro.2015.08.006.

Flint H.J., Scott K.P., Duncan S.H., Louis P., Forano E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes.* 2012;3(4):289-306. DOI 10.4161/gmic.19897.

Fujimura K.E., Slusher N.A., Cabana M.D., Lynch S.V. Role of the gut microbiota in defining human health. *Expert Rev. Anti. Infect. Ther.* 2010;8(4):435-454. DOI 10.1586/eri.10.14.

Gradel K.O., Nielsen H.L., Schønheyder H.C., Ejlertsen T., Kristensen B., Nielsen H. Increased short- and long-term risk of inflammatory bowel disease after salmonella or campylobacter gastroenteritis. *Gastroenterology.* 2009;137(2):495-501. DOI 10.1053/j.gastro.2009.04.001.

Hufnagl K., Pali-Schöll I., Roth-Walter F., Jensen-Jarolim E. Dysbiosis of the gut and lung microbiome has a role in asthma. *Semin Immunopathol.* 2020;42(1):75-93. DOI 10.1007/s00281-019-00775-y.

Jalanka-Tuovinen J., Salojärvi J., Salonen A., Immonen O., Garsed K., Kelly F.M., Zaitoun A., Palva A., Spiller R.C., de Vos W.M. Faecal microbiota composition and host-microbe cross-talk following gastroenteritis and in postinfectious irritable bowel syndrome. *Gut.* 2014;63(11):1737-1745. DOI 10.1136/gutjnl-2013-305994.

Jeffery I.B., O'Toole P.W., Öhman L., Claesson M.J., Deane J., Quigley E.M., Simrén M. An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut.* 2012;61(7):997-1006. DOI 10.1136/gutjnl-2011-301501.

Kovaleva J., Degener J.E., van der Mei H.C. Methylobacterium and its role in health care-associated infection. *J. Clin. Microbiol.* 2014;52(5):1317-1321. DOI 10.1128/JCM.03561-13.

Lai C.C., Cheng A., Liu W.L., Tan C.K., Huang Y.T., Chung K.P., Lee M.R., Hsueh P.R. Infections caused by unusual Methylobacterium species. *J. Clin. Microbiol.* 2011;49(9):3329-3331. DOI 10.1128/JCM.01241-11.

Lee-Sarwar K.A., Lasky-Su J., Kelly R.S., Litonjua A.A., Weiss S.T. Gut microbial-derived metabolomics of asthma. *Metabolites.* 2020;10(3):97. DOI 10.3390/metabo10030097.

Ley R.E. Gut microbiota in 2015: Prevotella in the gut: choose carefully. *Nat. Rev. Gastroenterol. Hepatol.* 2016;13(2):69-70. DOI 10.1038/nrgastro.2016.4.

Machiels K., Joossens M., Sabino J., De Preter V., Arijs I., Eeckhaut V., Ballet V., Claes K., Van Immerseel F., Verbeke K., Ferrante M., Verhaegen J., Rutgeerts P., Vermeire S. A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut.* 2014;63(8):1275-1283. DOI 10.1136/gutjnl-2013-304833.

Malinen E., Rinttilä T., Kajander K., Mättö J., Kassinen A., Krogius L., Saarela M., Korpela R., Palva A. Analysis of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR. *Am. J. Gastroenterol.* 2005;100(2):373-382.

Manichanh C., Borruel N., Casellas F., Guarner F. The gut microbiota in *IBD. Nat. Rev. Gastroenterol. Hepatol.* 2012;9(10):599-608. DOI 10.1038/nrgastro.2012.152.

O'Hara A.M., Shanahan F. The gut flora as a forgotten organ. *EMBO Rep*. 2006;7(7):688-693. DOI 10.1038/sj.embor.7400731.

Pace F., Pace M., Quartarone G. Probiotics in digestive diseases: focus on Lactobacillus GG. *Minerva Gastroenterol. Dietol.* 2015;61(4):273-292.

Parkes G.C., Rayment N.B., Hudspith B.N., Petrovska L., Lomer M.C., Brostoff J., Whelan K., Sanderson J.D. Distinct microbial populations exist in the mucosa-associated microbiota of sub-groups of irritable bowel syndrome. *Neurogastroenterol. Motil.* 2012;24(1):31-39. DOI 10.1111/j.1365-2982.2011.01803.x.

Pozuelo M., Panda S., Santiago A., Mendez S., Accarino A., Santos J., Guarner F., Azpiroz F., Manichanh C. Reduction of butyrate- and methane-producing microorganisms in patients with Irritable Bowel Syndrome. *Sci. Rep.* 2015;5:12693. DOI 10.1038/srep12693.

Qin J., Li R., Raes J., Arumugam M., Burgdorf K.S., Manichanh C., Nielsen T., Pons N., Levenez F., Yamada T., Mende D.R., Li J., Xu J., Li S., Li D., Cao J., Wang B., Liang H., Zheng H., Xie Y., Tap J., Lepage P., Bertalan M., Batto J.M., Hansen T., Le Paslier D., Linneberg A., Nielsen H.B., Pelletier E., Renault P., Sicheritz-Ponten T., Turner K., Zhu H., Yu C., Li S., Jian M., Zhou Y., Li Y, Zhang X., Li S., Qin N., Yang H., Wang J., Brunak S., Doré J., Guarner F., Kristiansen K., Pedersen O., Parkhill J., Weissenbach J.; MetaHIT Consortium, Bork P., Ehrlich S.D., Wang J. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature.* 2010;464(7285):59-65. DOI 10.1038/nature08821.

Rajilić-Stojanović M., Biagi E., Heilig H.G., Kajander K., Kekkonen R.A., Tims S., de Vos W.M. Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology.* 2011;141(5):1792-1801. DOI 10.1053/j.gastro.2011.07.043 2011.

Rigsbee L., Agans R., Shankar V., Kenche H., Khamis H.J., Michail S., Paliy O. Quantitative profiling of gut microbiota of children with diarrhea-predominant irritable bowel syndrome. *Am. J. Gastroenterol.* 2012;107(11):1740-51. DOI 10.1038/ajg.2012.287.

Rodiño-Janeiro B.K., Vicario M., Alonso-Cotoner C., Pascua-García R., Santos J. A Review of Microbiota and Irritable Bowel Syndrome: Future in Therapies. *Adv. Ther.* 2018;35(3):289-310. DOI 10.1007/s12325-018-0673-5.

Saebo A., Vik E., Lange O.J., Matuszkiewicz L. Inflammatory bowel disease associated with Yersinia enterocolitica O:3 infection. *Eur. J. Intern. Med.* 2005;16(3):176-182.

Shen Z.H., Zhu C.X., Quan Y.S., Yang Z.Y., Wu S., Luo W.W., Tan B., Wang X.Y. Relationship between intestinal microbiota and ulcerative colitis: Mechanisms and clinical application of probiotics and fecal microbiota transplantation. *World J. Gastroenterol.* 2018;24(1):5-14. DOI 10.3748/wjg.v24.i1.5.

Shin N.R., Whon T.W., Bae J.W. Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.* 2015;33:496-503. DOI 10.1016/j.tibtech.2015.06.011.

Sonnenberg A., Genta R.M. Low prevalence of *Helicobacter pylori* infection among patients with inflammatory bowel disease. *Aliment. Pharmacol. Ther.* 2012;35(4):469-476. DOI 10.1111/j.1365-2036.2011.04969.x.

Stiemsma L.T., Arrieta M.C., Dimitriu P.A., Cheng J., Thorson L., Lefebvre D.L., Azad M.B., Subbarao P., Mandhane P., Becker A.,

Sears M.R., Kollmann T.R., Canadian Healthy Infant Longitudinal Development (CHILD) Study Investigators; Mohn W.W., Finlay B.B., Turvey S.E. Shifts in *Lachnospira* and *Clostridium sp.* in the 3-month stool microbiome are associated with preschool age asthma. *Clin. Sci.* (*Lond*). 2016;130(23):2199-2207. DOI 10.1042/CS20160349.

Su T., Rongbei L., Lee A., Long Ya., Du L., Lai S., Chen X., Wang L., Si J., Chung O., Chen S. Altered intestinal microbiota with increased abundance of *Prevotella* is associated with high risk of diarrhea-predominant irritable bowel syndrome. *Gastroenterol. Research Pract.* 2018;2018:6961783. DOI 10.1155/2018/6961783.

Tana C., Umesaki Y., Imaoka A., Handa T., Kanazawa M., Fukudo S. Altered profiles of intestinal microbiota and organic acids may be the origin of symptoms in irritable bowel syndrome. *Neurogastroenterol. Motil.* 2010;22(5):512-519, e114-5. DOI 10.1111/j.1365-2982.2009.01427.x.

Tap J., Derrien M., Törnblom H., Brazeilles R., Cools-Portier S., Doré J., Störsrud S., Le Nevé B., Öhman L., Simrén M. Identification of an intestinal microbiota signature associated with severity of irritable bowel syndrome. *Gastroenterol.* 2017;152(1):111-123.e8. DOI 10.1053/j.gastro.2016.09.049.

Wang Y., Qian P.Y. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One.* 2009;4:e7401. DOI 10.1371/journal.pone.0007401.

Załęski A., Banaszkiewicz A., Walkowiak J. Butyric acid in irritable bowel syndrome. *Prz. Gastroenterol.* 2013;8(6):350-353. DOI 10.5114/pg.2013.39917.

Zhuang X., Xiong L., Li L., Li M., Chen M. Alterations of gut microbiota in patients with irritable bowel syndrome: A systematic review and meta-analysis. *J. Gastroenterol. Hepatol.* 2017;32(1):28-38. DOI 10.1111/jgh.13471.

ORCID ID

A.Y. Tikunov orcid.org/0000-0000-0001-5613-5447
A.N. Shvalov orcid.org/0000-0001-6890-1575
I.V. Babkin orcid.org/orcid.org/0000-0001-7158-3774
A.V. Bardasheva orcid.org/0000-0002-3872-4122
V.V. Morozova orcid.org/ 0000-0002-0869-3476
V.V. Vlasov orcid.org/ 0000-0003-2845-2992
N.V. Tikunova orcid.org/0000-0002-1687-8278

Acknowledgements. This study was supported by the Russian Science Foundation, project 18-74-00082, and the Program of basic research for national academies of sciences in 2013–2020, project AAAA-A17-117020210027-9.

Conflict of interest. The authors declare no conflict of interest.

Received May 4, 2020. Revised February 19, 2021. Accepted March 17, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):874-881
DOI 10.18699/VJ21.101

Original Russian text www.bionet.nsc.ru/vogis/

Revealing the diversity of *Fusarium* micromycetes in agroecosystems of the North Caucasus plains for replenishing the State Collection of Phytopathogenic Microorganisms of the All-Russian Scientific Research Institute of a Phytopathology

N.S. Zhemchuzhina[1], M.I. Kiseleva[1], T.M. Kolomiets[1] ✉, I.B. Ablova[2], A.P. Glinushkin[1], S.A. Elizarova[1]

[1] All-Russian Scientific Research Institute of a Phytopathology (ARSRIP), Moscow region, Russia
[2] National Center of Grain named after P.P. Lukyanenko, Krasnodar, Russia
✉ lomi1@yandex.ru

Abstract. In order to prevent crop yield losses from the most dangerous and economically important pathogenic organisms, it is necessary not only to monitor the virulence gene pool, but also to study the nature of pathogen variability and determine the potential for the emergence of new genes and races. This requires centralized collections of fungal cultures characterized by a set of stable strains to provide for phytopathological, immunological, breeding, genetic, toxicological, parasitological and other studies. The State Collection of Phytopathogenic Microorganisms of the ARSRIP is the State Depository of Phytopathogenic microorganisms that are non-pathogenic to humans or farmed animals. Currently, it has more than 4,500 accessions of plant pathogenic strains of fungi, oomycetes, bacteria, viruses, phytoplasmas, and the collection is updated annually. For this purpose, the study of the inter- and intraspecific genetic diversity of genus *Fusarium* was carried out in agricultural systems of the Krasnodar Territory. In 2020, the State Collection of Phytopathogenic Microorganisms was supplemented with 13 strains of *Fusarium* fungi isolated from tissues of winter wheat plants collected in several locations of the Krasnodar region. The complex of *Fusarium* fungi revealed on winter wheat usually included *Fusarium oxysporum*, *F. culmorum*, *F. lolii*, *F. graminearum*, *F. fujikuroi*, *F. sporotrichioides*, etc. The effect of the preceding crop on the frequency of *Fusarium* species isolated from winter wheat was observed. After series cloning of collected isolates, 21 strains of different fungal species characterized by stable morphology traits and known pathogenic and phytotoxic properties were selected for collection replenishment. Significant differences in pathogenic activity were revealed between fungi belonging to either the same or different species; the manifestation of this activity varied from the absence of any effect of spore suspensions on seedling development to a complete inhibition of their growth. The phytotoxic activity towards wheat seedlings varied from medium to high. Species possessing a high intensity of phytotoxic activities are the most dangerous for wheat, since they promote accumulation of dangerous phytotoxins in plant tissues.
Key words: microorganism collections; micromycetes; genetic diversity; winter wheat; plant pathogens; *Fusarium*.

For citation: Zhemchuzhina N.S., Kiseleva M.I., Kolomiets T.M., Ablova I.B., Glinushkin A.P., Elizarova S.A. Revealing the diversity of *Fusarium* micromycetes in agroecosystems of the North Caucasus plains for replenishing the State Collection of Phytopathogenic Microorganisms of the All-Russian Scientific Research Institute of a Phytopathology. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8):874-881. DOI 10.18699/VJ21.101

Выявление разнообразия микромицетов рода *Fusarium* в агроэкосистемах равнинной части Северного Кавказа для пополнения Государственной коллекции фитопатогенных микроорганизмов ФГБНУ ВНИИФ

Н.С. Жемчужина[1], М.И. Киселева[1], Т.М. Коломиец[1] ✉, И.Б. Аблова[2], А.П. Глинушкин[1], С.А. Елизарова[1]

[1] Всероссийский научно-исследовательский институт фитопатологии (ВНИИФ), Московская область, Россия
[2] Национальный центр зерна им. П.П. Лукьяненко, Краснодар, Россия
✉ lomi1@yandex.ru

Аннотация. С целью предотвращения потерь урожая сельскохозяйственных культур от наиболее опасных и экономически значимых патогенных организмов нужно не только проводить контроль генофонда вирулентности, но и изучать характер изменчивости патогенов, определять потенциальную возможность появления новых генов и рас. Для этого необходимы централизованные коллекции живых культур, характеризующиеся набором

стабильных штаммов для обеспечения фитопатологических, иммунологических, селекционных, генетических, токсикологических, паразитологических и других исследований. Государственная коллекция фитопатогенных микроорганизмов Всероссийского научно-исследовательского института фитопатологии – государственный депозитарий фитопатогенных микроорганизмов, непатогенных для человека и сельскохозяйственных животных. В настоящее время она насчитывает более 4500 единиц хранения патогенных для растений штаммов – грибов, оомицетов, бактерий, вирусов, фитоплазм – и ежегодно пополняется. С этой целью было проведено изучение межвидового и внутривидового генетического разнообразия фузариевых грибов в агроэкосистемах Краснодарского края. В 2020 г. Государственная коллекция фитопатогенных микроорганизмов была пополнена штаммами 13 видов грибов рода *Fusarium*, выделенными из тканей растений озимой пшеницы из нескольких районов Краснодарского края. В комплексе фузариевых грибов озимой пшеницы наиболее часто выявляли *Fusarium oxysporum*, *F. culmorum*, *F. lolii*, *F. graminearum*, *F. fujikuroi*, *F. sporotrichioides* и др. Отмечено влияние предшественника озимой пшеницы на частоту выделяемых видов грибов рода *Fusarium*. После серии моноконидиальных клонирований изолятов в коллекцию был отобран 21 штамм грибов разных видов со стабильными морфолого-культуральными признаками и известными патогенными и фитотоксичными свойствами. Патогенная активность у грибов как между видами *Fusarium*, так и в пределах одного вида существенно различалась: от отсутствия признаков влияния споровых суспензий на развитие проростков до полного их угнетения. Фитотоксическая активность в отношении проростков пшеницы варьировала от средней до высокой. Наибольшую опасность для проростков пшеницы представляют виды, обладающие высокой интенсивностью проявления фитотоксической активности, так как они способствуют накоплению опасных фитотоксинов в тканях растений.

Ключевые слова: коллекции микроорганизмов; микромицеты; генетическое разнообразие; озимая пшеница; фитопатогены; *Fusarium*.

Introduction

To successfully solve the problems of food security of the country, it is necessary to create varieties resistant to particularly dangerous diseases. In order to prevent crop yield losses from the most dangerous and economically significant pathogenic organisms, it is necessary not only to monitor the virulence gene pool, but also to study the nature of pathogen variability, determine the potential for the appearance of new genes and possibly dangerous races in different fungi populations. This requires centralized collections of cultures characterized by a set of stable properties to provide for phytopathological, immunological, breeding, genetic, toxicological, parasitological and other studies. Such collections of phytopathogenic organisms have been created and are successfully functioning in most developed countries of the world.

The State Collection of Phytopathogenic Microorganisms of the All-Russian Scientific Research Institute of a Phytopathology (ARSRIP) is the main gene pool of races, biotypes, pathotypes of phytopathogenic fungi, bacteria and viruses distributed over the vast territory of the Russian Federation. This is the first such gene pool created in Russia. Until recently, there were only scattered working collections of individual species of phytopathogenic microorganisms in various institutions and departments of the institute. Collection of phytopathogenic microorganisms of ARSRIP by the decree of the Government of the Russian Federation "On measures for the conservation and rational use of collections of microorganisms" dated 24.06.1996. No. 725-47c was given the name "State Collection of Phytopathogenic Microorganisms and Varieties-Identifiers (Differentiators) of Pathogenic Strains of Microorganisms" and the status of the State Depository of phytopathogenic microorganisms that are not pathogenic to humans and farm animals was determined. Currently, it has more than 4,500 storage units of plant pathogenic strains – fungi, oomycetes, bacteria, viruses, phytoplasmas – and is updated annually. For this purpose, the study of inter- and intraspecific diversity of Fusarium fungi in agricultural systems of the Krasnodar Territory was carried out.

According to the literature, facultative parasites from the genus *Fusarium* are often observed on winter wheat. These micromycetes are well adapted to changing external environmental factors, which ensures their survival in a wide range of weather conditions, and therefore are distributed almost everywhere where winter wheat is cultivated (Rukavitsina, 2008; Chulkina et al., 2009; Toropova et al., 2013). Monitoring the structure and localization of *Fusarium* populations in wheat ecosystems is of great practical importance not only for selecting disease-resistant varieties, but also for increasing the effectiveness of protective measures and improving the environmental situation of agricultural crops.

Recently in the southern regions of Russia, where winter wheat is widely cultivated, there has been an increase in diseases caused by fungi of the genus *Fusarium* (Zhalieva, 2010). It is known that 28 species of fungi of this genus parasitize wheat. The species *Fusarium graminearum*, *F. poae*, *F. sporotrichioides*, *F. tricinctum*, *F. nivale* prevail on wheat in the North Caucasus. As a rule, they are observed as pathogens of root rot, causing the weakening and death of seedlings, reducing the productivity potential of affected adult plants.

In some years, Fusarium head is widely spread, causing significant damage to grain production. On vegetative and generative organs of plants, the species composition of fungi can be ambiguous depending on weather conditions, the stability of cultivated varieties, wheat precursors, agricultural technology and many other factors (Chulkina et al., 2009; Zhalieva, 2010).

Assessment of the diversity of morphological features of *Fusarium* spp., identification of the amplitude of their variability, including the level of pathogenicity and phytotoxicity, is important for the selection and replenishment of the collection with micromycete strains (Booth, 1971; Thrane et al., 2004; Kolomiets et al., 2018).

The need to preserve the material of *Fusarium* spp. strains and isolates and the constant selection of samples to replenish the collection is explained by the relevance of conducting scientific research for the development of methods of biolo-

gical protection, to study the dynamics of the development and settlement of fungi, to assess their pathogenic and toxic activity on host plants. The collection material is also necessary for the evaluation and selection of wheat samples for breeding for resistance to disease (Dubovoy et al., 2016; Zhemchuzhina, Elizarova, 2019).

In this regard, the main tasks of a collection of unique *Fusarium* spp. isolates is not only to preserve the viability of spores and their genetic stability according to morphological and cultural characteristics for a long time, but also to replenish the funds with new isolates with a different spectrum of pathogenic properties, as well as to expand the range of geographical territories for selecting isolates (Gagkaeva, Levitin, 2005; Gagkaeva et al., 2008).

To fulfill these tasks, samples of infected plants received annually from various regions of the country are mycology studied, and the most pathogenic and phytotoxic samples are selected for the collection. Strains and isolates from the genus *Fusarium* have an important place in the collection, which serves, as has already been mentioned, to maintain the strains of these microorganisms in a viable state, preserve their pathogenic properties and provide infectious material for phytopathological, immunological, breeding, genetic and toxicological studies (Kolomiets, Zhemchuzhina, 2018).

The purpose of these studies was to assess the species and intraspecific diversity of micromycetes of the genus *Fusarium* in wheat crops in agroecosystems of the Krasnodar Territory for the selection and replenishment in 2020 of the State Collection of Phytopathogenic Microorganisms (SCPM) with strains and isolates of fungi isolated from the roots and leaves of vegetating winter wheat plants in several districts of the Krasnodar Territory.

Materials and methods

The research material was plants of zoned varieties of winter wheat with signs of fungal infections on the leaves and roots. The samples were selected in the second decade of May 2019 from winter wheat crops for different predecessors in the Pavlovsky, Korenovsky, Ust-Labinsky, Kanevsky and Primorsko-Akhtarsky districts of the Krasnodar Territory. The samples contained 10–20 wheat plants in the earing-grain formation phase. All the works were performed using the equipment of the Center of Collective Usage SCPM ARSRIP (http://www.vniif.ru/vniif/page/ckp-gkmf/1373).

Fungi were isolated from the affected plants using 2 % potato-glucose and potato-carrot agar. Fungi from wheat samples were isolated according to the standard method (Bilai, 1977; Bilai, Ellanskaya, 1982). The infected plants of each sample washed with tap water were cut into fragments of 5–10 mm in size, sterilized in 50 % alcohol for 20–30 seconds and, under aseptic conditions, laid out on the surface of 2 % potato-glucose agar in Petri dishes (4–6 fragments each). Each sample was represented by at least 150–00 fragments of the affected tissue. Petri dishes were placed in a thermostat with a temperature of 22–24 °C.

Observation of the development of fungi was carried out daily. As the fungal colonies grew, a piece of mycelium was sifted onto the nutrient medium in the center of the Petri dish. Fungal cultures were examined under a microscope. The species of fungi of the genus *Fusarium* were determined by the main morphological features of colonies and spores: by the growth rate, the color of the mycelium and its structure, pigmentation; by the shape and size of the apical and basal cells of macroconidia, by the presence of microconidia. 300 conidia were estimated for an average size of macroconidia. For determining the species of the fungus (Gerlach, Nirenberg, 1982; Leslie, Summerell, 2006; Dictionary..., 2008; Watanabe, 2010) were used as reference literature.

The assessment of the degree of sporulation was carried out on 14-day colonies of the fungus. At the same time, the results of sporulation were determined by the average value of the number of spores per cup when flushing from 10 Petri dishes of one morphotype. The sporulating ability of fungal colonies was determined by the standard method of counting spores in the Goryaev chamber (Bilai, 1977).

Series of monospore cloning of micromycete isolates were carried out according to the generally accepted method for the selection of strains of Fusarium fungi stable by morphological and cultural characteristics (Bilai, 1977; Bilai, Ellanskaya, 1982).

Fungal isolates isolated from the affected wheat samples were stored in a refrigerator at a temperature of 7–10 °C in biological test tubes on the nutrient medium – potato-glucose agar (Bilai, Ellanskaya, 1982).

The pathogenic and toxic properties of the strains were studied using the method of bioassay on seeds. The pathogenicity of spore suspensions and phytotoxicity of filtrates of culture fluids (FCF) of fungi of the genus *Fusarium* were tested on wheat seeds (Mironovskaya 808 variety). The degree of pathogenicity and toxicity of the strains was judged by the effect of suspensions of conidia and FCF on the germination of wheat seeds, the development of sprouts and primary roots of seedlings. However, the main indicator was considered the length of the roots.

The degree of pathogenicity and toxicity was determined on the 5th day from the beginning of seed germination. If the length of seedlings and roots (in mm) in the experimental version was 0–30 % of the control length, this indicated strong pathogenic (P) and strong toxic (T) activity of the fungus; 31–50 % – moderate pathogenicity (MP) and moderate toxicity (MT); 51–70 % – weak pathogenicity (WP) and weak toxicity (WT); 71–100 % – non-pathogenic (NP) and non-toxic (NT) properties of the isolate. The length of the sprouts and primary roots of seeds sprouted in water was considered as control and was taken as 100 %.

Results and discussion

During the mycological studies of experimental samples it was noted that fungi of the genus *Fusarium* had the same symptoms on plant organs, but the pieces of tissue of different organs, washed and flamed over the fire, decomposed into wet chambers, formed a characteristic mycelium and conidia for 3–5 days, which made it possible to identify the type of micromycete. The study of the main micro- and macromorphological features of fungi in culture by the presence, shape and size of macroconidia and microconidia (if present), the growth rate of the colony, the color and structure of the mycelium, carried out on more than 400 isolates of fungi, allowed us to identify the following 13 species of the genus *Fusarium*: *F. oxysporum* Schlecht., *F. culmorum* (Sm.) Sacc.,

Table 1. Species of the genus *Fusarium* found on winter wheat crops in the Krasnodar Territory, 2019, (in %)

Species of fungus	Precursor					Frequency of occurrence of species
	wheat	pure soil	sunflower	corn	peas	
F. acuminatum	1.2	0.9	0.7	0	0	2.8
F. avenaceum	1.2	0	0.7	1.2	0.9	3.9
F. chlamydosporum	0.7	0	0.5	0	0	1.2
F. culmorum	5.8	1.2	5.3	2.8	3.5	18.2
F. equiseti (*F. gibbosum*)	0	0	1.2	0.7	0	1.8
F. graminearum	5.5	1.2	0.5	5.1	0	12.0
F. lolii	6.2	2.8	3.5	2.8	0	15.8
F. moniliforme (*F. fujikuroi*)	2.1	1.2	2.3	4.2	0	9.7
F. oxysporum	7.2	1.2	5.5	2.3	4.2	20.1
F. poae	1.2	0.7	0.7	0	1.2	3.6
F. sambucinum (*F. roseum*)	1.2	0.5	1.2	0.5	0	3.2
F. solani	0	0	0.2	0	0.7	0.9
F. sporotrichioides	1.8	1.8	0.9	2.5	0	7.1
Total isolates, %	33.9	11.3	22.6	21.7	10.4	100
Total isolates, unit	147	49	98	94	45	433

F. lolii (Wm. G. Sm.) Sacc., *F. graminearum* Schwabe, *F. moniliforme* J. Sheld. (syn. *F. fujikuroi* Nirenberg), *F. sporotrichioides* Swerb., *F. avenaceum* (Fr.) Sacc., *F. poae* (Peck) Wollenw., *F. sambucinum* Fuckel (syn. *F. roseum* Link), *F. acuminatum* Ellis & Everh., *F. equiseti* (Corda) Sacc. (syn. *F. gibbosum* Appel & Wollenw.), *F. chlamydosporum* Wollenw. & Reinking, *F. solani* (Mart.) Sacc. (Table 1).

In the complex of Fusarium fungi of winter wheat in the Krasnodar Territory in 2019, isolates of *F. oxysporum* (Schlecht.) Snyd. et Hans. were detected most often (20.1 %). The fungus was recorded on winter wheat crops in the Ust-Labinsk, Pavlovsky, Korenovsky, Primorsko-Akhtarsky districts. Undoubtedly, the culture preceding winter wheat had an impact on the frequency of occurrence of *F. oxysporum*. Thus, the share of those isolates according to the wheat precursor was 7.2 %, sunflower – 5.5, peas – 4.2, corn –1.2 %. Often, cultures of *F. avenaceum*, *F. acuminatum*, *F. sambucinum* were noted together with this fungus, as well as *Alternaria* spp., bacteria, etc.

The second place in frequency of occurrence was occupied by *F. culmorum* (Sm.) Sacc. Isolates of the fungus were obtained from samples of the affected roots and the basal part of the stem of winter wheat in almost all the surveyed areas. The share of this pathogen in the complex of fungi of the genus *Fusarium* was 18.2 %. Often, isolates of *F. culmorum* were found in samples of winter wheat, the precursors of which were wheat, sunflower and peas.

F. lolii (Wm. G. Sm.) Sacc. (teleomorph *Gibberella cyanea* (Sollm.) Wr., syn. *F. heterosporum* Nees.), as a rule, were isolated from the highly rot-damaged and dried roots of winter wheat, the precursors of which were wheat (6.2 %) and sunflower (3.5 %).

Isolates of *F. graminearum* Schwabe (teleomorph *G. zeae* (Schwein.) Petch.) were discovered on the roots and basal stems of winter wheat in most areas of infectious material, and the fungus was isolated more frequently if predecessors were wheat and corn (5.1 and 5 %, respectively).

Isolates of *F. moniliforme* J. Sheld. (teleomorph *G. moniliformis* Wineland; syn. *F. fujikuroi* Nirenberg), the causative agent of pink mold and root rot of cereals, were found on the leaves, stems and roots of winter wheat in the Primorsko-Akhtarsky, Pavlovsky and Kanevsky districts. More often, the fungus was isolated from wheat, the precursors of which were corn (4.2 %), sunflower (2.3 %) and wheat (2.1 %).

Isolates of *F. sporotrichioides* Sherb. were isolated from the affected roots, root neck and stems of winter wheat (7.1 %) from the Pavlovsky and Korenovsky districts.

Isolates of *F. avenaceum* (Fr.) Sacc. (teleomorph: *G. avenacea* Cook) were found on leaves, ground parts of stems and roots of winter wheat from the Pavlovsky and Korenovsky districts. In the complex of fungi from the genus *Fusarium* isolated from winter wheat samples, the frequency of occurrence of *F. avenaceum* was 3.9 %. The fungus was observed in samples of winter wheat, the previous crops of which were wheat and corn.

The isolates of *F. poae* (Peck) Wollenw. were found with low frequency (3.6 %) in the Pavlovsky district. More often, isolates of this type were noted on samples of wheat, the precursors of which were wheat and peas.

It should be noted that in some cases two or more phytopathogens from the genus *Fusarium* were isolated from one sample of the affected winter wheat tissue. Such isolates of *F. acuminatum* Ellis & Verh. (telemorph *G. acuminate* Wr.) were isolated together with *F. oxysporum* more than half of the cases.

Table 2. Characteristics of fungal strains of 13 species of the genus *Fusarium* selected in the SCPM ARSRIP, according to macro- and micromorphological properties

Fungus	Code of strains	Morphological characteristics of colonies of fungal strains of 13 species of the genus *Fusarium*		Sporulation, million/ Petri dish	The size of macroconidia, microns			The presence of micro-conidia
					longitudinal		cross	
		Mycelium	Reverse		X_{min}–X_{max}	LSD_{05}	X_{min}–X_{max}	
F. avenaceum	CK-7l	Dark pink, weak developed	Brown	195.0 ± 4.6	30.5–85.7	35.7	3.3–6.1	–
F. acuminatum	CK-13k	White-pink, low with lysis sectors	Light brown	210.3 ± 6.8	18.5–22.4	5.7	4.0–5.0	–
	CK-16k	White-pink, cotton-like	Dark brown	195.5 ± 5.2	18.1–22.7	5.9	4.0–5.0	–
F. chlamydosporum	CK-3k	Light cream, velvety	Cream	45.7 ± 3.2	30.5–40.5	11.1	3.6–4.1	+
F. culmorum	CK-14k-1	Olive-red, loose-flaky	Brown	201.5 ± 5.8	16.0–48.5	18.5	3.9–6.5	–
	CK-14k-5	Pale pink, fluffy		115.7 ± 2.7	16.7–49.3	21.1	3.8–6.7	–
F. equiseti	CK-8l	White-cream, low, loose-fluffy	Brown	195.1 ± 7.8	15.5–70.5	38.8	4.0–4.5	+
F. graminearum	CK-10k-2	Olive-pink, flaky-fluffy	Brown	112.0 ± 2.2	21.5–75.0	31.5	4.3–4.5	–
	CK-11k-3		Dark brown	175.5 ± 3.7	23.1–77.7	28.7	4.3–4.5	–
F. lolii	CK-6k-1	Light cream, fluffy	Light cream	55.2 ± 3.2	20.5–35.5	13.9	4.0–4.3	–
	CK-6l	Milk light cream		140.5 ± 4.2	19.7–35.5	15.3	4.0–4.3	–
	CK-6k-5			109.1 ± 8.7	19.1–36.7	14.5	4.1–4.3	–
F. moniliforme	CK-4l	Light cream to purple, creeping	Light brown	75.3 ± 3.3	23.0–60.5	25.8	3.6–4.0	+
F. oxysporum	CK-5k	Pale lilac, creeping	Light brown	23.3 ± 4.1	28.3–35.2	17.5	3.5–4.5	+
	CK-9l	White with purple areas, cotton-like	Dark purple	110.7 ± 3.4	26.5–37.1	23.1	3.3–4.5	+
	CK-9k		Pale purple	95.5 ± 2.5	29.4–40.0	12.7	3.2–4.6	+
F. poae	CK-7k	Cream-pink, creeping	Light cream	24.7 ± 7.7	17.2–40.5	22.6	3.5–5.5	+
	CK-15k	Olive-pink, fluffy	Brown	10.5 ± 2.9	17.0–40.1	20.3	3.5–5.5	+
F. sambucinum	CK-5k-1	Light cream, loose	Cream	70.0 ± 4.3	17.5–24.5	8.2	3.6–4.5	–
F. solani	CK-13k-1	Cream-pink, felt-fluffy	Brown	105.1 ± 5.0	21.5–42.5	19.4	3.5–4.9	+
F. sporotrichioides	CK-4k	White-pink, fluffy	Brown	150.4 ± 5.8	26.0–45.0	15.1	3.5–5.0	+

Table 3. Characteristics of *Fusarium* fungal strains by pathogenicity of spore suspensions and phytotoxicity of culture liquid on wheat seedlings of cultivar Mironovskaya 808 (in % of control)

Code of strain	Pathogenicity (spore suspension)				Toxicity (culture liquid)			
	Seed germination, %	Sprout length, %	Root length, %	Degree of influence	Seed germination, %	Sprout length, %	Root length, %	Degree of influence
F. avenaceum (Fr.) Sacc.								
CK-7l	96.7	92.3 ± 1.5	63.5 ± 1.4	WP	100.0	96.9 ± 2.1	71.3 ± 2.0	NT
F. acuminatum Ellis & Everh.								
CK-13k	100.0	102.3 ± 1.9	93.9 ± 2.1	NP	96.7	85.0 ± 3.7	49.2 ± 2.4	MT
CK-16k	100.0	101.9 ± 3.1	85.7 ± 2.3	NP	100.0	90.6 ± 4.0	61.4 ± 2.5	WT
F. chlamydosporum Wollenw. & Reinking								
CK-3k	100.0	90.9 ± 1.8	68.9 ± 1.8	WP	96.7	88.8 ± 2.3	48.8 ± 2.2	MT
F. culmorum (Sm.) Sacc.								
CK-14k-1	86.7	79.5 ± 3.2	29.1 ± 2.5	P	85.5	69.4 ± 5.0	13.0 ± 1.2	T
CK-14k-5	95.5	84.1 ± 2.6	47.5 ± 3.1	MP	83.3	63.3 ± 1.5	22.3 ± 1.5	T
F. equiseti (Corda) Sacc.								
CK-8l	100.0	94.4 ± 2.2	87.4 ± 2.1	NP	100.0	87.1 ± 2.5	87.7 ± 2.7	NT
F. graminearum Schwabe								
CK-10k-2	96.4	38.5 ± 5.2	12.6 ± 5.1	P	63.3	36.2 ± 3.1	5.7 ± 2.3	T
CK-11k-3	96.7	72.0 ± 5.5	56.7 ± 5.2	MP	70.0	24.4 ± 1.6	10.9 ± 1.2	T
F. lolii (Wm. G. Sm.) Sacc.								
CK-6k-1	100.0	98.1 ± 2.5	60.8 ± 2.2	WP	100.0	92.2 ± 1.8	44.4 ± 1.7	MT
CK-6l	100.0	97.2 ± 2.4	64.9 ± 1.6	WP	96.7	84.3 ± 3.8	52.8 ± 2.9	WT
CK-6k-5	95.5	67.5 ± 4.2	27. ± 2.7	P	80.0	71.4 ± 5.0	28.6 ± 2.2	T
F. moniliforme J. Sheld.								
CK-4l	100.0	101.9 ± 3.0	103.9 ± 4.0	NP	100.0	96.2 ± 3.1	65.0 ± 1.6	WT
F. oxysporum Schlecht.								
CK-5k	100.0	93.1 ± 3.8	67.7 ± 3.0	NP	100.0	39.0 ± 1.7	13.0 ± 1.7	T
CK-9l	80.0	33.3 ± 3.1	27.7 ± 2.3	P	100.0	79.0 ± 1.8	63.0 ± 1.9	WT
CK-9k	91.5	64.3 ± 2.4	28.1 ± 1.7	P	100.0	86.7 ± 2.4	61.3 ± 1.7	WT
F. poae (Peck) Wollenw.								
CK-7k	100.0	101.3 ± 1.8	93.8 ± 2.2	NP	100.0	92.3 ± 2.6	65.4 ± 1.8	WT
CK-15k	100.0	99.1 ± 2.8	95.3 ± 4.1	NP	100.0	96.9 ± 2.0	74.2 ± 2.9	NT
F. sambucinum Fuckel								
CK-5k-1	90.0	58.6 ± 5.1	19.5 ± 1.6	P	90.0	83.1 ± 4.1	61.6 ± 2.7	WT
F. solani (Mart.) Sacc.								
CK-13k-1	100.0	89.1 ± 3.8	90.3 ± 2.1	NP	96.7	84.3 ± 2.8	62.8 ± 1.9	WT
F. sporotrichioides Swerb.								
CK-4k	100.0	38.5 ± 3.5	14.8 ± 1.9	P	100.0	29.7 ± 1.8	6.3 ± 1.0	T

Note. NP/NT is non-pathogenic/non-toxic; WP/WT – weakly pathogenic/weakly toxic; MP/MT – moderately pathogenic/moderately toxic; P/T – pathogenic/toxic.

A similar pattern was noted for *F. sambucinum* Fuckel (teleomorph *G. pulicaris* (Fr.) Sacc.). This fungus, regardless of the location on the plant (leaf, stem, root), was always isolated together with *F. oxysporum*, while it was often accompanied by a bacterial infection.

F. acuminatum isolates were found in the Ust-Labinsk and Pavlovsky districts with low frequency (2.8 %) in the affected roots of winter wheat, the precursors of which were wheat, steam and sunflower.

Isolates of *F. equiseti* (Corda) Sacc. (teleomorph *G. intracans* Wollenw.; syn. *F. gibbosum* Appel & Wollenw. Emend Bilai) were detected mainly on browned winter wheat stalks in the Ust-Labinsk and Korenovsky districts.

The proportion of isolates of *F. chlamydosporum* Wollenw. & Reinking in the complex of Fusarium fungi did not exceed 1.2 %. The fungus was isolated from the roots of two wheat samples from the Korenovsky district. Along with *F. chlamydosporum*, saprophytic and pathogenic fungal species were abundantly isolated from the same roots.

Several isolates of *F. solani* (Mart.) Sacc. (teleomorph *Nectria haematococca* Berk. & Broome) were found in the Pavlovsky district on the roots of wheat.

As it was noted earlier, based on the study of morphological features, the obtained isolates of fungi of the genus *Fusarium* are assigned to 13 taxonomic groups. After a series of monoconidial cloning of isolates, strains of fungi of different species with stable morphological and cultural characteristics were selected for the collection. When selecting fungi cultures for the collection, special attention was paid to the macro- and micromorphological features characteristic of each species.

The *Fusarium* spp. strains differed in the morphology, in the size and shape of macro- and microconidia, and in the sporulation of colonies. Differences between strains within the same species were often noted only when studying macromorphological features – the color and structure of the mycelium, sporulation. When analyzing the data of micromorphological features, i. e., the shape and size of conidia, the method of their formation, the differences between the strains of the same species were minimal.

21 strains of *Fusarium* spp. identified on winter wheat crops of the North Caucasus in 2019 were transferred to the collection (Table 2).

The obtained biological material of *Fusarium* spp. was studied by the degree of pathogenic and phytotoxic severity. The results of the influence of spore suspensions and metabolites of filtrates of culture fluids of 21 strains of fungi from the genus *Fusarium* on the development of wheat seedlings in cultivar Mironovskaya 808 (seed germination, sprout and root length) were shown in Table 3.

It was shown that the strains of fungi within the same species differed in the pathogenic or phytotoxic degrees. The strains of *F. oxysporum* (CK-5k, CK-9l, CK-9k) and *F. lolii* (CK-6k-1, CK-6l, CK-6k-5) had a wide intraspecific diversity according to these characteristics. Among them, there were different categories – from pathogenic/toxic to non-pathogenic/slightly toxic.

Phytotoxic and pathogenic isolates of *F. culmorum* and *F. graminearum* suppressed the development of seedlings of cv. Mironovskaya 808 to a wide extent.

Isolates of *F. acuminatum* (CK-13k, CK-16k) were found to be non-pathogenic to the seedlings of the tester variety, but had weak and moderate phytotoxicity.

Isolates of *F. avenaceum* (CK-7l), *F. equiseti* (CK-8l), *F. poae* (CK-7k, CK-6k-1) and *F. chlamydosporum* (CK-3k) and others were characterized by very weak pathogenic and phytotoxic properties.

It was found that spore suspensions of fungal isolates had little effect on seed germination (80–100 %), but subsequently affected the development of seedlings: pathogenic isolates of fungi inhibited their growth (up to 33.3 % strain of *F. oxysporum* CK-9l) or non-pathogenic ones stimulated it (up to 102.3 % strain of *F. acuminatum* CK-13k). A stronger effect of spore suspensions on the growth and development of primary roots was noted (12.6–95.3 %).

Seeds' treatment of filtrates of culture fluids of *Fusarium* strains had a weak effect on their germination (63.3–100 %), although in the future the intensity of development of seed seedlings significantly slowed down. The length of seedlings under the action of filtrates of fungal culture fluids compared to the control was 24.4–96.9 %. The average length of the primary roots was 5.7–74.2 %, which allowed the isolates to be grouped according to the degree of toxicity.

The obtained results indicate that the Krasnodar populations of Fusarium differ by morphological, pathogenic and phytotoxic characteristics.

Conclusion

The influence of the precursor of winter wheat, from which experimental samples were taken, on the frequency of isolated of the genus *Fusarium* was noted. It is shown that the pathogenic activity of fungi both between *Fusarium* species and within the same species differs significantly: from the absence of signs of the influence of spore suspensions on the development of seedlings to their complete suppression. Phytotoxic activity against wheat seedlings varied from medium to high. The greatest danger for wheat seedlings is represented by species with a high intensity of phytotoxic activity associated with the fact that they contribute to the accumulation of dangerous toxins in plant tissues.

Based on the results of the data obtained, the strains of 13 species of the genus *Fusarium* from the agroecosystems of the lowland part of the North Caucasus were selected and placed in the collection. All strains are characterized by morphological, pathogenic and phytotoxic properties.

References

Bilai V.I. Fusaria. Kiev: Naukova Dumka Publ., 1977. (in Russian)

Bilai V.I., Ellanskaya I.A. Basic mycological methods in plant pathology. In: Methods of Experimental Mycology. Kiev: Naukova Dumka Publ., 1982;418-431. (in Russian)

Booth C. The Genus Fusarium. Kew, Surrey, England: Commonwealth Mycological Inst., 1971.

Chulkina V.A., Toropova E.Yu., Stetsov G.Ya. Integrated Plant Protection: Phytosanitary Systems and Technologies. Moscow: Kolos Publ., 2009. (in Russian)

Dictionary of the Fungi. Eds. P.M. Kirk, P.F. Cannon, D.W. Minter, J.A. Stalpers. 10st ed. CAB International, 2008.

Dubovoy V., Zhemchuzhina N., Elizarova S., Gorelov P. State Collection of Phytopathogenic Microorganisms from All-Russian Research Institute of Phytopdthology. *Analitika = Analytics.* 2016;1: 76-78. (in Russian)

Gagkaeva T.Yu., Gavrilova O.P., Levitin M.M. Modern taxonomy of fungi from the genus *Fusarium* of the section Sporotrichiella. *Mikologiya i Fitopatologiya = Mycology and Phytopathology.* 2008; 42(3):201-214. (in Russian)

Gagkaeva T.Yu., Levitin M.M. The current state of the taxonomy of fungi of the *Gibberella fujikuroi* complex. *Mikologiya i Fitopatologiya = Mycology and Phytopathology.* 2005;39(6):3-14. (in Russian)

Gerlach W., Nirenberg H. The Genus *Fusarium* – a Pictoral Atlas. Mitt. Biol. Bundesans Land-Forstwirtsch Berlin-Dahlem, 1982.

Kolomiets T., Zhemchuzhina N. Genetic resources of the State Collection of Phytopathogenic Microorganisms of the All-Russian Research Institute of Phytopathology (ARRIP). In: Conf. proc. of XXXVII Annual Meeting of the European Culture Collection Organisation, 13-15 September 2018, Moscow. 2018;45-46.

Kolomiets T.M., Zhemchuzhina N.S., Pankratova L.F., Alexandrova A.V., Kiseleva M.I., Elizarova S.A. Micromycetes of the State Collection of phytopathogenic microorganisms causing mycoses and mycotoxicoses in humans and animals. *Uspekhi Meditsinskoy Mikologii = Advances in Medical Mycology.* 2018;19:371-372. (in Russian)

Leslie J.F., Summerell B.A. The Fusarium Laboratory Manual. 2st ed. Iowa: Blackwell Publishing, 2006.

Rukavitsina I.V. Biology and Ecology of Alternariosis, Fusariosis and Helminthosporiosis of Wheat. Shortandy, 2008. (in Russian)

Thrane U., Adler A., Clasen P.E. Diversity in metabolite production by *Fusarium langsethiae*, *Fusarium poae*, and *Fusarium sporotrichioides*. *Int. J. Food Microbiol.* 2004;95:257-266. DOI 10.1016/j.ijfoodmicro.2003.12.005.

Toropova E.Yu., Kazakova O.A., Vorobyeva I.G., Selyuk M.P. Fusarium root rot of grain crops in West Siberia and Transuralia. *Zashchita i Karantin Rasteniy = Plant Protection and Quarantine.* 2013;9:23-26. (in Russian)

Watanabe T. Pictorial Atlas of Soil and Seed Fungi. Morphologies of Cultured Fungi and Key to Species. 3st ed. New York: CRC Press, 2010.

Zhalieva L.D. Fungi of the genus *Fusarium* in the agrocenosis of winter wheat in the Krasnodar region: Parasitism and symbiosis. *Immunopatologiya, Allergologiya, Infektologiya = Immunopathology, Allergology, Infectology*. 2010;1:101. (in Russian)

Zhemchuzhina N.S., Elizarova S.A. The use of strains of the State Collection of Phytopathogenic Microorganisms of the All-Russia Research Institute of Phytopathology in phytopathological, breeding, and genetic studies. In: Science, Production, Business: current state and ways of innovative development of the agricultural sector by the example of the Baiserke-Agro Agricultural Holding. Proceedings of the International Scientific and Practical Conference dedicated to the 70th anniversary of the Honored Worker of the Republic of Kazakhstan Temirkhan Dosmukhambetov (April 4-5, 2019, Almaty, Kazakhstan). Almaty: Luxe Media Group Publ., 2019;64-66. (in Russian)

ORCID ID

T.M. Kolomiets orcid.org/0000-0002-1897-2380
I.B. Ablova orcid.org/0000-0002-3454-9988

Conflict of interest. The authors declare no conflict of interest.

Received July 16, 2021. Revised August 30, 2021. Accepted August 31, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):882-888
DOI 10.18699/VJ21.102

Original Russian text www.bionet.nsc.ru/vogis/

Creation of a recombinant *Komagataella phaffii* strain, a producer of proteinase K from *Tritirachium album*

A.B. Beklemishev[1, 2] ✉, M.B. Pykhtina[1, 2], Ya.M. Kulikov[1, 2], T.N. Goryachkovskaya[2], D.V. Bochkov[2], S.V. Sergeeva[2], A.R. Vasileva[2], V.P. Romanov[2], D.S. Novikova[3], S.E. Peltek[2] ✉

[1] Federal Research Center of Fundamental and Translational Medicine, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Efirnoe Joint-Stock Company, Alekseyevka, Belgorod region, Russia
✉ beklem@niibch.ru; peltek@bionet.nsc.ru

Abstract. The objects of the study were recombinant clones of *Komagataella phaffii* K51 carrying the heterologous proteinase K (PK-w) gene from *Tritirachium album* integrated into their genome as well as samples of recombinant proteinase K isolated from these clones. The aims of this work were i) to determine whether it is possible to create recombinant *K. phaffii* K51 clones overexpressing functionally active proteinase K from *T. album* and ii) to analyze the enzymatic activity of the resulting recombinant enzyme. The following methods were used: computational analysis of primary structure of the proteinase K gene, molecular biological methods (PCR, electrophoresis of DNA in an agarose gel, electrophoresis of proteins in an SDS polyacrylamide gel under denaturing conditions, spectrophotometry, and quantitative assays of protease activity), and genetic engineering techniques (cloning and selection of genes in bacterial cells *Escherichia coli* TOP10 and in the methylotrophic yeast *K. phaffii* K51). The gene encoding natural proteinase K (*PK-w*) was designed and optimized for expression in *K. phaffii* K51. The proteinase K gene was synthesized and cloned within the plasmid pPICZα-A vector in *E. coli* TOP10 cells. The proteinase K gene was inserted into pPICZα-A in such a way that – at a subsequent stage of transfection into yeast cells – it was efficiently expressed under the control of the promoter and terminator of the *AOX1* gene, and the product of the exogenous gene contained the signal peptide of the *Saccharomyces cerevisiae* α-factor to ensure the protein's secretion into the culture medium. The resultant recombinant plasmid (pPICZα-A/PK-w) was transfected into *K. phaffii* K51 cells. A recombinant *K. phaffii* K51 clone was obtained that carried the synthetic proteinase K gene and ensured its effective expression and secretion into the culture medium. An approximate productivity of the yeast recombinant clones for recombinant proteinase K was 25 μg/mL after 4 days of cultivation. The resulting recombinant protease has a high specific proteolytic activity: ~5000 U/mg.
Key words: proteinase K; gene cloning; *Komagataella phaffii*; gene expression; enzymatic activity.

For citation: Beklemishev A.B., Pykhtina M.B., Kulikov Ya.M., Goryachkovskaya T.N., Bochkov D.V., Sergeeva S.V., Vasileva A.R., Romanov V.P., Novikova D.S., Peltek S.E. Creation of a recombinant *Komagataella phaffii* strain, a producer of proteinase K from *Tritirachium album*. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8):882-888. DOI 10.18699/VJ21.102

Получение рекомбинантного штамма *Komagataella phaffii* – продуцента протеиназы К из *Tritirachium album*

А.Б. Беклемишев[1, 2] ✉, М.Б. Пыхтина[1, 2], Я.М. Куликов[1, 2], Т.Н. Горячковская[2], Д.В. Бочков[2], С.В. Сергеева[2], А.Р. Васильева[2], В.П. Романов[2], Д.С. Новикова[3], С.Е. Пельтек[2] ✉

[1] Федеральный исследовательский центр фундаментальной и трансляционной медицины, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Акционерное общество «Эфирное», г. Алексеевка, Белгородская область, Россия
✉ beklem@niibch.ru; peltek@bionet.nsc.ru

Аннотация. Объектами исследования являлись рекомбинантные штаммы *Komagataella phaffii* K51, несущие интегрированный в их геном гетерологичный ген протеиназы К (PK-w) из *Tritirachium album*, а также препарат рекомбинантной протеиназы К, полученный из этих штаммов. Целью работы было изучение возможности получения рекомбинантных штаммов *K. phaffii* K51, обеспечивающих высокий уровень синтеза функционально активной протеиназы К из *T. album*, и анализ ферментативной активности полученного рекомбинантного энзима. В работе использованы методы компьютерного анализа первичной структуры гена протеиназы К, молекулярно-биологические методы (ПЦР, электрофорез ДНК в агарозных гелях, электрофорез белков в SDS-ПААГ в денатурирующих условиях, спектрофотометрия, методы количественного определения активности протеаз), генно-инженерные методы (методы клонирования и селекции генов в бактериальных клетках *Escherichia*

coli str. TOP10 и в метилотрофных дрожжах *K. phaffii* str. K51). Спроектирован ген природной протеиназы К (*PK-w*), оптимизированный для экспрессии в дрожжах *K. phaffii* K51. Осуществлены синтез и клонирование синтезированного гена протеиназы К в составе вектора pPICZα-A в клетках *E. coli* str. TOP10. Ген протеиназы К встроен в векторную плазмиду pPICZα-A таким образом, чтобы на последующем этапе переклонирования в клетках дрожжей обеспечить его эффективную экспрессию под контролем промотора и терминатора гена *AOX1*, а продукт экспрессии клонированного гена содержал сигнальный пептид альфа-фактора *Saccharomyces cerevisiae* для обеспечения секреции белка в культуральную жидкость. Проведено переклонирование рекомбинантной плазмиды (pPICZα-A/PK-w) в клетках дрожжей *K. phaffii* str. K51. Получен рекомбинантный штамм *K. phaffii* K51, несущий синтетический ген протеиназы К и обеспечивающий его экспрессию в дрожжах и секрецию в культуральную среду. Приблизительный выход рекомбинантной протеиназы К после четырех суток культивирования дрожжевых рекомбинантных клонов составил 25 мкг/мл. Полученный препарат рекомбинантной протеазы обладает высокой удельной протеолитической активностью, составляющей ~5000 Ед/мг.
Ключевые слова: протеиназа К; клонирование гена; *Komagataella phaffii*; экспрессия гена; активность фермента.

Introduction

More than 70 % of enzymes used in various fields of industry are hydrolases (Kudryavtseva et al., 2008; Yin et al., 2014), and proteases account for more than 30 % of the total market of industrial enzymes (Kulkarni et al., 1999; Gupta et al., 2002; Koga et al., 2014; Singh et al., 2016). This is due to their widespread use in various areas of industry, in particular in the production of detergents, in waste disposal, and in the food, dairy, leather, pharmaceutical, and textile industries. The increased demand for the production of proteases in recent years is caused by the urgent need to manufacture high-quality effective detergents as well as new food products from agricultural waste of plant raw materials and from waste of meat and fish processing. In the past three decades, proteinases from various sources (bacteria, bacilli and fungi) have found many applications in various fields of industry and in clinical practice. The most studied group of proteolytic enzymes is bacterial serine proteinases. They are actively used in the pharmaceutical industry, in tissue engineering, and systemic enzyme therapies (Gupta et al., 2002; Kudryavtseva et al., 2008; Yin et al., 2014).

Formulations containing proteinases are widely employed in many fields of medicine: in surgery – for the treatment of trophic ulcers, abscesses, phlegmons, osteomyelitis, and other purulent-inflammatory processes; in dentistry – for the treatment of caries, pulpitis, periodontitis, periodontal disease, and its complications, and in pulmonology – as a mucolytic drug for the treatment of various types of pneumonia and bronchitis (administration via inhalation).

Numerous independent studies confirm that serine proteases hold promise for medical purposes (Yariswamy et al., 2013; Muthu et al., 2017; Belov et al., 2018; Abaturov, 2020; Osmolovsky et al., 2020). In particular, one of the main problems that physicians face when treating skin wounds and burns in people with compromised immunity is the formation of a surface biofilm generated by conditionally pathogenic microorganisms (*Staphylococcus aureus* and *S. epidermidis*, micrococci, and *Pseudomonas*); under this biofilm, the microbes cannot be reached by antibiotics, and as a consequence, wound healing slows down.

For the degradation of various components of the biofilm extracellular matrix, various formulations are currently being designed based on a mixture of enzymes: proteases (including proteinase K), glycosidases, and deoxyribonucleases (Abaturov, 2020). One of the promising areas for application of proteinases is the creation of thrombolytic drugs on the basis of these enzymes. Thus, the development of new effective therapeutics based on enzymes of bacterial origin has good potential for modern medicine, microbiology, and biotechnology.

Thermostable proteases are the most popular in this regard because, firstly, they are characterized by a higher rate of catalysis, and secondly, they provide protection of the reaction mixture and products of enzymatic conversion from microbial contamination because these enzymes catalyze the reactions at high temperatures. Both bacterial and yeast strains that are recombinant superproducers of thermostable proteases have been constructed, and in most studies, it has been shown that the methylotrophic yeast *Komagataella phaffii* generates a larger amount of recombinant proteases than bacterial strains do (Kim et al., 2005; Latiffi et al., 2013; Yu et al., 2014; Ma et al., 2016; Shu et al., 2016; Kangwa et al., 2018; Pereira et al., 2020). In addition, the proteases produced by yeasts are usually secreted into the culture medium in a soluble functionally active state (Yang et al., 2016). Of particular interest are proteinases that exert their activity in a wide range of temperatures and pH of the medium.

Accordingly, a solution to the problem of obtaining a yeast superproducer of a proteinase from *Tritirachium album* (proteinase K) is undoubtedly intriguing because this proteinase has a number of important practical advantages: it has broad specificity, is most active at high reaction temperatures (37 to 60 °C), is functional across a wide pH range (4–12), and is not inhibited by ionic or nonionic detergents. The present study is aimed at solving this problem.

Materials and methods

Materials. All chemical reagents of analytical purity were purchased from Sigma-Aldrich (USA) or Reachem (Moscow, Russia), and restriction endonucleases from SibEnzyme (Novosibirsk, Russia). DNA ligase T4 and DNA polymerase Phusion were acquired from Thermo Fisher Scientific Inc. (USA), and oligonucleotides – Biosintez (Novosibirsk, Russia). Yeast extract, bactopeptone, and tryptone from Difco were utilized to prepare the Luria-Bertani (LB) medium for growing *Escherichia coli* cells. Yeast culture media (YPD, BMGY, BMM2, and BMM10) were prepared as described in the manufacturer's protocol (Easy Select™ Pichia Expression Kit (Invitrogen, USA). Modified Eagle medium (MEM) was bought from Biolot (Russia), dithiothreitol and iodoacetamide,

from Bio-Rad (USA), and porcine trypsin, from Promega (Trypsin Gold, Mass Spectrometry Grade, USA). DEAE-Sepharose FF and SP Sepharose FF ion exchange resins were purchased from GE Healthcare Bioscience (Sweden). The water used in the work was deionized and autoclaved.

Bacterial and yeast strains and plasmid vectors. The yeast *K. phaffii* K51 strain was obtained from the Russian National Collection of Industrial Microorganisms (cat. No. Y-4935), whereas *E. coli* TOP10 and the pPICZα-A plasmid vector were acquired from Invitrogen Inc. (USA).

Buffers and culture media. Solutions and buffers were prepared from deionized autoclaved water. *E. coli* clones carrying the pPICZα-A plasmid or its derivatives were selected on low-salt LB agar plates (1 % of tryptone, 0.5 % of yeast extract, 0.5 % of NaCl, 1.8 % of Bacto-agar, and 50 μg/mL zeocin). Yeast cells were grown in the YPD medium (2 % of yeast peptone, 1 % of yeast extract, and 2 % of dextrose). Yeast transformants were cultured and selected on YPD agar plates (2 %) with various concentrations of zeocin (500 or 2000 μg/mL). The selected yeast clones were also cultivated in the BMGY medium (1 % of yeast extract, 2 % of peptone, 100 mM potassium phosphate pH 6.0, 1.34 % of YNB, 4×10^{-5} % of biotin, and 2 % of glycerol). To induce the *AOX1* gene promoter, the clones were cultivated first in the BMM2 medium (1.34 % of YNB, 4×10^{-5} % of biotin, and 1 % of methanol) and then in the BMM10 medium (1.34 % of YNB, 4×10^{-5} % of biotin, and 5 % of methanol).

Construction of recombinant plasmid (pPICZα-A/PK-w). The nucleotide sequence of a synthetic gene encoding natural proteinase K from *T. album* (i. e., protease K, endopeptidase K; E.C.3.4.21.64) – hereinafter referred to as PK-w – was designed and optimized for expression in the yeast *K. phaffii*. The optimized PK-w protease gene was synthesized by GenScript (USA). The PK-w protease gene was cloned in *E. coli* cells after insertion into the pPICZα-A plasmid at the *Xho*I and *Xba*I sites.

Small-scale preparation of recombinant proteinase K (PK-w). The genetically modified yeast strain was grown in 250 mL of the BMGY medium with 1 % of glycerol in 1 L flasks on an orbital shaker at 250 rpm for 48 h at 28 °C. Next, protein biosynthesis was induced with 1 % methanol (every day, 25 mL of 10 % methanol was added) for 4 days. On the 4th day, proteolytic activity in the culture liquid was determined.

The protein concentration in solutions was determined by three methods: a) via absorption measurement in the protein solution at 280 nm, taking into account the extinction coefficient for the protein; b) by densitometry of the colored protein band in a gel; c) by the Bradford assay with the Quick Start™ Bradford 1×Dye Reagent (Bio-Rad) according to the manufacturer's instructions.

Determination of protease activity of the recombinant proteinase K toward casein. Proteolytic activity was determined by the Kunitz method using casein from cow's milk (Sigma-Aldrich) as a substrate (Bisswanger, 2010). For this purpose, 0.4 mL of a 2 mg/mL casein solution in 10 mM Tris-HCl buffer (pH 8.0) containing 10 mM $CaCl_2$ was heated to 55 °C, and 0.2 mL of the enzyme solution in the same buffer was added. The mixture was incubated at 55 °C for 10 min, and then the reaction was stopped by the addition of 1 mL of 1.2 M trichloroacetic acid. A control solution was subjected to the same procedures, except that the enzyme solution was introduced into the casein solution after the addition of trichloroacetic acid. The samples were centrifuged at 10000 g for 5 min at 5 °C, and supernatant absorbance was determined at a wavelength of 275 nm. One unit of activity was defined as the amount of the protease that leads – in 1 min at 55 °C – to the same absorption value as does 1 μmol of tyrosine (according to a calibration curve). The calibration curve was built within the appropriate range of tyrosine concentrations.

Quantitation of protease activity in the culture liquid toward azocasein. A culture liquid was centrifuged at 4 °C for 10 min at 10000 g to pellet the cells, and the supernatant was collected for the protease activity assay.

The reaction mixture consisting of 0.5 mL of a 0.2 % solution of azocasein in 50 mM Tris-glycine buffer (pH 8) and 0.25 mL of the supernatant was heated in a water bath at 55 °C for 40 min. The reaction was stopped by the addition of 1 mL of 1.2 M trichloroacetic acid. A control solution containing 0.5 mL of a 0.2 % solution of azocasein in 50 mM Tris-glycine buffer (pH 8) without the supernatant was heated in a water bath at 55 °C for 40 min. The reaction was stopped by the addition of 1 mL of 1.2 M trichloroacetic acid, after which 0.25 mL of the supernatant was introduced into the mixture. The product of the azocasein hydrolysis was quantified spectrophotometrically by means of absorption at 440 nm.

Results

Design of the plasmid for the expression of the proteinase K gene in the yeast *K. phaffii* K51

To obtain a yeast producer of proteinase K (i. e., protease K, endopeptidase K; E.C. 3.4.21.64), the proteinase gene from *T. album* was optimized for expression in the methylotrophic yeast *K. phaffii* K51. The amino acid sequence of the proteinase K precursor protein and the layout of its domains are shown in Fig. 1.

The synthesized gene encoded only the prepeptide and amino acid sequence of the mature protein. The signal peptide of the *Saccharomyces cerevisiae* α-factor encoded by a fragment of the pPICZα-A vector (data not shown) served as a signal for enzyme secretion.

Cloning of the recombinant plasmid pPICZα-A/PK-w in *E. coli* TOP10 cells

Electrocompetent *E. coli* TOP10 cells were transformed with recombinant plasmids (pPICZα-A/PK-w) using an electroporator (Bio-Rad). The transformed cells were inoculated into 1 mL of the LB medium and incubated at 37 °C for 1 h on an orbital shaker at 140 rpm. The cell suspensions were plated on agar Petri dishes containing 50 μg/mL zeocin and incubated for 16 h at 37 °C. From 150 to 200 colonies emerged in each plate. Ten colonies from each plate were transferred (with puncturing) to separate Petri dishes containing the agar low-salt LB medium with 50 μg/mL zeocin and were utilized to prepare thermolysates for colony PCR detecting the recombinant plasmid (pPICZα-A/PK-w). The PCR was carried out with a pair of primers specific for the regions of the pPICZα-A vector flanking the inserted gene: forward primer No. 324-AOX1-F, 5′-GACTGGTTCCAATTGACAAGC-3′; and reverse primer No. 325-AOX1-R, 5′-GCAAATGGCATTCTGACATCC-3′.

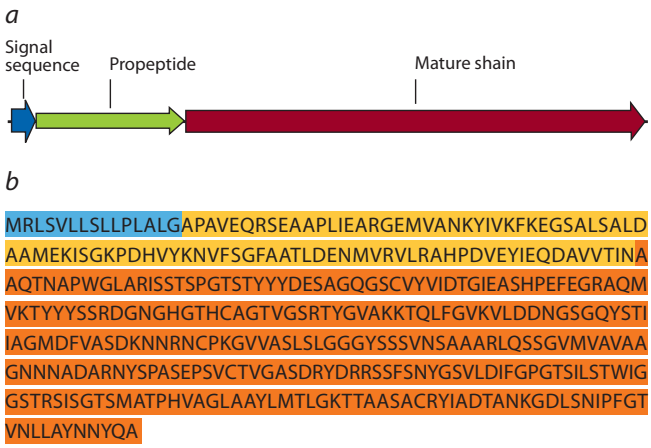


Fig. 1. Domain structure of proteinase K (*a*) and amino acid sequence of the PK-w precursor protein (*b*).

The signal peptide is highlighted in blue, the prepeptide (prodomain) is highlighted in yellow, and the mature protein is orange.

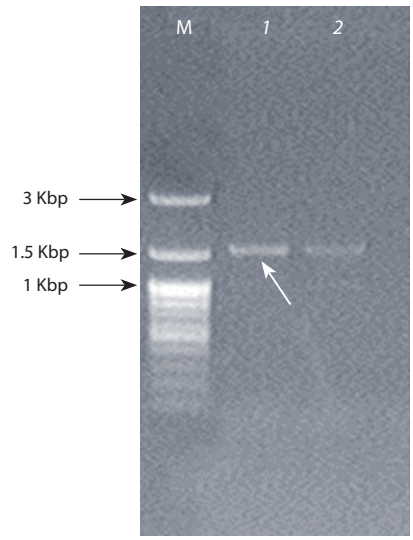


Fig. 2. Electropherogram of the PCR products obtained via amplification of a portion of the vector containing the proteinase K gene (designated as the pPICZα-A/PK-w plasmid).

The arrow indicates the DNA fragment used for subsequent work. Lanes: M – SibEnzyme DNA molecular weight markers (100–3000 bp); *1* and *2* – proteinase K gene amplicons from two recombinant clones carrying plasmids with the *PK-w* gene insert.

Amplicon sizes were analyzed by electrophoresis in a 0.8 % agarose gel containing ethidium bromide. PCR-positive clones were selected for subsequent lab scale production of recombinant plasmids with the aim of their subsequent transfection into yeast cells.

Fig. 2 shows the results of testing *E. coli* TOP10 clones by colony PCR for the presence of the pPICZα-A plasmid carrying an insert of the proteinase K gene (named as the pPICZα-A/PK-w plasmid).

As readers can see in Fig. 2, the amplicons obtained from two recombinant plasmids containing the inserted proteinase K gene have the theoretically expected size: ~1626 bp. The selected clone containing the pPICZα-A/PK-w plasmid with the insertion of the proteinase K gene (*PK-w*) was used for further procedures: lab scale production of the plasmid, its linearization with restriction endonuclease *Bst*X1, and subsequent transfection into *K. phaffii* K51 yeast cells.

Transfection of the proteinase K gene into *K. phaffii* K-51 yeast cells and screening of transfectants

At the first stage, the selected *E. coli* clone was propagated in 100 mL of the LB medium, followed by isolation of pPICZα-A/PK-w plasmid DNA from the cells by means of the GenElute™ HP Plasmid Midiprep Kit. The isolated plasmid DNA was analyzed by electrophoresis in a 0.8 % agarose gel stained with ethidium bromide.

DNA concentration in the plasmid sample was determined on a Qubit fluorometer (Invitrogen). As a result of the isolation procedure, ~25 μg of the purified pPICZα-A/PK-w plasmid was obtained. Approximately 5–10 μg of the isolated pPICZα-A/PK-w plasmid was linearized by digestion with restriction endonuclease *Bst*X1 and used for transfection (electroporation of *K. phaffii* K51, see below). At the end of the restriction reaction, phenol-chloroform extraction of DNA was carried out, followed by its precipitation with isopropanol and washing with 70 % ethanol. DNA pellets were dissolved in 10 μL of double-distilled H_2O, frozen, and stored at –20 °C. The completeness of the plasmid hydrolysis reaction was verified by electrophoresis of restriction products in a 0.8 % agarose gel stained with ethidium bromide. Judging by the results of the electrophoretic analysis, the bulk of the plasmid sample was successfully hydrolyzed by restriction endonuclease *Bst*XI. For the transformation of electrocompetent *K. phaffii* K51 cells, 13–15 μg of plasmid DNA was employed, dissolved in 10 μL of double-distilled H_2O. Electroporation was performed using a Gene Pulser Xcell Total System Electroporator (Bio-Rad).

After preliminary cultivation on an orbital shaker at 200 rpm for 2 h at 27 °C in test tubes containing 1 mL of the YPD medium, the transformed cells were plated on Petri dishes with the agar YPD medium containing 500 or 2000 μg/mL zeocin. The culture plates were placed in a thermostat at 30 °C for 3–5-day incubation. On the 4th day after the cell transfection with the pPICZα-A/PK-w plasmid, many separate colonies were visible on the culture plates with 500 μg/mL of zeocin, whereas on the plates with 2000 μg/mL zeocin, there were 50 to 100 colonies. Such a large number of colonies on the plates with 2000 μg/mL zeocin is apparently due to partial degradation of the antibiotic, which had been stored at 4 °C for a long time.

Zeocin-resistant transformants grown on the culture plates with 2000 μg/mL zeocin were evaluated for their ability to synthesize and secrete the desired protein: we cultured the selected clones in 96-well deep well plates (Axygen Scientific). For screening, 20 colonies were randomly chosen, which were placed into the wells of these plates. In parallel, the same colonies, grown on agar plates with a zeocin concentration of 2000 μg/mL, were transferred to separate Petri dishes (containing an agar medium with the same zeocin concentration) by puncturing sites labeled with numbers.

The chosen clones were cultivated individually in 96-well deep well plates in 300 μL of the BMGY medium on an orbital shaker at 250 rpm for 48 h at 28 °C. Then, 250 μL of

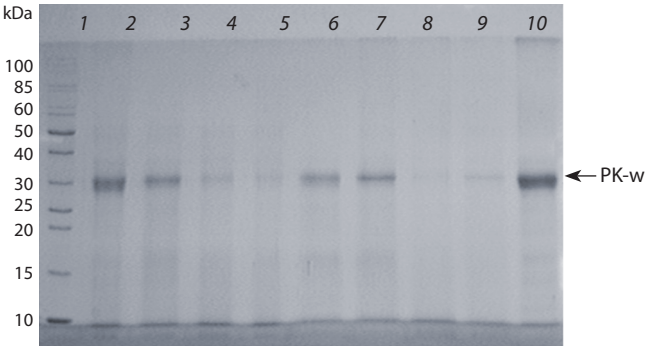


Fig. 3. An electropherogram of the proteins produced by the tested yeast clones transformed with the pPICZα-A/PK-w plasmid.

Lanes: *1* – Thermo Scientific Molecular Weight Markers (10–200 kDa); *2–10* – proteins produced and secreted by the analyzed clones.

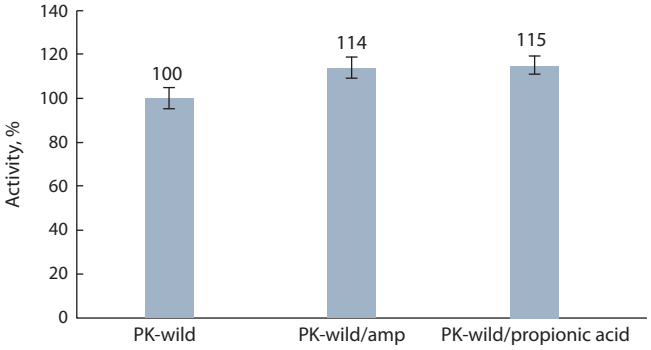


Fig. 4. Proteolytic activity of the culture liquid of recombinant yeast clone No. 10 as a result of its small-scale propagation in the presence of ampicillin (PK-wild/amp) or propionic acid (PK-wild/propionic acid) or without any antibacterial agents (PK-wild). The proteolytic activity of the culture liquid obtained without the antibacterial agents (PK-wild) was set to 100 %.

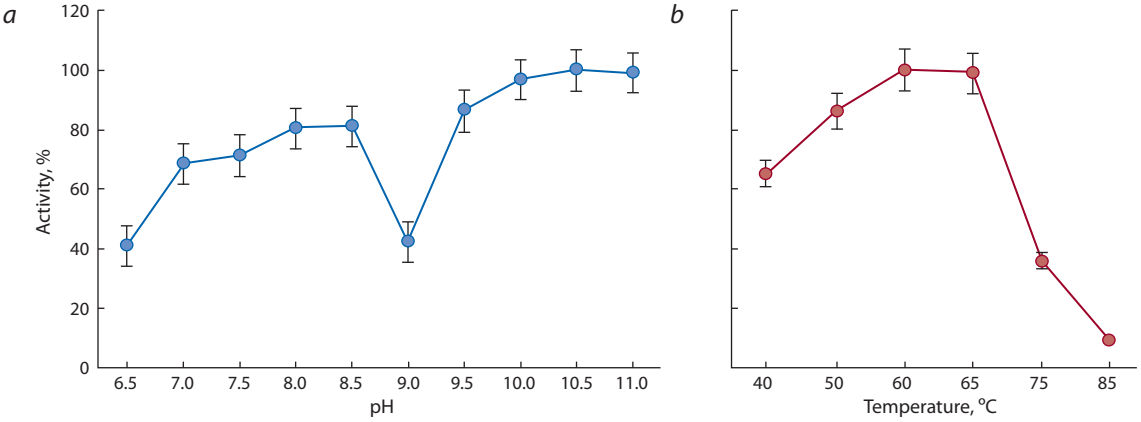


Fig. 5. Dependence of the recombinant proteinase K's activity on pH and temperature.

the BMM2 medium was added into each well. On each of the next 3 days, 50 μL of the BMM10 medium was introduced into the wells. On day 4, the culture liquid from each well was centrifuged at 6000 rpm for 5 min to pellet the cells, and the resulting supernatants were analyzed by SDS-PAGE for the presence of the target protein.

Samples for the electrophoresis were prepared as follows: 10 trichloroacetic acid was added to the supernatants to concentrate the proteins. The protein precipitates were washed with acetone, resuspended in 1× TGB, and 4× denaturing buffer was added, followed by boiling of the samples, and then the proteins were separated in a 12.5 % polyacrylamide gel. According to the electrophoresis results, we chose culture liquids of the clones producing the largest amount of proteins with a molecular weight corresponding to that of natural proteinase K (~30 kDa). The purpose was to determine the concentration of recombinant proteinase K and assess its enzymatic activity. The results of electrophoretic analysis of proteins produced and secreted by the recombinant yeast clones are depicted in Fig. 3.

As presented in Fig. 4, all zeocin-selected clones produce a major protein with a molecular weight of ~29–30 kDa, which corresponds to the molecular weight of mature proteinase K: 28903 Da. According to the electrophoresis results, clone No. 10 was chosen for further experiments because it produced the largest amount of a recombinant protein with a molecular weight of ~29–30 kDa. The yield of recombinant protease PK-w after 4 days of cultivation of recombinant yeast clone No. 10 in the 96-well plate was 25 μg/mL.

Propionic acid was applied to control bacterial contamination during lab scale production of large amounts of the recombinant protein. Fig. 4 shows that the use of propionic acid at a concentration of 0.025 % is comparable to the addition of ampicillin at 0.2 mg/mL.

Microscopic analysis of the cultures did not reveal the presence of bacteria; however, the use of ampicillin and propionic acid somewhat increased the production of the target protein.

Recombinant yeast clone No. 10 was utilized for preliminary small-scale preparation of the enzyme.

Quantification of the protease activity of the recombinant protein produced by clone No. 10

As a result of these measurements, it was found that the obtained batch of the recombinant protein with a molecular weight of ~29–30 kDa had a high specific proteolytic activity, ~5000 U/mg. This finding indicated that this recombinant protein is proteinase K (PK-w).

The dependence of the obtained recombinant proteinase K's activity on pH of the medium and on reaction temperature was investigated next (Fig. 5). The optimum of enzymatic activity

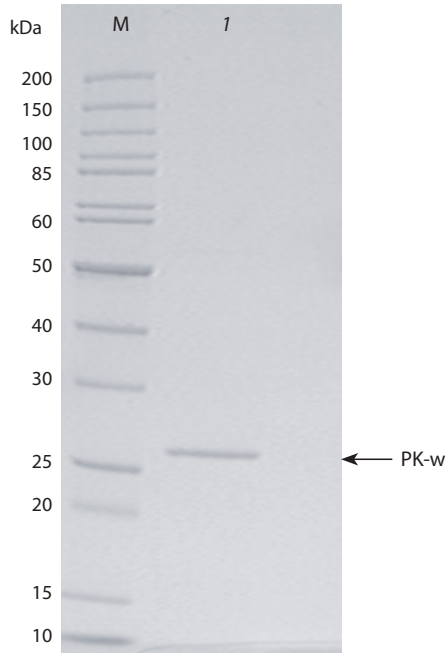


Fig. 6. An electropherogram of the purified recombinant protein from the batch produced by yeast clone No. 10.

Electrophoresis was performed in an SDS 12.5 % polyacrylamide gel. Lanes: M – Thermo Scientific Molecular Weight Markers (10–200 kDa); *1* – the chromatographically purified recombinant protein.

of the obtained recombinant proteinase K is in the range of pH 10–11, although the enzyme is also active at pH from 9.5 to 6.5 (see Fig. 5, *a*). A sharp drop of the activity was observed at pH 9.0 and at pH ≤ 4.0. The optimal temperature range for the manifestation of this protease activity turned out to be 40–65 °C (see Fig. 5, *b*).

Chromatographic purification of the obtained recombinant proteinase K

All procedures were conducted at a temperature ≤ 5 °C. The culture fluid of recombinant clone No. 10 was separated from the cells by centrifugation for 25 min at 4000 rpm. Low-molecular-weight impurities were removed and the supernatant was concentrated 20-fold via ultrafiltration by means of centrifugal concentrators.

Protein impurities were removed from the enzyme sample by ion exchange chromatography on anion exchange resin DEAE-Sepharose 6HF. Elution was performed with a buffer composed of 50 mM sodium chloride and 50 mM Tris-HCl (pH 7.2). Fractions showing proteolytic activity were pooled, concentrated using the centrifuge concentrators, and either lyophilized or stored in 50 % glycerin in a freezer of a fridge.

The purified recombinant protein was analyzed by electrophoresis in SDS 12.5 % polyacrylamide gel (Fig. 6). As presented in Fig. 6, on the gel, the recombinant protein is represented by one major band with a size in the range of ~26.5–27.0 kDa. There are no protein impurities on the gel. The activity of the purified protein was 49 800 U/mg toward azocasein and 5000 U/mg toward casein. Thus, a highly purified batch of recombinant proteinase K was obtained.

Conclusion

The design and optimization of the nucleotide sequence encoding the precursor protein of natural proteinase K (PK-w) from *T. album* were performed to ensure its efficient expression in the yeast *K. phaffii*. The synthesized proteinase K gene was cloned within the pPICZα-A vector in *E. coli* str. TOP10 cells, and then the plasmid was isolated and transfected into yeast *K. phaffii* str. K51 cells.

A recombinant clone of *K. phaffii* K51 that carries the gene of recombinant proteinase K and successfully expresses and secretes this enzyme into the culture medium was obtained. A lab scale batch of the recombinant proteinase K was prepared. Protease activity of the obtained recombinant proteinase K (PK-w) was determined with casein and azocasein as substrates. The enzyme batch has a high specific proteolytic activity: ~5000 U/mg. The optimal enzymatic activity of the obtained recombinant proteinase K is in a pH range of 10–11 and a temperature range of 40–65 °C.

References

Abaturov A.E. Proteases that degrade the biofilm matrix. *Zdorov'ye Rebenka = Child's Health.* 2020;15(3):187-194. DOI 10.22141/2224-0551.15.3.2020.204554. (in Russian)

Belov A.A., Vaniushenkova A.A., Dosadina E.E., Khanafina A.A. New textile dressings based on biodegradable polymers containing proteinases for wounds and burns treatment. *Wounds and Wound Infections. The prof. B.M. Kostyuchenok Journal.* 2018;5(1):16-26. DOI 10.25199/2408-9613-2018-5-1-16-26. (in Russian)

Bisswanger H. Practical Enzymology. Moscow: Binom Publ., 2010; 154-156. (in Russian)

Gupta R., Beg Q., Lorenz P. Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biotechnol.* 2002;59(1):15-32. DOI 10.1007/s00253-002-0975-y.

Kangwa M., Salgado J.A.G., Fernandez-Lahore H.M. Identification and characterization of N-glycosylation site on a *Mucor circinelloides* aspartic protease expressed in *Pichia pastoris*: effect on secretion, activity and thermo-stability. *AMB Express.* 2018;8(1):157. DOI 10.1186/s13568-018-0691-3.

Kim T., Lei X.G. Expression and characterization of a thermostable serine protease (TfpA) from *Thermomonospora fusca* YX in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 2005;68(3):355-359. DOI 10.1007/s00253-005-1911-8.

Koga Y., Tanaka S.-I., Sakudo A., Tobiume M., Aranishi M., Hirata A., Takano K., Ikuta K., Kanaya S. Proteolysis of abnormal prion protein with a thermostable protease from *Thermococcus kodakarensis* KOD1. *Appl. Microbiol. Biotechnol.* 2014;98(5):2113-2120. DOI 10.1007/s00253-013-5091-7.

Kudryavtseva O., Dunaevsky Y., Kamzolkina O., Belozersky M. Fungal proteolytic enzymes: features of the extracellular proteases of xylotrophic basidiomycetes. *Microbiology*. 2008;77(6):643-653. DOI 10.1134/S0026261708060015.

Kulkarni N., Shendye A., Rao M. Molecular and biotechnological aspects of xylanases. *FEMS Microbiol. Rev.* 1999;23(4):411-456. DOI 10.1111/j.1574-6976.1999.tb00407.x.

Latiffi A.A., Salleh A.B., Rahman R.N., Oslan S.N., Basri M. Secretory expression of thermostable alkaline protease from *Bacillus stearothermophilus* FI by using native signal peptide and α-factor secretion signal in *Pichia pastoris*. *Genes Genet. Syst.* 2013;88(2):85-91. DOI 10.1266/ggs.88.85.

Ma X., Liu Y., Li Q., Liu L., Yi L., Ma L., Zhai C. Expression, purification and identification of a thermolysin-like protease, neutral protease I, from *Aspergillus oryzae* with the *Pichia pastoris* expression system. *Protein Expr. Purif.* 2016;128:52-59. DOI 10.1016/j.pep.2016.08.008.

Muthu S., Gopal V.B., Soundararajan S., Nattarayan K., Narayan K.S., Lakshmikanthan M., Malairaj S., Perumal P. Antibacterial serine

protease from *Wrightia tinctoria*: Purification and characterization. *Plant Physiol. Biochem.* 2017;112:161-172. DOI 10.1016/j.plaphy.2017.01.003.

Osmolovsky A.A., Orekhova A.V., Konti E., Kreyer V.G., Baranova N.A., Egorov N.S. Production and stability of the proteinase complex from *Aspergillus ochraceus* L-1 with fibrinolytic and anticoagulant activity. *Moscow University Biological Sciences Bulletin.* 2020;75(3):130-135.

Pereira W.E.S., da Silva R.R., de Amo G.S., Ruller R., Kishi L.T., Boscolo M., Gomes E., da Silva R. A collagenolytic aspartic protease from *Thermomucor indicae-seudaticae* expressed in *Escherichia coli* and *Pichia pastoris*. *Appl. Biochem. Biotechnol.* 2020;191(3):1258-1270. DOI 10.1007/s12010-020-03292-z.

Shu M., Shen W., Yang S., Wang X., Wang F., Wang Y., Ma L. High-level expression and characterization of a novel serine protease in *Pichia pastoris* by multi-copy integration. *Enzyme Microb. Technol.* 2016;92:56-66. DOI 10.1016/j.enzmictec.2016.06.007.

Singh R., Kumar M., Mittal A., Mehta P.K. Microbial enzymes: industrial progress in 21st century. *3 Biotech.* 2016;6(2):174. DOI 10.1007/s13205-016-0485-8.

Yang H., Zhai C., Yu X., Li Z., Tang W., Liu Y., Ma X., Zhong X., Li G., Ma L. High-level expression of proteinase K from *Tritirachium album* limber in *Pichia pastoris* using multi-copy expression strains. *Protein Expr. Purif.* 2016;122:38-44. DOI 10.1016/j.pep.2016.02.006.

Yariswamy M., Shivaprasad H.V., Joshi V., Nanjaraj Urs A.N., Nataraju A., Vishwanath B.S. Topical application of serine proteases from *Wrightia tinctoria* R. Br. (Apocyanaceae) latex augments healing of experimentally induced excision wound in mice. *J. Ethnopharmacol.* 2013;149(1):377-383. DOI 10.1016/j.jep.2013.06.056.

Yin C., Zheng L., Chen L., Tan Q., Shang X., Ma A. Cloning, expression, and characterization of a milk-clotting aspartic protease gene (*Po-Asp*) from *Pleurotus ostreatus*. *Appl. Biochem. Biotechnol.* 2014;172(4):2119-2131. DOI 10.1007/s12010-013-0674-4.

Yu X., Wang X., Zhong X., Tang W., Zhai C., Chen W., Ma L. Expression of the thermostable carboxypeptidase Taq gene in *Pichia pastoris* GS115. *Sheng Wu Gong Cheng Xue Bao*. 2014;30(11):1791-1795.

ORCID ID

A.B. Beklemishev orcid.org/0000-0001-9732-1246
M.B. Pykhtina orcid.org/0000-0002-7808-9274
Ya.M. Kulikov orcid.org/0000-0001-6916-6875
T.N. Goryachkovskaya orcid.org/0000-0002-7445-0608
D.V. Bochkov orcid.org/0000-0002-0764-5800
S.E. Peltek orcid.org/0000-0002-7122-2546

Acknowledgements. This work was done with financial support from the Ministry of Science and Higher Education of the Russian Federation as part of the implementation of a comprehensive project on the topic "Creation of a high-tech production workflow for high-quality nutritional proteins of plant origin" (in accordance with the agreement No. 075-11-2020-036 of December 15, 2020, about a subsidy from the federal budget for the development of cooperation between a state scientific institution and an enterprise from a real-world sector of the economy in order to implement a comprehensive project for setting up high-tech production) at the ICG SB RAS within the framework of Decree No. 218 of the Russian Government (April 9, 2010). A part of this work was performed on the equipment of the Multi-Access Center "Proteomic Analysis" at Federal Research Center of Fundamental and Translational Medicine (Novosibirsk, Russia).

The authors are grateful to A.S. Rozanov (senior researcher in the Laboratory of Molecular Biotechnology, ICG SB RAS) for providing the *K. phaffii* K51 yeast strain.

Conflict of interest. The authors declare no conflict of interest.

Received November 16, 2021. Revised November 29, 2021. Accepted November 29, 2021.

Вавиловский журнал генетики и селекции. 2021;25(8):889-895
DOI 10.18699/VJ21.103

CLARITY and Light-Sheet microscopy sample preparation in application to human cerebral organoids

T.A. Shnaider ✉, I.E. Pristyazhnyuk

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ shnayder@bionet.nsc.ru

Abstract. Cerebral organoids are three-dimensional cell-culture systems that represent a unique experimental model reconstructing early events of human neurogenesis *in vitro* in health and various pathologies. The most commonly used approach to studying the morphological parameters of organoids is immunohistochemical analysis; therefore, the three-dimensional cytoarchitecture of organoids, such as neural networks or asymmetric internal organization, is difficult to reconstruct using routine approaches. Immunohistochemical analysis of biological objects is a universal method in biological research. One of the key stages of this method is the production of cryo- or paraffin serial sections of samples, which is a very laborious and time-consuming process. In addition, slices represent only a tiny part of the object under study; three-dimensional reconstruction from the obtained serial images is an extremely complex process and often requires expensive special programs for image processing. Unfortunately, staining and microscopic examination of samples are difficult due to their low permeability and a high level of autofluorescence. Tissue cleaning technologies combined with Light-Sheet microscopy allows these challenges to be overcome. CLARITY is one of the tissue preparation techniques that makes it possible to obtain opaque biological objects transparent while maintaining the integrity of their internal structures. This method is based on a special sample preparation, during which lipids are removed from cells and replaced with hydrogel compounds such as acrylamide, while proteins and nucleic acids remain intact. CLARITY provides researchers with a unique opportunity to study three-dimensional biological structures while preserving their internal organization, including whole animals or embryos, individual organs and artificially grown organoids, in particular cerebral organoids. This protocol summarizes an optimization of CLARITY conditions for human brain organoids and the preparation of Light-Sheet microscopy samples.
Key words: cerebral organoids; CLARITY; Light-Sheet microscopy; immunohistochemistry; tissue clearing; tissue imaging.

For citation: Shnaider T.A., Pristyazhnyuk I.E. CLARITY and Light-Sheet microscopy sample preparation in application to human cerebral organoids. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(8):889-895. DOI 10.18699/VJ21.103

CLARITY и Light-Sheet микроскопия применительно к органоидам головного мозга человека

Т.А. Шнайдер ✉, И.Е. Пристяжнюк

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ shnayder@bionet.nsc.ru

Аннотация. Церебральные органоиды – это трехмерные системы культивирования клеток, представляющие собой уникальную экспериментальную модель, которая позволяет реконструировать ранние события нейрогенеза человека *in vitro* в норме и при различных патологиях. На сегодняшний день для изучения морфологических параметров органоидов чаще всего применяют иммуногистохимический анализ. В связи с этим аспекты трехмерной цитоархитектуры органоидов, такие как нейронные сети или асимметричная внутренняя организация, трудно реконструировать при использовании рутинных подходов. Иммуногистохимический анализ биологических объектов является универсальным методом в биологических исследованиях. Один из ключевых этапов данного подхода – изготовление крио- или парафиновых серийных срезов образцов. Это очень трудоемкий и времязатратный процесс. Кроме того, срезы представляют собой лишь небольшую часть исследуемого объекта, а трехмерная реконструкция из полученных серийных изображений является крайне сложной процедурой и часто требует специальных дорогостоящих программ для обработки изображений. К сожалению, окрашивание и микроскопирование целых образцов затруднено из-за их низкой проницаемости и высокого уровня автофлуоресценции. Технологии очистки тканей в сочетании с Light-Sheet микроскопией дают возможность преодолеть эти проблемы при работе. CLARITY – это одна из технологий подготовки тканей, позволяющая сделать непрозрачные биологические объекты прозрачными с сохранением целостности их внутренней структуры. Метод основан на специальной пробоподготовке, во

время которой из клеток удаляются липиды и заменяются гидрогелевыми соединениями, такими как акриламид; при этом белки и нуклеиновые кислоты остаются интактными. Технология CLARITY предоставляет исследователям уникальную возможность изучать объемные биологические структуры с сохранением их внутренней организации, включая целых животных или эмбрионы, отдельные органы и искусственно выращенные органоиды, в частности церебральные. Данный протокол обобщает оптимизацию условий CLARITY для органоидов головного мозга человека и особенности подготовки образцов Light-Sheet микроскопии.
Ключевые слова: церебральные органоиды; CLARITY; Light-Sheet микроскопия; иммуногистохимия; очистка тканей.

Introduction

Biological tissues and organs present a complex three-dimensional structure. Due to their opacity and high level of autofluorescence, three-dimensional reconstruction of such objects is an extremely laborious, but necessary process. To date, a number of protocols (more than a dozen) have been developed for making tissue transparent: SeeDB (Ke et al., 2013), ScaleA2 (Hama et al., 2011), uDISCO (Pan et al., 2016), CLARITY (Chung, Deisseroth, 2013), CUBIC (Susaki et al., 2015) and others. In general, all protocols can be divided into three groups, depending on the chemicals used for tissue clearance: organic solvents (hydrophobic reagent)-based protocols (BABB, 3DISCO, ECi method), hydrophilic reagent-based protocols (ClearT, Scale, FUnGI, Fructoseglycerol, CUBIC and other) and hydrogel-tissue chemistry-based protocol (CLARITY, SWITCH and SHIELD) (Ueda et al., 2020; Susaki, Takasato, 2021). Some of them have different advantages like quality and speed of clearing or simplicity of the procedure. But on the other hand, some of the protocols involve using toxic and corrosive chemicals that require special objectives to avoid damage to the microscope or require other special equipment. Most of these protocols have been developed to clarify entire organs or their big fragments.

Recently, a new method of artificial mini-organ or organoids generation from induced pluripotent stem cells (iPSC) was developed (Lancaster et al., 2013) and now many different types of organoids have already been produced (brain, lung, liver, intestine, pancreas, kidney and others). Organoids are widely used both to recreate the three-dimensional architecture and functional activity of the original organs during normal embryonic development and at various disorders and to test the biological activity of various drugs, chemical and biological agents. Usually, organoids are opaque, which makes investigating them rather difficult. For this purpose, it is advisable to use the combination of tissue clearing and 3D imaging technologies. However, it is important to select the clarifying technology that would match organoids size and fragility as much as possible and would produce sufficient resolution for investigation of tiny structures.

Various techniques have been used for organoid tissue clearing and several studies have compared different clarifying methods which could be applied to mini-organs (Susaki, Takasato, 2021). Some techniques, such as the hydrophilic clearing protocols (ClearT2 and ScaleS) are most acceptable for clearing small spheroids such as neurospheres (Boutin, Hoffman-Kim, 2015) or cancer cell spheroids (Boutin et al., 2018). Others, such as RapiClear, Fructoseglycerol and FUnGI, also using hydrophilic components, are designed and optimized for handling small and fragile, predominantly hollow organoid structures such as intestinal organoids. It should be noted that these protocols are very convenient and take only three days without application of harmful chemicals (Dekkers et al., 2019; van Ineveld et al., 2020). For complex and dense brain organoids, stronger clearing protocols including delipidation procedure are usually used (Susaki, Takasato, 2021). Applying organic solvent-based methods like 2Eci (2nd generation Ethyl cinnamate-based clearing method) (Masselink et al., 2019; Goranci-Buzhala et al., 2020) or BABB method for midbrain organoids (Renner H. et al., 2020) can get a relatively quick (within a few days) result. However, most of the organic components used in these protocols are quite toxic (for example, a mixture of benzyl alcohol and benzyl benzoate in BABB method (Renner H. et al., 2020)).

The use of hydrogel-tissue chemistry sometimes provides more opportunities for preserving the structure of organoids and increasing the optical resolution of tiny objects. That is due to the tissue hydrogel scaffold preparation by cross-linking hydrogel monomers to native biomolecules (Gradinaru et al., 2018). The creation of such a polymer frame in the brain organoids allows combining these protocols with additional procedures with sodium dodecyl sulfate and physical electrophoresis, as well as with high-resolution imaging of Expansion Microscopy with a general microscopy setup (Wassie et al., 2019; Susaki, Takasato, 2021).

Thus, it is quite important to choose the most optimal and effective tissue clearing technique for samples, especially for such complex objects as cerebral organoids.

One of the most convenient and lab-friendly techniques is CLARITY (Clear Lipid-exchanged Acrylamide-hybridized Rigid Imaging/Immunostaining/*In situ* hybridization-compatible Tissue-hYdrogel). CLARITY was developed in 2013 for obtaining high-resolution information from complex 3D structures, such as the whole mouse brain (Chung, Deisseroth, 2013). Application of this technique enabled to obtain intact-tissue imaging of long-range projections, local circuit wiring, cellular relationships, subcellular structures, protein complexes, and neurotransmitters. CLARITY protocol includes replacing lipids with hydrophilic polymers (acrylamide and bis-acrylamide), which help to stabilize tissue but make it optically transparent and permeable. It is very important that molecules like nucleic acids and proteins stuck in the hydrogel keep their structures and locations. Thus, CLARITY allows combining tissue clearing techniques with immunostaining and *in situ* hybridization and explores the internal structure of large three-dimensional objects without damaging their integrity. There is only one article in which CLARITY technique was used for cerebral organoid clarifying (Sakaguchi et al., 2019), but without a detailed description. Thus, the aim of our work was optimization of CLARITY protocol in application to cerebral organoids and detailed description of samples preparation for Light-Sheet microscopy.

Materials and methods

Reagents

1. Acrylamide (PanReac AppliChem, catalogue number: A1090).
2. Agarose D1, low EEO (Life science products, catalogue number: 1932.0025).
3. Bisacrylamide (PanReac AppliChem, catalogue number: A3636).
4. Boric acid (PanReac AppliChem, catalogue number: A2940).
5. ddH_2O.
6. Glue (Henkel, catalogue number: 2340344).
7. Parafilm M (Pechiney Plastic Packaging Company, catalogue number: PM 996).
8. Paraformaldehyde (PFA) (Sigma Aldrich, catalogue number: 158127).
9. Phosphate buffer saline (PBS) (VWR Life Science AMRESCO, catalogue number: Am-E404-100).
10. Sodium azide (Sigma Aldrich, catalogue number: S8032).
11. Sodium dodecyl sulfate (PanReac AppliChem, catalogue number: A1112).
12. Triton X-100 (VWR Life Science AMRESCO, catalogue number: Am-O694-0.1).
13. VA044 (Wako, catalogue number: 011-19365).
14. Serological pipets 5, 10, 25 ml (Corning, catalogue number: 4050, 4100, 4250).
15. 1-ml syringe (B. Braun, catalogue number: 9161635S).
16. 2 ml tube (Eppendorf, catalogue number: 0030120094).
17. 5 ml tube (Axygen, catalogue number: SCT-5ML-S).
18. Glass bottle 100 and 500 ml (Rasotherm, catalogue number: 95206001 and 95206003).
19. Syringe filter, 0.22 μm (TTP, catalogue number: 99722).
20. 4′,6-diamidino-2-phenylindole (DAPI) (Sigma Aldrich, catalogue number: D-9542).
21. Antibodies (Table 1).

Table 1. Primary and secondary antibodies used in the protocol

Antibodies	Producer	Catalogue number	Host	Dilution
anti-CTIP2	Abcam	ab18465	Rat	1:100
anti-bTubb3	Covance	MMS-435P	Mouse	1:200
Anti-Rat IgG (Alexa Fluor® 488 conjugated)	Jackson ImmunoResearch	712-545-150	Donkey	1:200
Anti-Mouse IgG (Alexa Fluor® 488 conjugated)	Jackson ImmunoResearch	715-545-150	Donkey	1:200

Equipment

1. Light-Sheet Z1 microscope (Zeiss).
2. Orbital shaker (Biosan, catalogue number: OS-20).
3. Roller shaker (Selecta, catalogue number: 7001723).
4. pH meter (OHAUS, catalogue number: 00000032755).
5. Magnetic stirrer (Biosan, catalogue number: MSH-300i).
6. Standard microwave.
7. Thermometer.
8. Forceps.
9. Chemical spoons.
10. Fume hood.
11. Icebox.

Software

1. ImageJ (NIH, https://imagej.nih.gov/ij/index.html)
2. ZEN (Zeiss, https://www.zeiss.com/)

Procedure

Fixation of human cerebral organoid

Note: For any manipulation with organoids, use cut 1 ml tips or wide orifice 1 ml tips to protect samples from damage.

1. Transfer cerebral organoids in 5-ml tubes and wash with 1X PBS solution 2 times.
2. Replace 1X PBS solution with freshly prepared 4 % PFA solution.
3. Place the tubes on the roller/orbital shaker and incubate at room temperature for 2–3 h.
4. Wash samples 3 times with 1X PBS solution for 30 min.

Note: At this step, cerebral organoids can be kept at +4 °C in 1X PBS solution. For keeping more than 1 week, we recommend adding sodium azide to a final concentration of 0.01 % to prevent sample contamination with bacteria and fungi.

Hydrogel embedding

1. Precool all solutions, equipment, and samples on ice to prevent premature polymerization of the hydrogel solution.

Note: If you use a frozen aliquot of hydrogel solutions, thaw the vial on ice in a fridge overnight. After thawing, gently mix and check for the absence of precipitation.

2. Fill the 2-ml tube with the hydrogel solution and transfer cerebral organoids in the tube having previously gently removed leftovers of the PBS with a paper towel.

Note: 2-ml tube format is acceptable for 1–3 organoids. For a large number of organoids, we recommend using a bigger tube.

3. Incubate the samples in the hydrogel solution at +4 °C at the lowest speed of roller/orbital shaker for 24 h.
4. Refill the tube with fresh hydrogel solution and incubate at +37 °C for 4 h.

Note: Fill the tube with hydrogel solution completely. Oxygen inhibits hydrogel polymerization, thus all bubbles should be removed. Additionally, we recommend covering the tube with Parafilm to prevent air access.

5. Very gently extract the samples from the polymerized hydrogel by carefully rolling samples on a paper towel.
6. Transfer the samples into the 5-ml tube and wash with Clearing Solution 4 times at room temperature for 24 h.

Passive clearing

1. Change Clearing Solution every 2 days and incubate at +37 °C with agitation. Continue clearing until samples become transparent.

Note: We strongly recommend using +37 °C for lipid removal. Room temperature slows this process down to several months!
Note: The time of tissue clearing depends on the size of organoids. Cerebral organoids ≤0.5 cm become transparent during ~2 weeks, for organoids ≥0.5 cm it can take up to 3 weeks.

2. Wash samples in PBST for 48 h. Change solution 2–3 times per day.

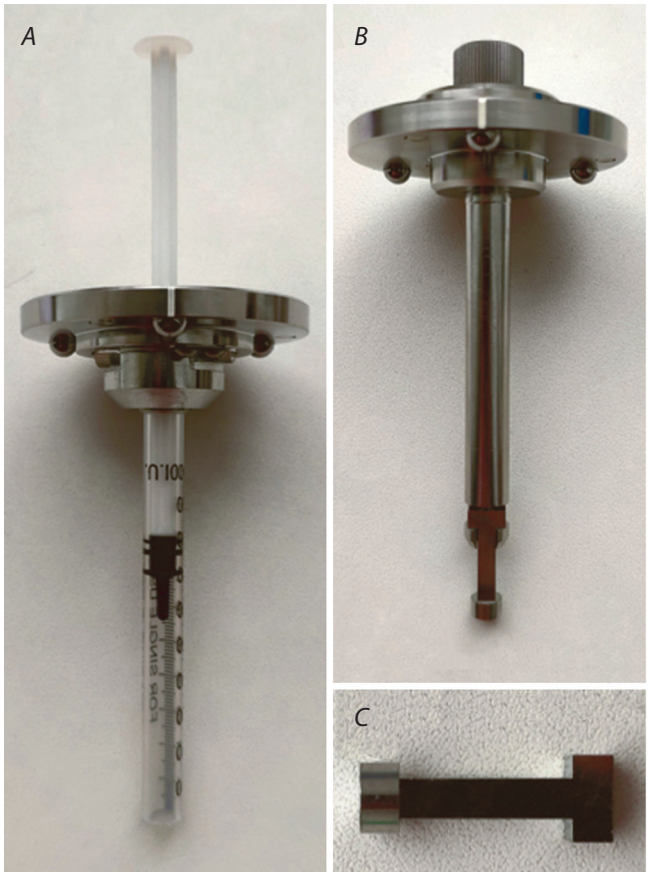


Fig. 1. *A*, ready-to-use sample holder with 1-ml syringe for agarose embedded samples; *B*, sample holder for hanging samples; *C*, the bar for sample sticking.

Staining

1. Incubate samples with primary antibodies in PBST at room temperature on a shaker for 3 days.
2. Wash samples with PBST for 2 days, changing PBST every 4 h.
3. Incubate with secondary antibodies and DAPI at room temperature for 2 days.
4. Wash samples with PBSR for 2 days, changing PBST every 4 h.

Note: Antibody consumption for staining of CLARITY samples is very high. We recommend reducing the volume to the minimum at which the samples in the tube are completely covered with the staining buffer with constant stirring on an orbital or roller shaker.

Note: For larger organoids, we recommend extending each staining step by at least 1 day.

Sample preparation for Light-Sheet microscopy

Organoid sizes can vary greatly. Therefore, we recommend using a different fixation method for Light-Sheet microscopy depending on the size.

Agarose embedding samples (for smaller samples)

Note: Use the agarose with a low melting point temperature only.

Note: The percentage of agarose solutions depends on the size of the organoid. For larger organoids, use 1.5 % agarose solution.

1. Prepare the 1-ml syringe by cutting off the top (Fig. 1, *A*).
2. Weigh the required amount of agarose (at the rate of 1 g per 100 ml) and dissolve in 1X PBS or ddH_2O. Prepare agarose solution by melting in the microwave. Usually, for a 1-ml syringe, 1.5 ml of agarose is enough.
3. Pour the hot agarose solution into a 12-well plate or any other laboratory glassware or plasticware. When agarose solution cools down to +40 °C, transfer samples and gently mix. Put the samples in agarose solution into the 1-ml syringe.
4. Assemble the 1-ml syringe with a sample holder (see Fig. 1, *A*).
5. Proceed to Light-Sheet microscopy.

Note: Fill the microscope chamber with ddH_2O or 1X PBS. No great differences were observed between the two solutions.

Agarose-free or hanging samples (for bigger samples)

1. Glue the sample to the bar (see Fig. 1, *B*, *C*). The area of adhesion can be increased by attaching a small piece of filter paper. Keep samples in ddH_2O or 1X PBS before placing them into the microscope chamber.
2. Proceed to Light-Sheet microscopy.

Note: It is imperative to check and rinse the rod and sample holder for glue residues. If there are any, we strongly recommend that you soak in soapy water and mechanically remove any glue residue.

Recipes

Note: Most solutions and reagents from this protocol are toxic and biohazardous. Do not forget about your safety and work in protective laboratory clothing and only under a fume hood!

10X PBS solution

To prepare a 10X stock solution, dissolve 10 tablets of PBS in 100 ml of ddH_2O.

PFA solutions

- 16 % PFA stock solution

To prepare stock solution, dissolve 16 g of PFA in 80 ml of 1X PBS using a magnetic stirrer. Adjust pH to 7.4–7.5 and add 1X PBS up to 100 ml. Filter the solution through a 0.40 μm filter and aliquote into 5 ml tubes. Keep stock solution at +4 °C for short storage (up to 2 weeks) or at –20 °C for long storage.

- 4 % PFA working solution

To prepare 4 % PFA working solution, dilute stock solution with 1X PBS.

Hydrogel solution

Note: All solutions and equipment have to be pre-cooled to prevent premature polymerization of hydrogel solution.

1. Mix all components on ice according to Table 2.
2. Aliquote hydrogel solution and keep at –20 °C for long storage or use freshly prepared solution.

Table 2. Hydrogel solution composition

Component	Stock	Quantity	Final concentration
Acrylamide	40 %	10 ml	4 %
Bisacrylamide	2 %	1.25 ml	0.025 %
PBS	×10	10 ml	×1
ddH_2O	–	78.5 ml	–
VA-044	–	0.25 g	0.25 %

Clearing solution

1. Mix all components on ice according to Table 3.
2. Keep the solution in a glass bottle at room temperature.

Table 3. Clearing solution composition

Component	Quantity	Final concentration
Sodium dodecyl sulfate	40 g	4 %
Boric acid	12.366 g	200 mM
Sodium hydroxide	–	to pH 8.5
ddH_2O	1000 ml	–

PBST

1. Add Triton-X100 to the final concentration of 0.1 % using a magnetic stirrer.
2. Keep the solution in a glass bottle at room temperature.

Results

Human cerebral organoids were generated according to a protocol from Lancaster et al. (2013) with small modifications. 2- and 3-month-old cerebral organoids were used for tissue clearing protocol. At this stage, there are dense spheres more than 2 mm in diameter (Fig. 2, *A*). We noted that the time of tissue clearing depends on the size of organoids. Cerebral organoids ≤0.5 cm become transparent during ~2 weeks, for organoids ≥0.5 cm up to 3 weeks. This time may vary from sample to sample, however, continue cleaning until the samples become transparent (see Fig. 2, *B*, *C*).

For immunostaining we chose two proteins with different subcellular localisation such as nuclear CTIP2 (Fig. 3, *A*) and cytoplasmic bTubb3 (see Fig. 3, *C*). We did not find a significant difference between penetration of antibodies into different

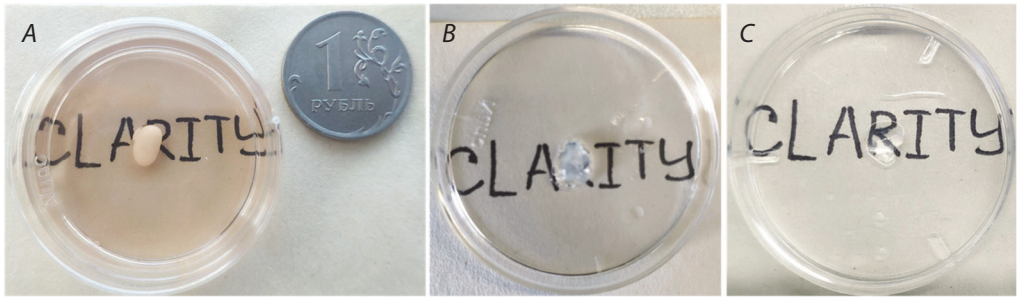


Fig. 2. *A*, intact cerebral organoid before CLARITY; *B*, hydrogel embedded cerebral organoid before tissue clearing; *C*, cerebral organoids after 2 weeks of tissue clearing.

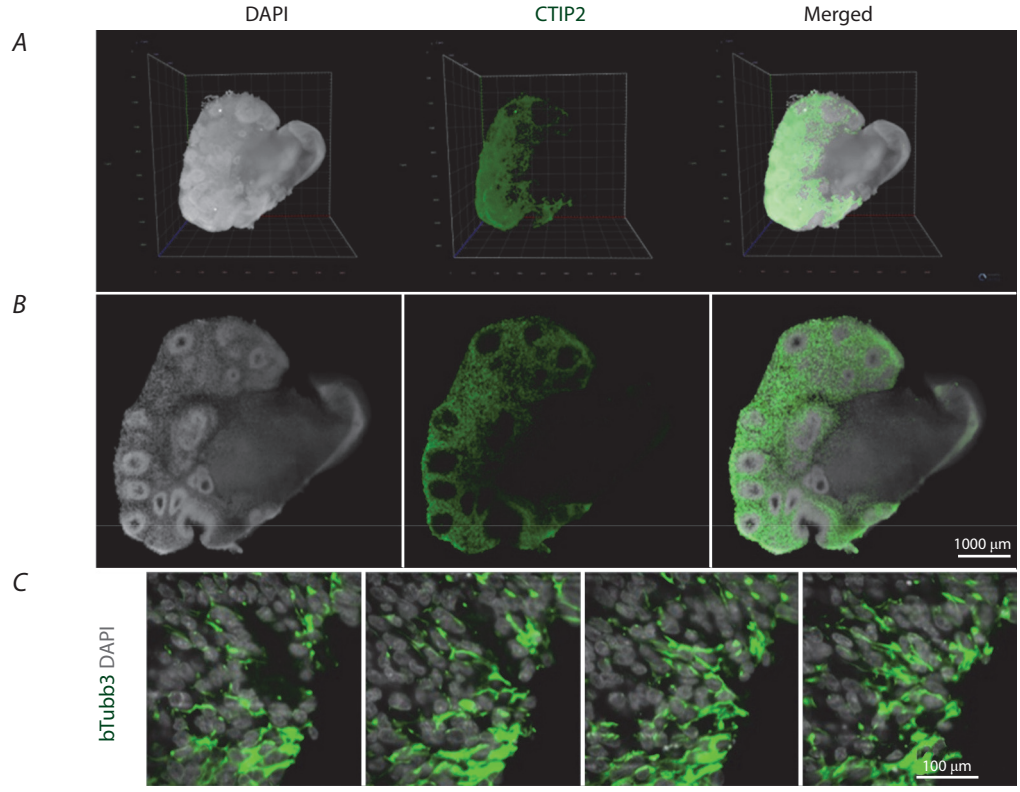


Fig. 3. Light-Sheet imaging of a cerebral organoid after CLARITY: *A*, 3D reconstructed cerebral organoid, 5×, NA 0.16, water immersion; *B*, optical section of the middle part of the cerebral organoid, 5×, NA 0.16, water immersion; *C*, optical sections of a small part of the cerebral organoid, 10×, NA 0.5, water immersion.

cellular compartments. In both cases, we observed specific staining throughout the entire thickness of the organoid (see Fig. 3, *B*).

Conclusions

Cerebral organoids are a unique novel technology that allows the reconstruction of early human neurogenesis. The outstanding feature of this *in vitro* system is the reproduction of the three-dimensional organization of the human embryonic brain. Standard histological methods of analysis do not allow reconstructing the internal structure of the cerebral organoids and result in information loss. The tissue clearing technique helps to overcome these limitations, allowing to recreate a three-dimensional model of cerebral organoids and explore their fine organization without internal structure destruction. This is especially important for the investigation of brain organoids since they contain a dense network of long processes of nerve cells, which is very difficult to study by serial sections (Dodt et al., 2007).

Based on the various tissue clearance techniques analysis, we settled on the use of hydrogel-tissue chemistry as a clearing agent. Generally hydrophobic and hydrophilic reagent-based protocols are applied to the investigation of spheroids or hollow organoids such as intestinal organoids (Susaki, Takasato, 2021), while hydrogel reagents are used for the clarifying of human iPSC-derived retinal organoids (Cora et al., 2019) and iPSC-derived cerebral organoids (Renner M. et al., 2017; Sakaguchi et al., 2019; Albanese et al., 2020). Hydrogel-tissue chemistry-based protocols maximize the preservation of the internal structure of organoids and allow to achieve high optical resolution and low background at fluorescent microscopy.

Currently, there are at least three known hydrogel-tissue chemistry-based methods that use different delipidation and dehydration chemicals: SWITCH (Glutaraldehyde cross-linking (Delipidation) Diatrizoic acid N-methyl-D-glucamine Iodixanol (dehydration)), SHIELD (Polyepoxy cross-linking (Delipidation), Diatrizoic acid N-methyl-D-glucamine Iodixanol (dehydration)) and CLARITY (Hydrogel embedding (Delipidation), HistodenzTM Glycerol (dehydration)) (Susaki, Takasato, 2021; Yu et al., 2021). Therefore in our choice of a suitable technique, we also focused on the availability of the appropriate reagents, the simplicity of the protocol and the lack of need for special equipment.

Of course, a significant disadvantage of CLARITY technique is the relatively long tissue clearance procedure (approximately three weeks for 90-days cerebral organoids), but this obstacle is compensated by a quite simple protocol. To our knowledge, there is a single report in which the CLARITY technique was used for cerebral organoid clarifying (Sakaguchi et al., 2019); however, a detailed description of this technique applied to brain organoids has not been previously performed. For the first time, we make a detailed description of the human cerebral organoid samples preparation for investigation of CLARITY-treated samples for Light-Sheet microscopy.

References

Albanese A., Swaney J.M., Yun D.H., Evans N.B., Antonucci J.M., Velasco S., Sohn C.H., Arlotta P., Gehrke L., Chung K. Multiscale 3D phenotyping of human cerebral organoids. *Sci. Rep.* 2020;10(1):21487. DOI 10.1038/s41598-020-78130-7. https://www.ncbi.nlm.nih.gov/pubmed/33293587.

Boutin M.E., Hoffman-Kim D. Application and assessment of optical clearing methods for imaging of tissue-engineered neural stem cell spheres. *Tissue Eng. Part C Methods.* 2015;21(3):292-302. DOI 10.1089/ten.TEC.2014.0296. https://www.ncbi.nlm.nih.gov/pubmed/25128373.

Boutin M.E., Voss T.C., Titus S.A., Cruz-Gutierrez K., Michael S., Ferrer M. A high-throughput imaging and nuclear segmentation analysis protocol for cleared 3D culture models. *Sci. Rep.* 2018;8(1):11135. DOI 10.1038/s41598-018-29169-0. https://www.ncbi.nlm.nih.gov/pubmed/30042482.

Chung K., Deisseroth K. CLARITY for mapping the nervous system. *Nat. Methods.* 2013;10(6):508-513. DOI 10.1038/nmeth.2481. https://www.ncbi.nlm.nih.gov/pubmed/23722210.

Cora V., Haderspeck J., Antkowiak L., Mattheus U., Neckel P.H., Mack A.F., Bolz S., Ueffing M., Pashkovskaia N., Achberger K., Liebau S.A. Cleared view on retinal organoids. *Cells.* 2019;8(5):391. DOI 10.3390/cells8050391. https://www.ncbi.nlm.nih.gov/pubmed/31035373.

Dekkers J.F., Alieva M., Wellens L.M., Ariese H.C.R., Jamieson P.R., Vonk A.M., Amatngalim G.D., Hu H., Oost K.C., Snippert H.J.G., Beekman J.M., Wehrens E.J., Visvader J.E., Clevers H., Rios A.C. High-resolution 3D imaging of fixed and cleared organoids. *Nat. Protoc.* 2019;14(6):1756-1771. DOI 10.1038/s41596-019-0160-8. https://www.ncbi.nlm.nih.gov/pubmed/31053799.

Dodt H.U., Leischner U., Schierloh A., Jährling N., Mauch C.P., Deininger K., Deussing J.M., Eder M., Zieglgänsberger W., Becker K. Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain. *Nat. Methods.* 2007;4(4):331-336. DOI 10.1038/nmeth1036. https://www.ncbi.nlm.nih.gov/pubmed/17384643.

Goranci-Buzhala G., Mariappan A., Gabriel E., Ramani A., Ricci-Vitiani L., Buccarelli M., D'Alessandris Q.G., Pallini R., Gopalakrishnan J. Rapid and efficient invasion assay of glioblastoma in human brain organoids. *Cell Rep.* 2020;31(10):107738. DOI 10.1016/j.celrep.2020.107738. https://www.ncbi.nlm.nih.gov/pubmed/32521263.

Gradinaru V., Treweek J., Overton K., Deisseroth K. Hydrogel-tissue chemistry: principles and applications. *Annu. Rev. Biophys.* 2018;47:355-376. DOI 10.1146/annurev-biophys-070317-032905. https://www.ncbi.nlm.nih.gov/pubmed/29792820.

Hama H., Kurokawa H., Kawano H., Ando R., Shimogori T., Noda H., Fukami K., Sakaue-Sawano A., Miyawaki A. Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. *Nat. Neurosci.* 2011;14(11):1481-1488. DOI 10.1038/nn.2928. https://www.ncbi.nlm.nih.gov/pubmed/21878933.

Ke M.T., Fujimoto S., Imai T. SeeDB: a simple and morphology-preserving optical clearing agent for neuronal circuit reconstruction. *Nat. Neurosci.* 2013;16(8):1154-1161. DOI 10.1038/nn.3447. https://www.ncbi.nlm.nih.gov/pubmed/23792946.

Lancaster M.A., Renner M., Martin C.A., Wenzel D., Bicknell L.S., Hurles M.E., Homfray T., Penninger J.M., Jackson A.P., Knoblich J.A. Cerebral organoids model human brain development and microcephaly. *Nature.* 2013;501(7467):373-379. DOI 10.1038/nature12517. https://www.ncbi.nlm.nih.gov/pubmed/23995685.

Masselink W., Reumann D., Murawala P., Pasierbek P., Taniguchi Y., Bonnay F., Meixner K., Knoblich J.A., Tanaka E.M. Broad applicability of a streamlined ethyl cinnamate-based clearing procedure. *Development.* 2019;146(3):dev166884. DOI 10.1242/dev.166884. https://www.ncbi.nlm.nih.gov/pubmed/30665888.

Pan C., Cai R., Quacquarelli F.P., Ghasemigharagoz A., Lourbopoulos A., Matryba P., Plesnila N., Dichgans M., Hellal F., Ertürk A. Shrinkage-mediated imaging of entire organs and organisms using uDISCO. *Nat. Methods.* 2016;13(10):859-867. DOI 10.1038/nmeth.3964. https://www.ncbi.nlm.nih.gov/pubmed/27548807.

Renner H., Grabos M., Becker K.J., Kagermeier T.E., Wu J., Otto M., Peischard S., Zeuschner D., TsyTsyura Y., Disse P., Klingauf J., Leidel S.A., Seebohm G., Schöler H.R., Bruder J.M. A fully auto-

mated high-throughput workflow for 3D-based chemical screening in human midbrain organoids. *Elife.* 2020;9:e52904. DOI 10.7554/eLife.52904. https://www.ncbi.nlm.nih.gov/pubmed/33138918.

Renner M., Lancaster M.A., Bian S., Choi H., Ku T., Peer A., Chung K., Knoblich J.A. Self-organized developmental patterning and differentiation in cerebral organoids. *EMBO J.* 2017;36(10):1316-1329. DOI 10.15252/embj.201694700. https://www.ncbi.nlm.nih.gov/pubmed/28283582.

Sakaguchi H., Ozaki Y., Ashida T., Matsubara T., Oishi N., Kihara S., Takahashi J. Self-organized synchronous calcium transients in a cultured human neural network derived from cerebral organoids. *Stem Cell Reports.* 2019;13(3):458-473. DOI 10.1016/j.stemcr.2019.05.029. https://www.ncbi.nlm.nih.gov/pubmed/31257131.

Susaki E.A., Tainaka K., Perrin D., Yukinaga H., Kuno A., Ueda H.R. Advanced CUBIC protocols for whole-brain and whole-body clearing and imaging. *Nat. Protoc.* 2015;10(11):1709-1727. DOI 10.1038/nprot.2015.085. https://www.ncbi.nlm.nih.gov/pubmed/26448360.

Susaki E.A., Takasato M. Perspective: extending the utility of three-dimensional organoids by tissue clearing technologies. *Front. Cell Dev. Biol.* 2021;9:679226. DOI 10.3389/fcell.2021.679226. https://www.ncbi.nlm.nih.gov/pubmed/34195197.

Ueda H.R., Dodt H.U., Osten P., Economo M.N., Chandrashekar J., Keller P.J. Whole-brain profiling of cells and circuits in mammals by tissue clearing and light-sheet microscopy. *Neuron.* 2020;106(3):369-387. DOI 10.1016/j.neuron.2020.03.004. https://www.ncbi.nlm.nih.gov/pubmed/32380050.

van Ineveld R.L., Ariese H.C.R., Wehrens E.J., Dekkers J.F., Rios A.C. Single-cell resolution three-dimensional imaging of intact organoids. *J. Vis. Exp.* 2020;160:e60709. DOI 10.3791/60709. https://www.ncbi.nlm.nih.gov/pubmed/32568249.

Wassie A.T., Zhao Y., Boyden E.S. Expansion microscopy: principles and uses in biological research. *Nat. Methods.* 2019;16(1):33-41. DOI 10.1038/s41592-018-0219-4. https://www.ncbi.nlm.nih.gov/pubmed/30573813.

Yu T., Zhu J., Li D., Zhu D. Physical and chemical mechanisms of tissue optical clearing. *iScience.* 2021;24(3):102178. DOI 10.1016/j.isci. 2021.102178. https://www.ncbi.nlm.nih.gov/pubmed/33718830.

ORCID ID

I.E. Pristyazhnuk orcid.org/0000-0003-0226-4213

Acknowledgements. We thank the Microscopic Center of the Siberian Branch of the Russian Academy of Sciences for granting access to microscopic equipment. This study was supported by the Russian Foundation for Basic Research (grant No. 19-29-04067) and by Budget Project (No. 0259-2021-0016).

Ethics. Written informed consent was obtained from the patient. The study was approved by the Scientific Ethics Committee of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (protocol No. 131 from 01.10.2019).

Conflict of interest. The authors declare no competing financial interests.

Received April 11, 2021. Revised October 4, 2021. Accepted October 11, 2021.

Алфавитный указатель авторов статей, опубликованных в журнале в 2021 г.

Прием статей через электронную редакцию на сайте http://vavilov.elpub.ru/index.php/jour
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции»/"Vavilov Journal of Genetics and Breeding"
до 2011 г. выходил под названием «Информационный вестник ВОГиС»/
"The Herald of Vavilov Society for Geneticists and Breeding Scientists".

Регистрационное свидетельство ПИ № ФС77-45870 выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 20 июля 2011 г.

«Вавиловский журнал генетики и селекции» включен ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Российский индекс научного цитирования, ВИНИТИ, базы данных Emerging Sources Citation Index (Web of Science), Zoological Record (Web of Science), Scopus, PubMed Central, Ebsco, DOAJ, Ulrich's Periodicals Directory, Google Scholar, Russian Science Citation Index на платформе Web of Science, каталог научных ресурсов открытого доступа ROAD.

Открытый доступ к полным текстам:
на сайте ИЦиГ СО РАН – bionet.nsc.ru/vogis/
платформе Elpub – vavilov.elpub.ru/index.php/jour
платформе Научной электронной библиотеки – elibrary.ru/title_about.asp?id=32440
PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/journals/3805/

Подписку на «Вавиловский журнал генетики и селекции» можно оформить в любом почтовом отделении России. Индекс издания 42153 по каталогу «Пресса России».

✉ e-mail: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.
Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218.
Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.
Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.

Подписано в печать 21.12.2021. Выход в свет 24.12.2021. Формат 60 × 84 $^{1}/_{8}$. Усл. печ. л. 12.55.
Уч.-изд. л. 13.5. Тираж 150 экз. (1-й завод 1–50 экз.) Заказ № 92. Цена свободная.

Отпечатано в Сибирском отделении РАН, Морской проспект, 2, Новосибирск, 630090.