2023 • 27 • 5

Print ISSN 2500-0462
Online ISSN 2500-3259

Сетевое издание

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.
Периодичность 8 выпусков в год
DOI 10.18699/VJGB-23-51

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

2023 • 27 • 5

Print ISSN 2500-0462
Online ISSN 2500-3259

Online edition

VAVILOV JOURNAL OF GENETICS AND BREEDING

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

Founded in 1997
Published 8 times annually
DOI 10.18699/VJGB-23-51

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 • 2023 • 27 • 5

VAVILOV JOURNAL OF GENETICS AND BREEDING

CONTENTS • 2023 • 27 • 5

Вавиловский журнал генетики и селекции. 2023;27(5):433-439
DOI 10.18699/VJGB-23-52

Original Russian text https://vavilovj-icg.ru/

Identification of new nucleotide sequences of the *Glu-B1-1* gene encoding x-type glutenins in bread wheat

A.A. Galimova[1, 2] ✉, B.R. Kuluev[1, 2]

[1] Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
[2] Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia
✉ aiz.galimova@yandex.ru

Abstract. Studies of the genetic base and polymorphism of bread wheat cultivars aimed at identifying alleles of genes associated with high baking and other economically valuable traits seem to be relevant, since bread wheat, along with all representatives of the Triticeae tribe, has a huge genetic potential for creating cultivars with high technological and rheological properties of grain flour. The aim of this study was sequencing and analysis of the nucleotide sequences of the *Glu-B1-1* gene, and analysis of the predicted amino acid sequences of its protein product in three cultivars of bread wheat. Thus, in the course of genotyping cultivars and lines of bread wheat for the *Glu-B1-1* gene, in the cultivars 'Avesta', 'Leningradka krupnozernaya' and line C-75094, previously undescribed changes in the size of amplifiable regions of the *Glu-B1-1* gene for high-molecular-weight glutenins were found. Comparative analysis of the nucleotide sequences of these genes with known sequences showed the presence of two deletions in 'Avesta' and C-75094 and the presence of seven single-nucleotide substitutions in 'Leningradka krupnozernaya'. Alignment of the predicted Glu-B1 amino acid sequences of the studied accessions and the standard cultivar carrying the *Glu-B1-a* allele showed that deletions in the amino acid sequences of 'Avesta' and C-75094 accessions are localized in the central domain of the protein and affect the amount of tri-, hexa-, and nonapeptides, and in 'Leningradka krupnozernaya', a decrease in GQQ and PGQGQQ by one unit was revealed. In addition, substitutions of five amino acids were found in 'Leningradka krupnozernaya'. Thus, we have found previously undescribed deletions and substitutions in the nucleotide sequences of the *Glu-B1-1* gene for high-molecular-weight glutenins, which lead to changes in amino acid sequences in functionally important regions, namely, in the central domains of protein molecules. The identified mutations can be used for genotyping bread wheat cultivars.
Key words: baking quality; high-molecular-weight glutenin subunits; *Glu-1* genes; genotyping.

For citation: Galimova A.A., Kuluev B.R. Identification of new nucleotide sequences of the *Glu-B1-1* gene encoding x-type glutenins in bread wheat. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):433-439. DOI 10.18699/VJGB-23-52

Идентификация новых нуклеотидных последовательностей гена *Glu-B1-1*, кодирующего глютенины x-типа у мягкой пшеницы

А.А. Галимова[1, 2] ✉, Б.Р. Кулуев[1, 2]

[1] Институт биохимии и генетики – обособленное структурное подразделение Уфимского федерального исследовательского центра Российской академии наук, Уфа, Россия
[2] Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР), Санкт-Петербург, Россия
✉ aiz.galimova@yandex.ru

Аннотация. Мягкая пшеница, наряду с другими представителями трибы Пшениciевых, обладает огромным генетическим потенциалом для создания сортов с высокими технологическими и реологическими свойствами муки. Поэтому исследования генетической базы полиморфизма сортов мягкой пшеницы и выявление аллелей генов, ассоциированных с высокими хлебопекарными признаками, представляются актуальными. Цель данной работы – анализ нуклеотидных последовательностей гена субъединиц x-типа высокомолекулярных глютенинов *Glu-B1-1* и анализ предсказанных аминокислотных последовательностей его белкового продукта у трех генотипов мягкой пшеницы. В ходе генотипирования по гену *Glu-B1-1* у сортов мягкой пшеницы Авеста, Ленинградка крупнозерная и линии С-75094 были обнаружены ранее не описанные изменения в размерах амплифицируемых участков. Сравнительный анализ нуклеотидных последовательностей этих

генов с известными опубликованными последовательностями показал наличие двух делеций у генотипов Авеста и С-75094 и семи однонуклеотидных замен у сорта Ленинградка крупнозерная. Выравнивание предсказанных аминокислотных последовательностей Glu-B1 рассматриваемых генотипов и стандартного сорта, несущего аллель *Glu-B1-a*, показало, что делеции в аминокислотных последовательностях у сорта Авеста и линии С-75094 локализуются в центральном домене белка и влияют на количество три-, гекса- и нонапептидов. У сорта Ленинградка крупнозерная было выявлено уменьшение количества трипептида GQQ и гексапептида PGQGQQ на одну единицу; кроме того, выявлены замены пяти аминокислот. Таким образом, нами обнаружены ранее не описанные делеции и замены в нуклеотидных последовательностях гена высокомолекулярных глютенинов *Glu-B1-1*, которые приводят к изменениям аминокислотных последовательностей в функционально значимых участках, а именно в центральных доменах белковых молекул. Выявленные мутации могут быть использованы при генотипировании сортообразцов мягкой пшеницы.
Ключевые слова: хлебопекарные качества; высокомолекулярные субъединицы глютенина; гены *Glu-1*; генотипирование.

Introduction

High-molecular-weight glutenin subunits (HMW-GS) play an important role in determining the viscoelastic properties of bread wheat grains as they contribute to the formation of larger gluten polymers and are major determinants of dough elasticity (Shewry et al., 1989, 1992, 1995, 1997). Therefore, the characterization of the HMW-GS composition is an important task in bread wheat breeding programs aimed at improving grain quality. This makes it possible to predict the baking qualities of bread wheat cultivars (Payne, 1987; Nucia et al., 2019).

Until recently, SDS electrophoresis of storage proteins was the main method for determining HMW-GS composition, which revealed a huge allelic diversity of HMW-GS in the Triticeae tribe. For example, to date, 52 alleles have been identified for the *Glu-A1* locus of subgenome A, 83 alleles for the *Glu-B1* locus of subgenome B, and 70 alleles for the *Glu-D1* locus of subgenome D (McIntosh et al., 2013).

Recently, protein SDS electrophoresis has been replaced by methods of molecular genetics, which make it possible to distinguish subunits of high-molecular-weight glutenins with similar molecular weights at the genetic level (Vafin et al., 2018; Nucia et al., 2019). However, the nucleotide sequences of most HMW-GS alleles identified using protein electrophoregrams have not been characterized and deposited in databases still. Studies aimed at determining the nucleotide sequences of alleles of genes associated with high or low grain quality are relevant, since their results can be used in marker-assisted and genomic selection of bread wheat.

In the course of genotyping 95 bread wheat cultivars according to the composition of storage protein genes (Galimova et al., 2023), we identified genotypes carrying previously unknown x-type nucleotide sequences encoded by the *Glu-B1-1* locus (designation in accordance with the Catalogue of Gene Symbols for Wheat (McIntosh et al., 2013)). They were found in cultivars Avesta, Leningradka krupnozernaya and line C-75094. This study describes these new deletions and nucleotide substitutions, as well as some characteristics of the amino acid sequences' predicted fragments of the subunits of these high-molecular-weight glutenins allelic variants.

Materials and methods

The materials of the study were winter bread wheat cultivars Sterlinskaya (used as a control sample – a cultivar carrying the allele of the x-type subunit – Bx7), Avesta, spring cultivar Leningradka krupnozernaya and line C-75094, obtained from the VIR collection. According to the VIR, the cultivar Leningradka krupnozernaya and the line C-75094 have low baking qualities. In accordance with the data given in the Russian State Register of Breeding Achievements (https://reestr.gossortrf.ru/sorts/9358556/, accessed 10/15/2022), the Avesta cultivar is characterized by good baking qualities.

Isolation of total DNA from dried bread wheat leaves was performed using CTAB (Doyle J.J., Doyle J.L., 1987). The BxF/BxR primer pair was used to amplify the *Glu-B1-1* gene fragment (Ma et al., 2003). The BxF forward primer is annealed at two regions of the *Glu-B1-1* gene, forming during PCR, together with the reverse primer, two reaction products with sizes of 766 and 630 bp. DNA amplification was carried out according to the program: initial denaturation for 5 min at 95 °C; 35 cycles of denaturation at 95 °C for 40 sec, primer annealing at 58 °C for 40 sec, elongation at 72 °C for 1 min and final elongation for 3 min at 72 °C. Amplification results were visualized in 1.6 % agarose gels with DNA fragment length markers of 100 bp (Evrogen, Russia).

For sequencing PCR products, an average of 500 ng of each PCR product obtained above was used. The products were purified using the following reaction: 1 U alkaline phosphatase (NEB, USA) and 10 U exonuclease I (NEB, USA) in a final volume of 10 µl at 37 °C for 15 min, followed by enzyme inactivation at 85 °C for 15 min. 1 µl (~50 ng) of each purified sample was directly used as a template for sequencing. The reaction was set up using 10 pmol primer and 0.5 µl BigDye™ Terminator v3.1 Ready Reaction Mix in a final volume of 10 µl. The cycles of the sequencing reaction: denaturation at 96 °C for 10 sec, primer annealing at 58 °C for 5 sec, elongation at 60 °C for 4 min for all 30 cycles. Fluorescently labeled PCR products were analyzed using an Applied Biosystems 3500 genetic analyzer (Thermo Fisher Scientific, USA).

Three biological replicates were used when sequencing fragments of the studied genes of each sample. Sequencing was performed at both ends using primers BxF and BxR. Further, for each sample, by aligning the three obtained sequences, one consensus sequence was compiled. This procedure was carried out primarily to avoid possible errors in sequencing. Alignment of nucleotide sequences by the Clustal W method and detection of putative mutations were performed using the MEGA program version 11.0.8 (Molecular Evolutionary Genetics Analysis version 11).

Results

In the course of genotyping 95 cultivars and lines of bread wheat for the *Glu-1* genes of subgenomes A, B, and D with genome-specific primers (Galimova et al., 2023), we found previously undescribed deletions in the nucleotide sequence of the *Glu-B1* gene encoding HMW-GS x-type (*Glu-B1-1*). When analyzing the allelic state of the *Glu-B1-1* gene, the expected products of the amplification reaction were 766 and 630 bp amplicons. The production of reaction products of the indicated sizes would show that the cultivar carries the allele of the x-type subunit Bx7 (in this study, the Sterlinskaya cultivar was taken as a control). However, in the case of three samples (Avesta, Leningradka krupnozernaya cultivars and C-75094 line), during genotyping, reaction products were found that differed from those expected – one reaction product instead of two with a size of 766 and ~669 bp, not previously described in the literature. Amplification of the genomic DNA of the Leningradka krupnozernaya cultivar produced only one reaction product 766 bp in size. Amplification of the DNA of the Avesta cultivar and the C-75094 line also resulted in the formation of one reaction product, while an amplicon larger than 630 bp and less than 700 bp was detected on the electrophoregram. There are data in the literature on the formation of PCR products with a size of 669 bp when using the BxF/BxR primer pair (Ma et al., 2003). This gave reason to believe that the PCR products of the Avesta and C-75094 genotypes represent the previously described fragment of the *Glu-B1-1* gene with a size of 669 bp (Galimova et al., 2023). Therefore, below, the size of this PCR product was designated as ~669 bp (Fig. 1, Table 1).

To determine the nucleotide sequences of the *Glu-B1-1* gene detected fragments, their sequencing was carried out. Comparative analysis of the nucleotide sequences of the *Glu-B1-1* gene fragments of the Avesta and C-75094 genotypes with known sequences from the GenBank database containing annotated DNA and RNA sequences did not reveal complete identity between them. Alignment of the nucleotide sequence of the *Glu-B1-1* gene fragments of the studied samples (C-75094, Avesta) showed their similarity to the x-type subunit of the *i* allele, which has three deletions relative to the *a* allele. In the samples we studied, only two of them were identified (Fig. 2, *a–c*). Thus, the size of the amplified and sequenced fragment of the *Glu-B1-1* gene was 687 bp.

Analysis of the nucleotide sequence of the *Glu-B1-1* gene fragment of the cultivar Leningradka krupnozernaya, for which an amplification product of 766 bp was detected, showed that two single-nucleotide substitutions occurred in one of the two annealing regions of the BxF forward primer (G→A and A→G), which probably prevent annealing of the forward primer (Fig. 3). Presumably, as a result of this, only one reaction product is formed, instead of the expected two.

Since the *Glu-B1-1* gene of the studied genotypes (Avesta, Leningradka krupnozernaya, C-75094) was not completely sequenced, it was supplemented at both ends with flanking regions of the *Glu-B1a* allele (GenBank BK006773) for comparative analysis of their amino acid sequences. Thus, the analysis of the predicted amino acid sequences of the Glu-B1-1 protein of the studied genotypes will be carried out on the basis of data obtained as a result of sequencing of the fragment of the *Glu-B1-1* gene.

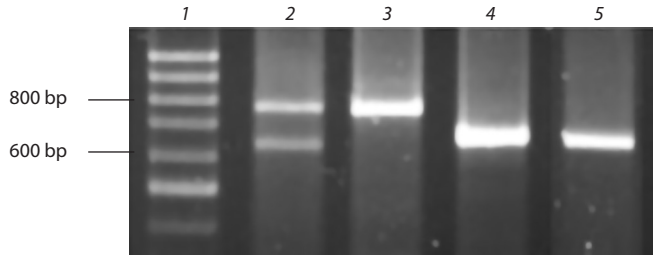


Fig. 1. Electrophoregram of the *Glu-B1-1* gene fragment amplification results with primers BxF/BxR:

1 – 100 bp DNA ladder (Evrogen, Russia); *2* – cultivar Sterlinskaya with expected amplicon sizes of 766 + 630 bp; *3* – cultivar Leningradka krupnozernaya (766 bp); *4* – cultivar Avesta (~669 bp); *5* – line C-75094 (~669 bp).

Table 1. Expected and actual amplicon lengths during genotyping of the studied bread wheat genotypes using BxF/BxR primers

Cultivar/line	Amplicon size, bp	
	expected	actual
Sterlinskaya	766 + 630	766 + 630
Leningradka krupnozernaya		766
Avesta		~669
C-75094		

The central part of the Glu-B1-1 protein is represented by repeating motifs of the tri-, hexa-, and nonapeptides GQQ, PGQGQQ, and GYYPTSPQQ. Despite the fact that the amino acid sequences of the studied genotypes Avesta and C-75094 have a number of amino acids different from the number of amino acids of the *Glu-B1i* allele carrier cultivar, all three compared amino acid sequences have the same number of tri-, hexa-, and nonapeptides repeats (Table 2).

Significant differences in the number of amino acid residues and motifs are observed when comparing the amino acid sequences of the three mentioned cultivars and lines with the amino acid sequence of the cultivar carrying the *Glu-B1a* allele. Thus, in the amino acid sequence of the *Glu-B1a* allele, there are 44 GQQ tripeptides, 18 PGQGQQ hexapeptides, and 11 GYYPTSPQQ nonapeptides, while in the amino acid sequences of the cultivar carrying the *Glu-B1i* allele and in the studied samples Avesta and C-75094, the number of peptides decreases by 3, 1 and 2 motives, respectively (Fig. 4, see Table 2).

Analysis of the number of amino acids in the predicted amino acid sequences of the four compared genotypes showed a decrease in the number of E (glutamic acid), H (histidine), Q (glutamine), P (proline), G (glycine) residues in the studied samples of Avesta, C-75094 and in the cultivar with the *Glu-B1i* allele, compared with the cultivar carrying the *Glu-B1a* allele (see Table 2). Thus, a significant difference in the number of glutamines was found in the amino acid sequences of the *Glu-B1a* allele and the amino acid sequences of the other

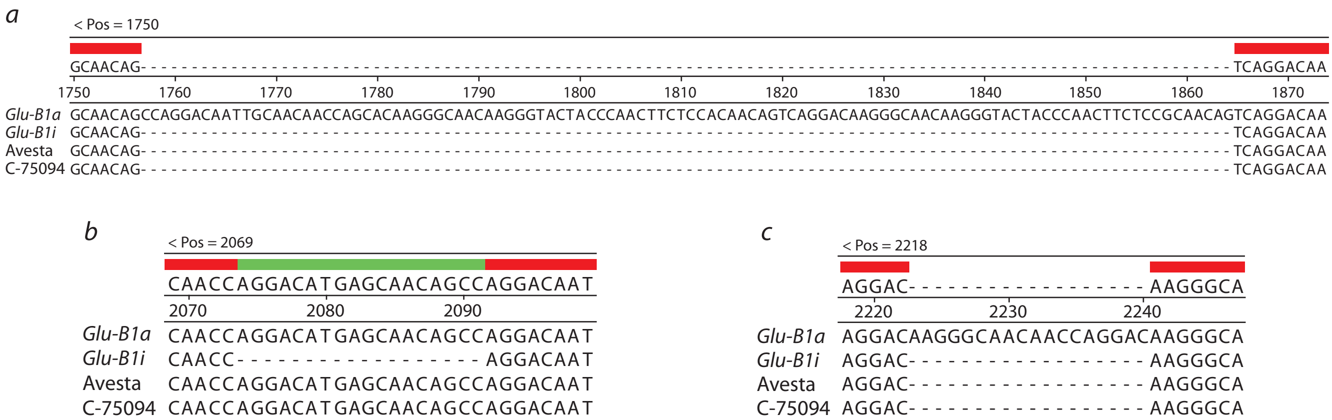


Fig. 2. Alignment of the nucleotide sequences of the *Glu-B1-1* gene fragments of the Avesta and C-75094 genotypes with the *Glu-B1-1* nucleotide sequences of the alleles *Glu-B1a* (GenBank BK006773) and *Glu-B1i* (GenBank AB263219).

a – deletion (1757–1864 nucleotides); *b* – deletion (2074–2091 nucleotides, this deletion distinguishes the nucleotide sequences of *Glu-B1-1* of the *Glu-B1i* allele and the studied genotypes Avesta and C-75094); *c* – deletion (2223–2240 nucleotides).

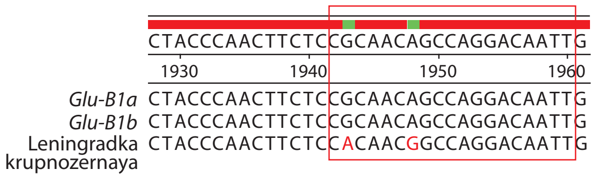


Fig. 3. Single nucleotide substitutions (G→A and A→G) in the nucleotide sequence of the *Glu-B1-1* gene of the Leningradka krupnozernaya cultivar at the site of the BxF forward primer annealing.

The annealing site of the BxF forward primer is marked with a red frame.

studied samples: *n*–18 glutamine residues in the *Glu-B1i* allele and *n*–16 in the amino acid sequences of the Avesta and C-75094 genotypes. In addition, the *Glu-B1-1* amino acid sequences of the Avesta and C-75094 samples show differences in the numbers of prolines (*n*–7) and glycines (*n*–8). The cultivar carrying the *Glu-B1i* allele (GenBank AB263219), in addition to glutamine, differs in the number of 4 more amino acids: glutamic acid, histidine, proline, and glycine (see Table 2, Fig. 4). From the analysis of the amino acid sequence predicted fragments, it can be seen that the sequenced glutenin fragment of the studied samples (Avesta, Leningradka krupnozernaya, C-75094) lacks newly formed cysteine residues that are significant for the formation of disulfide bonds (see Table 2, Fig. 4).

Alignment of the predicted amino acid sequences of cultivars carrying the alleles *Glu-B1a*, *Glu-B1i* and the studied samples Avesta and C-75094 showed that amino acid losses are localized in the central domain of the protein and affect the amount of tri-, hexa-, and nonapeptides (see Fig. 4). However, it should be noted that, in terms of the number of motifs, the amino acid sequences of the Avesta and C-75094 samples do not differ from the Glu-B1-1 amino acid sequence encoded by the *Glu-B1i* allele. At the same time, the Avesta and C-75094 genotypes differ from the *Glu-B1i* allele in the number of glutamic acid, histidine, glutamine, proline, and glycine residues.

A comparative analysis of the predicted amino acid sequences of the Leningradka krupnozernaya cultivar and the cultivar carrying the *Glu-B1a* allele showed a decrease in the amount of the GQQ tripeptide and PGQGQQ hexapeptide by one unit (see Table 2). In addition, substitutions of 5 amino acids were identified, among which there are substitutions of two glycine residues (G→R, G→W), as well as two substitu-

Table 2. The number of repeated HMW-GS motifs and amino acid residues in the studied region of the Glu-B1-1 protein

Allele/genotype	Number of repeats of *Glu-B1-1* locus motifs			Number of amino acid residues				
	Tripeptide GQQ	Hexapeptide PGQGQQ	Nonapeptide GYYPTSPQQ	E	H	Q	P	G
Glu-B1a	44	18	11	*n*	*n*	*n*	*n*	*n*
Glu-B1i	41	17	9	*n*–1	*n*–1	*n*–18	*n*–8	*n*–9
Glu-B1-1 (Avesta, C-75094)	41	17	9	*n*	*n*	*n*–16	*n*–7	*n*–8
Glu-B1-1 (Leningradka krupnozernaya)	43	17	11	*n*	*n*–1	*n*–1+1	*n*	*n*–2

Note. Amino acid residues abbreviations: E – glutamic acid, H – histidine, Q – glutamine, P – proline, G – glycine. *n* – number of amino acid residues of the Glu-B1-1 protein of the genotype (cultivar) carrying the *Glu-B1a* allele; *n*–1+1 – substitution leading to the formation of the amino acid glutamine (H⟶Q), and substitution of the amino acid glutamine with another amino acid (Q⟶R).

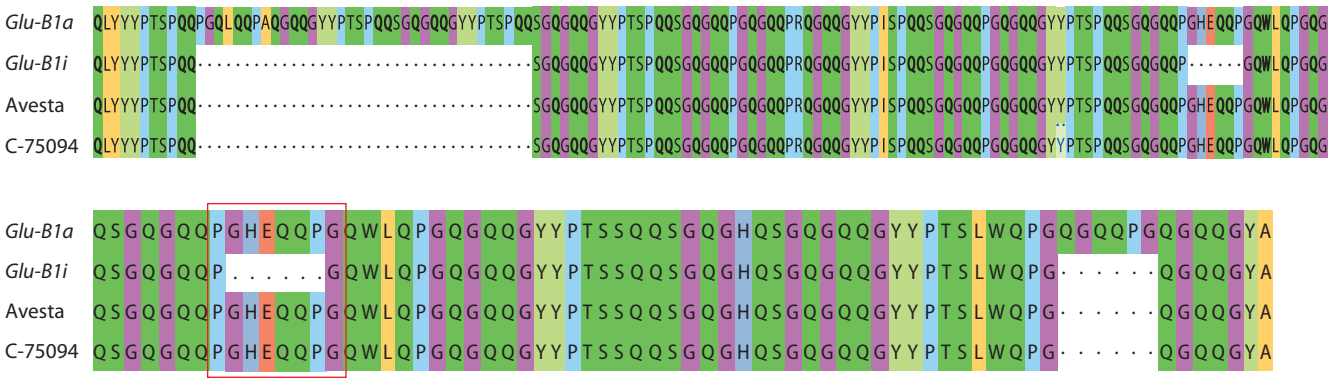


Fig. 4. The results of alignment of the predicted amino acid sequences of the Glu-B1-1 protein of cultivars carrying the alleles *Glu-B1a*, *Glu-B1i*, and the studied samples Avesta, C-75094.

The red frame highlights the region where the amino acid sequences of the studied Avesta and C-75094 samples differ from the amino acid sequence of the protein encoded by the *Glu-B1i* allele.

tions, one of which leads to the formation of the amino acid glutamine (H→Q), and the second, to the glutamine arginine substitution (Q→R).

Discussion

The HMW-GS gene polymorphism is most likely one of the reasons for the high genetic variability of bread wheat traits that affect the technological and rheological properties of flour, and, as a result, baking quality (Patil et al., 2015; Ravel et al., 2020). In the course of genotyping various cultivars and lines of bread wheat at the *Glu-B1-1* locus, we identified amplicons, the nucleotide sequence lengths of which did not correspond to the expected ones. The changes in the nucleotide composition of the *Glu-B1* x-type subunit gene found in this study have not been previously described.

HMW-GS consist of N- and C-terminal domains and a central domain that consists of repeating motifs (Shewry et al., 1992). The N- and C-terminal domains contain more charged residues than the central domain and include most or all of the cysteine residues present in the subunits. Repeat domains are characterized by tri-, hexa-, and nonapeptide motifs in x-type subunits (GQQ, PGQGQQ and GYYPTSPQQ) and hexa- and nonapeptide repeats in y-type subunits (PGQGQQ and GYYPTSLQQ) (Tatham et al., 1990). Thus, two features of the HMW-GS structure, the number and distribution of disulfide bonds, as well as the properties and interactions of the repeating motifs of the central domain, can be related to the determination of protein elasticity (Kohler et al., 1994).

Disulfide bonds are extremely important for the structure of gluten and are significant factors in determining the viscoelastic and rheological properties of the dough (Lindsay et al., 2000; Li et al., 2016). Intra- and intermolecular disulfide bonds form between cysteine residues (Wang et al., 2021). For the predicted fragments of the amino acid sequences of the studied genotypes, an analysis of the cysteine residues content was carried out, which showed no changes in their number.

Although interchain disulfide bonds are critical for stabilizing HMW-GS polymers, nuclear magnetic resonance studies indicate that hydrogen bonds mediated by glutamine side chains may also play an important role in gluten structure stabilization (Belton, 1994; Belton et al., 1995). A high content of glutamine residues has a high ability to form both intra- and intermolecular hydrogen bonds and positively affect dough elasticity (Belton, 1999; Guo et al., 2019). In the genotypes studied, changes in the content of glutamine were found (see Table 2). Note that samples Avesta and C-75094 are characterized by the presence of a greater number of glutamine residues compared to the cultivar carrying the *Glu-B1i* allele.

Variations in the central repeat domain of glutenin proteins are the main reasons for differences in the size of its subunits (Anderson, Greene, 1989; Halford et al., 1992; Shewry et al., 1992; D'Ovidio et al., 1995), which was also shown in our study. It can be seen from the analysis of the central region of the predicted HMW-GS protein in samples Avesta and C-75094 that they differ from the known amino acid sequence of cultivars carrying the *Glu-B1a* allele in the number of motifs (all three types), and, accordingly, in the number of amino acids. They differ from the amino acid sequence of the *Glu-B1i* allele only in the number of amino acids. Thus, the number of central domain motifs in samples Avesta and C-75094 and in the *Glu-B1i* allele is the same and equals 41 tri-, 17 hexa-, and 9 nonapeptides, but the number of amino acid residues in them is different (see Table 2, Fig. 4). For the spring cultivar Leningradka krupnozernaya, a decrease in the amount of tri- and hexapeptides was shown compared to the amino acid sequence of the cultivar carrying the *Glu-B1a* allele (see Table 2). Thus, the amino acid sequences of the cultivars Avesta, Leningradka krupnozernaya and line C-75094 have a lower number of motif repeats compared to the cultivar carrying the *Glu-B1a* allele.

It is known that the length of the central domain, that is, the number of repetitions of its motifs, affects dough elasticity (Gianibelli et al., 2001). It is possible that in the genotypes Avesta, C-75094, and Leningradka krupnozernaya, one of the factors of low baking qualities is a decrease in the number of repeats of the GQQ, PGQGQQ, GYYPTSPQQ motifs and the number of amino acid residues of glutamine and glycine in the Glu-B1-1 protein.

Conclusion

The study describes previously unknown nucleotide sequences of the x-type HMW-GS gene, *Glu-B1-1*, which were found during the genotyping of *Glu-1* gene alleles in the Avesta, Leningradka krupnozernaya cultivars and the C-75094 line. The identified mutations can be used for genotyping cultivars and lines of bread wheat for the HMW-GS genes. They can also be proposed as DNA markers in breeding, but this requires further detailed studies on the effect of the identified mutations on the baking quality of grain. Differences in the nucleotide sequences of the *Glu-B1-1* gene lead to changes in the predicted amino acid sequences of their proteins. Changes in the number of tri-, hexa-, and nonapeptide repeats of the central domain of the protein were predicted in the studied genotypes, and changes in the number of glutamine and glycine were revealed. Since the length of the central domain, as well as the amino acid composition of repetitive motifs, are significant in determining the intra- and intermolecular interactions of a protein molecule, the results of the study can be taken into account when analyzing the viscoelastic properties of the dough and economically valuable traits in the studied cultivars and lines.

References

Anderson O.D., Greene F.C. The characterization and comparative analysis of high-molecular-weight glutenin genes from genomes A and B of a hexaploid bread wheat. *Theor. Appl. Genet.* 1989;77:689-700. DOI 10.1007/BF00261246.

Belton P.S. A hypothesis concerning the elasticity of high molecular weight subunits. In: Wheat Kernel Proteins: Molecular and Functional Aspects. Proceedings of the International Meeting, S. Martino Al Cimino, Viterbo (Italy), September 28–30, 1994. Viterbo: Università degli studi della Tuscia, 1994;159-165.

Belton P.S. Mini review: on the elasticity of wheat gluten. *J. Cereal Sci*. 1999;29(2):103-107. DOI 10.1006/jcrs.1998.0227.

Belton P.S., Colquhoun I.J., Grant A., Wellner N., Field J.M., Shewry P.R., Tatham A.S. FTIR and NMR studies on the hydration of a high-M_r subunit of glutenin. *Int. J. Biol. Macromol.* 1995;17(2):74-80. DOI 10.1016/0141-8130(95)93520-8.

D'Ovidio R., Masci S., Porceddu E. Development of a set of oligonucleotide primers specific for genes at the *Glu-1* complex loci of wheat. *Theor. Appl. Genet.* 1995;91:189-194. DOI 10.1007/BF00220876.

Doyle J.J., Doyle J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 1987;19(1):11-15.

Galimova A.A., Kuluev A.R., Ismagilov K.R., Kuluev B.R. Genetic polymorphism of high-molecular-weight glutenin subunit loci in bread wheat varieties in the Pre-Ural steppe zone. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(4):297-305. DOI 10.18699/VJGB-23-36.

Gianibelli M.C., Larroque O.R., MacRitchie F., Wrigley C.W. Biochemical, genetic, and molecular characterization of wheat glutenin and its component subunits. *Cereal Chem.* 2001;78(6):635-646. DOI 10.1094/CCHEM.2001.78.6.635.

Guo H., Wu J., Lu Y., Yan Y. High-molecular-weight glutenin 1Bx17 and 1By18 subunits encoded by *Glu-B1i* enhance rheological properties and bread-making quality of wheat dough. *J. Food Qual*. 2019;2019:1958747. DOI 10.1155/2019/1958747.

Halford N.G., Field J.M., Blair H., Urwin P., Moore K., Robert L., Thompson R., Flavell R.B., Tatham A.S., Shewry P.R. Analysis of HMW glutenin subunits encoded by chromosome 1A of bread wheat (*Triticum aestivum* L.) indicates quantitative effects on grain quality. *Theor. Appl. Genet.* 1992;83:373-378. DOI 10.1007/BF00224285.

Kohler P., Keck B., Muller S., Wieser H. Disulphide bonds in wheat gluten. In: Wheat Kernel Proteins: Molecular and Functional Aspects. Proceedings of the International Meeting, S. Martino Al Cimino, Viterbo (Italy), September 28–30, 1994. Viterbo: Università degli studi della Tuscia, 1994;45-54.

Li X.J., Liu T.H., Song L.J., Zhang H., Li L.Q., Gao X. Influence of high-molecular-weight glutenin subunit composition at *Glu-A1* and *Glu-D1* loci on secondary and micro structures of gluten in wheat (*Triticum aestivum* L.). *Food Chem.* 2016;213:728-734. DOI 10.1016/j.foodchem.2016.07.043.

Lindsay M.P., Tamas L., Appels R., Skerritt J.H. Direct evidence that the number and location of cysteine residues affect glutenin polymer structure. *J. Cereal Sci.* 2000;31(3):321-333. DOI 10.1006/jcrs.2000.0309.

Ma W., Zhang W., Gale K. Multiplex-PCR typing of high molecular weight glutenin alleles in wheat. *Euphytica*. 2003;134:51-60. DOI 10.1023/A:1026191918704.

McIntosh R.A., Yamazaki Y., Dubcovsky J., Rogers W.J., Morris C., Appels R., Xia X.C. In: Catalogue of Gene Symbols for Wheat. Proceed. of the 12th International Wheat Genetics Symposium. Yokohama, 2013;84-95.

Nucia A., Okoń S., Tomczyńska-Mleko M. Characterization of HMW glutenin subunits in European spring common wheat (*Triticum aestivum* L.). *Genet. Resour. Crop Evol*. 2019;66:579-588. DOI 10.1007/s10722-018-00733-x.

Patil V.R., Talati J.G., Singh C., Parekh V.B., Jadeja G.C. Genetic variation in glutenin protein composition of aestivum and durum wheat cultivars and its relationship with dough quality. *Int. J. Food Properties.* 2015;18(9-12):2393-2408. DOI 10.1080/10942912.2014.980948.

Payne P.I. Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Ann. Rev. Plant Physiol*. 1987;38:141-153. DOI 10.1146/annurev.pp.38.060187.001041.

Ravel C., Faye A., Ben-Sadoun S., Ranoux M., Dardevet M., Dupuits C., Exbrayat F., Poncet Ch., Sourdille P., Branlard G. SNP markers for early identification of high molecular weight glutenin subunits (HMW-GSs) in bread wheat. *Theor. Appl. Genet.* 2020;133(3):751-770. DOI 10.1007/s00122-019-03505-y.

Shewry P.R., Halford N.G., Tatham A.S. The high molecular weight subunits of wheat, barley and rye: genetics, molecular biology, chemistry and role in wheat gluten structure and functionality. In: Miflin B.J. (Ed.) Oxford Surveys of Plant and Molecular Cell Biology. London: Oxford Univ. Press, 1989;163-219.

Shewry P.R., Halford N.G., Tatham A.S. High molecular weight subunits of wheat glutenin. *J. Cereal Sci.* 1992;15(2):105-120. DOI 10.1016/s0733-5210(09)80062-3.

Shewry P.R., Tatham A.S., Barro F., Barcelo P., Lazzeri P. Biotechnology of breadmaking: unravelling and manipulating the multiprotein gluten complex. *BioTechnology.* 1995;13:1185-1190. DOI 10.1038/nbt1195-1185.

Shewry P.R., Tatham A.S., Lazzeri P. Biotechnology of wheat quality. *J. Sci. Food Agric*. 1997;73(4):397-406. DOI 10.1002/(SICI)1097-0010(199704)73:4<397::AID-JSFA758>3.0.CO;2-Q.

Tatham A.S., Drake A.F., Shewry P.R. Conformational studies of synthetic peptides corresponding to the repetitive region of the high molecular weight (HMW) glutenin subunits of wheat. *J. Cereal Sci*. 1990;11(3):189-200. DOI 10.1016/S0733-5210(09)80163-X.

Vafin R., Rzhanova I., Askhadullin D., Askhadullin D., Vasilova N. Screening of the genotypes of bread wheat (*Triticum aestivum* L.) by the allelic variants of *Waxy* genes and HMW glutenin subunits. *Acta Agrobot.* 2018;71(4):1746. DOI 10.5586/aa.1746.

Wang L., Yu J.J., Wang C.C. Protein disulfide isomerase is regulated in multiple ways: consequences for conformation, activities, and pathophysiological functions. *BioEssays*. 2021;43(3):2000147. DOI 10.1002/bies.202000147.

ORCID ID

A.A. Galimova orcid.org/0000-0002-7068-3359
B.R. Kuluev orcid.org/0000-0002-1564-164X

Acknowledgements. Research on genotyping and DNA markers of bread wheat was supported by a grant from the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2021-1066 dated September 28, 2021), studies on sequencing and analysis of the nucleotide sequences of glutenin genes were carried out as part of state assignment No. 122030200143-8.

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

Received October 25, 2022. Revised April 27, 2023. Accepted May 12, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):440-446
DOI 10.18699/VJGB-23-53

Original Russian text https://vavilovj-icg.ru/

5′-UTR allelic variants and expression of the lycopene-ε-cyclase *LCYE* gene in maize (*Zea mays* L.) inbred lines of Russian selection

D.Kh. Arkhestova[1, 2], B.R. Shomakhov[2], A.V. Shchennikova[1] ✉, E.Z. Kochieva[1, 3]

[1] Institute of Bioengineering, Federal Research Center "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, Russia
[2] Institute of Agriculture – Branch of the Federal Scientific Center "Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences", Nalchik, Russia
[3] Lomonosov Moscow State University, Moscow, Russia
✉ shchennikova@yandex.ru

Abstract. In breeding, biofortification is aimed at enriching the edible parts of the plant with micronutrients. Within the framework of this strategy, molecular screening of collections of various crops makes it possible to determine allelic variants of genes, new alleles, and the linkage of allelic variants with morphophysiological traits. The maize (*Zea mays* L.) is an important cereal and silage crop, as well as a source of the main precursor of vitamin A – β-carotene, a derivative of the β,β-branch of the carotenoid biosynthesis pathway. The parallel β,ε-branch is triggered by lycopene-ε-cyclase LCYE, a low expression of which leads to an increase in provitamin A content and is associated with the variability of the 5′-UTR gene regulatory sequence. In this study, we screened a collection of 165 maize inbred lines of Russian selection for 5′-UTR *LCYE* allelic variants, as well as searched for the dependence of *LCYE* expression levels on the 5′-UTR allelic variant in the leaves of 14 collection lines. 165 lines analyzed were divided into three groups carrying alleles A2 (64 lines), A5 (31) and A6 (70), respectively. Compared to A2, allele A5 contained two deletions (at positions -267–260 and -296–290 from the ATG codon) and a G_{251}→T substitution, while allele A6 contained one deletion (-290–296) and two SNPs (G_{251}→T, G_{265}→T). Analysis of *LCYE* expression in the leaf tissue of seedlings from accessions of 14 lines differing in allelic variants showed no associations of the 5′-UTR *LCYE* allele type with the level of gene expression. Four lines carrying alleles A2 (6178-1, 6709-2, 2289-3) and A5 (5677) had a significantly higher level of *LCYE* gene expression (~0.018–0.037) than the other 10 analyzed lines (~0.0001–0.004), among which all three allelic variants were present.
Key words: *Zea mays* L.; maize inbred lines; lycopene-ε-cyclase; *LCYE* alleles; gene expression.

For citation: Arkhestova D.Kh., Shomakhov B.R., Shchennikova A.V., Kochieva E.Z. 5′-UTR allelic variants and expression of the lycopene-ε-cyclase *LCYE* gene in maize (*Zea mays* L.) inbred lines of Russian selection. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):440-446. DOI 10.18699/VJGB-23-53

Аллельные варианты 5′-UTR и экспрессия гена ликопин-ε-циклазы *LCYE* у инбредных линий кукурузы *Zea mays* L. российской селекции

Д.Х. Архестова[1, 2], Б.Р. Шомахов[2], А.В. Щенникова[1] ✉, Е.З. Кочиева[1, 3]

[1] Институт биоинженерии им. К.Г. Скрябина, Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук, Москва, Россия
[2] Институт сельского хозяйства – филиал Федерального научного центра «Кабардино-Балкарский научный центр Российской академии наук», Нальчик, Россия
[3] Московский государственный университет им. М.В. Ломоносова, Москва, Россия
✉ shchennikova@yandex.ru

Аннотация. Селекционная биофортификация направлена на обогащение съедобных частей растения микронутриентами. В рамках данной стратегии молекулярный скрининг коллекций различных культур позволяет определять аллельные варианты генов, новые аллели и сцепленность аллельных вариантов с морфофизиологическими признаками. Кукуруза *Zea mays* L. является важной зерновой и силосной культурой, а также источником основного предшественника витамина А – β-каротина, производного β,β-ветви пути биосинтеза каротиноидов. Параллельная β,ε-ветвь запускается ликопин-ε-циклазой LCYE, низкая экспрессия которой приводит к росту содержания провитамина А и связана с вариабельностью регуляторной последовательности 5′-UTR гена. В настоящем исследовании проведены скрининг коллекции 165 инбредных линий кукурузы российской селекции на варианты аллелей 5′-UTR *LCYE*, а также поиск зависимости уровня экспрессии гена *LCYE* от аллельного варианта 5′-UTR в листьях 14 коллекционных линий. Проанализированные 165 линий разделились на три группы, несущие аллели А2 (64 линии), А5 (31) и А6 (70). В сравнении с А2, аллель А5 содержал две делеции (в позициях -267–260 и -296–290 от ATG-кодона) и замену G_{251}→T, тогда как аллель А6 – одну делецию (-290–296) и две замены (G_{251}→T,

G_{265}→T). Анализ экспрессии гена *LCYE* в листовой ткани проростков образцов 14 линий, различающихся аллельными вариантами, показал отсутствие ассоциаций варианта аллеля 5′-UTR *LCYE* с уровнем экспрессии гена. Четыре линии, несущие аллели A2 (образцы 6178-1, 6709-2, 2289-3) и A5 (образец 5677), имели значительно более высокий уровень экспрессии гена *LCYE* (~0.018–0.037) по сравнению с остальными десятью проанализированными линиями (~0.0001–0.004), среди которых были представлены все три аллельных варианта.
Ключевые слова: *Zea mays* L.; инбредные линии кукурузы; ликопин-ε-циклаза; аллели *LCYE*; экспрессия гена.

Introduction

Maize *Zea mays* L. is an important world crop. Climatic conditions in Russia favor the predominant cultivation of corn for silage (immature cobs, leaves and stems), which makes up about 50 % of the dry matter of the main feed for farm animals (Cabiddu et al., 2019; Graulet et al., 2019; Mitani et al., 2021). As a grain crop, corn is grown only in the southern regions of Russia. According to the Ministry of Agriculture, 1.4 million tons of grain were harvested in 2021, which is ~50 times less compared to wheat (https://mcx.gov.ru/press-service/news/sbor-zernovykh-v-rossii-dostig-100-mln-tonn/).

Both grain and silage of maize are considered important sources of antioxidants, including provitamin A, represented by three carotenoid compounds: β-carotene (provides two units of retinol – active vitamin A, when oxidatively broken down), β-cryptoxanthin (provides one unit of retinol, but with greater bioavailability than β-carotene) and α-carotene (one unit of retinol) (LaPorte et al., 2022). In addition to dietary significance, enrichment with β-carotene and β-cryptoxanthin contributes to an essential reduction in aflatoxin contamination of corn grain (Suwarno et al., 2019). In the grain of the most popular varieties and hybrids, according to various data, carotenoids range from 9.55 to 62.96 μg/g (Trono, 2019), while in the leaves their content is already about 200 μg/g (Li et al., 2008; Suwarno et al., 2019).

β-Carotene and β-cryptoxanthin are products of the β,β-branch of the carotenoid biosynthetic pathway (Fig. 1): lycopene-β-cyclase (LCYB) catalyzes the formation of β-ionone rings at both ends of the all-*trans*-lycopene molecule with the formation of β-carotene, the hydroxylation of which leads to the synthesis of xanthophylls, including β-cryptoxanthin (Rosas-Saavedra, Stange, 2016). The α-carotene molecule, a product of the β,ε-branch triggered by lycopene-ε-cyclase (LCYE) (see Fig. 1), is characterized by a β-ring at one end and an ε-ring at the other end of the isoprenoid chain (Rosas-Saavedra, Stange, 2016). A signature of the predominance of the β,β- or β,ε-branch is the orange or light yellow, respectively, color of the corn grain (Harjes et al., 2008; Babu et al., 2013; Zunjare et al., 2018).

Maize breeding for provitamin A biofortification uses the *LCYE* gene, as well as the β-carotene hydroxylase 1 (*CrtRB1*) gene, which catalyzes the conversion of β-carotene to β-cryptoxanthin (LaPorte et al., 2022). A decrease in the expression level of the first, second, or both genes simultaneously leads to a shift in the metabolic pathway towards the biosynthesis of β-carotene as the most promising source of provitamin A (Harjes et al., 2008; Yan et al., 2010; Muthusamy et al., 2014; Liu et al., 2015; Zunjare et al., 2018; LaPorte et al., 2022).

One of the main conditions for successful breeding is the availability of donors of allelic variants linked to the desired economically valuable traits. Maize accessions, which are characterized by low grain expression of *LCYE* and/or *CrtRB1* genes, are used in breeding, including provitamin A biofortification (Pixley et al., 2013; Muthusamy et al., 2014; Liu et al., 2015; Menkir et al., 2017; Prasanna et al., 2020). It has been shown that a reduced level of *LCYE* transcripts may be associated with polymorphisms in the 5′-UTR sequence of the gene (Harjes et al., 2008; Babu et al., 2013; Zunjare et al., 2018).

With all the promise of *LCYE* alleles in maize biofortification, as well as the widespread use of maize for silage, studies of the gene activity are limited to corn grain and barely touch upon photosynthetic organs. Previously, we have shown an inverse relationship between the content of β-carotene and the level of *LCYE* gene expression in the leaf tissue of maize seedlings (Arkhestova et al., 2022).

In the study, we assessed the correlations between the level of *LCYE* expression and allelic variants of the 5′-UTR regulatory region of the gene in a collection of 165 inbred maize lines. We also analyzed the relationship between the level of expression and the allelic variant of the lycopene-ε-cyclase gene in the leaves of 14 accessions differing in the 5′-UTR *LCYE* alleles. The maize lines used in the work were obtained by breeders of two organizations in the Kabardino-Balkarian

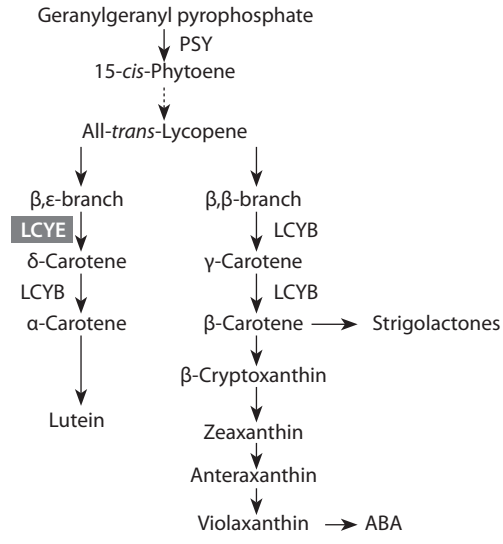


Fig. 1. A simplified scheme of xanthophylls biosynthesis.

Phytoene synthase PSY catalyzes the synthesis of phytoene, from which *trans*-lycopene is formed as a result of several successive reactions. Further, the metabolic pathway is divided into β,β- and β,ε-branches, which lead to the production of β-carotene and α-carotene, respectively, and then β-cryptoxanthin and xanthophylls – zeaxanthin, antheraxanthin, violaxanthin (β,β-carotenoids) and lutein (β,ε-carotenoids). The modification of β-carotene and violaxanthin under the action of carotenoid-cleaving dioxygenases leads to the synthesis of apocarotenoids – strigolactones and abscisic acid (ABA), respectively.

Republic (KBR, North Caucasian Federal District), which, due to climatic conditions, ranks first in the Russian Federation in the cultivation of corn, according to the allocated sown area (Rosstat data for 2021; https://rosstat.gov.ru/storage/mediabank/Census_agr_2021.pdf).

Materials and methods

Accessions of 165 *Z. mays* inbred lines from two breeding organizations (JSC "OTBOR" and the Institute of Agriculture KBSC RAS) were used for the study; the lines are currently being tested and are listed in the work under the numbers assigned to them by the breeders (see Supplementary Material)[1].

The seed material of plants grown in the field in 2022 (KBR, Russia) was kindly provided by the JSC "OTBOR" (KBR, Russia) and the Institute of Agriculture of the Branch of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences (IA KBSC RAS, KBR, Russia). According to the originators (JSC "OTBOR", IA KBSC RAS), the lines differ in grain color (Fig. 2, see Suppl. Material). Germinated grains were grown until the 4th true leaf appeared in moist soil under controlled conditions (23 °C/25 °C, 16/8 h day/night) of the experimental climate control facility in the Institute of Bioengineering (Research Center of Biotechnology, Russian Academy of Sciences). Leaf material was collected and used for analysis of *LCYE* allelic variants and expression.

To identify allelic variants, genomic DNA was isolated from the leaf material according to (Filyushin et al., 2023) and used as a template for PCR amplification of the 5′-UTR region of the *LCYE* gene under the following conditions: initial denaturation (5 min, 95 °C), 32 cycles (denaturation 1 min, 95 °C; annealing 30 s, 60 °C; synthesis 45 s, 72 °C). Amplification primer sequences were: F2 (5′-AAGCATCCGACCAAAATAACAG-3′) and R2 (5′-GAGAGGGAGACGACGAGACAC-3′) (Harjes et al., 2008). The generated fragments purified from the gel (Zymoclean™ Gel DNA Recovery Kit, ZymoResearch, USA) were sequenced (primer F2) on an ABI 310 Capillary DNA Analyzer (Applied Biosystems, USA; Core Facility Use Bioengineering, Russian Academy of Sciences). Structural analysis was performed using NCBI-BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and MEGA 7.0 (Kumar et al., 2016).

To analyze gene expression, total RNA was isolated from 50–100 mg of leaf tissue using the RNeasy Plant Mini Kit (QIAGEN, Germany), purified from genomic DNA impurities (RNase free DNasy set, QIAGEN), and used for cDNA synthesis (GoScript™ Reverse Transcription System, Promega, USA). RNA quality was assessed by electrophoresis in 1.5 % agarose gel. The concentration of RNA and cDNA preparations was determined fluorimetrically using Qubit 4 (Thermo Fisher Scientific, USA) and reagents Qubit RNA HS Assay Kit and Qubit DS DNA HS Assay Kit (Invitrogen, USA).

The level of transcripts of the lycopene-ε-cyclase gene *LCYE* in the leaves of maize seedlings was determined by quantitative (q) real-time (RT) PCR (qRT-PCR). The data were normalized to the level of *Z. mays polyubiquitin* gene transcripts (NM_001329666.1; primers ZmUBI-rtF 5′-ATCGTGGTTGTGGCTTCGTTG-3′ and ZmUBI-rtR 5′-GCTGCAGAAGAGTTTTGGGTACA-3′).

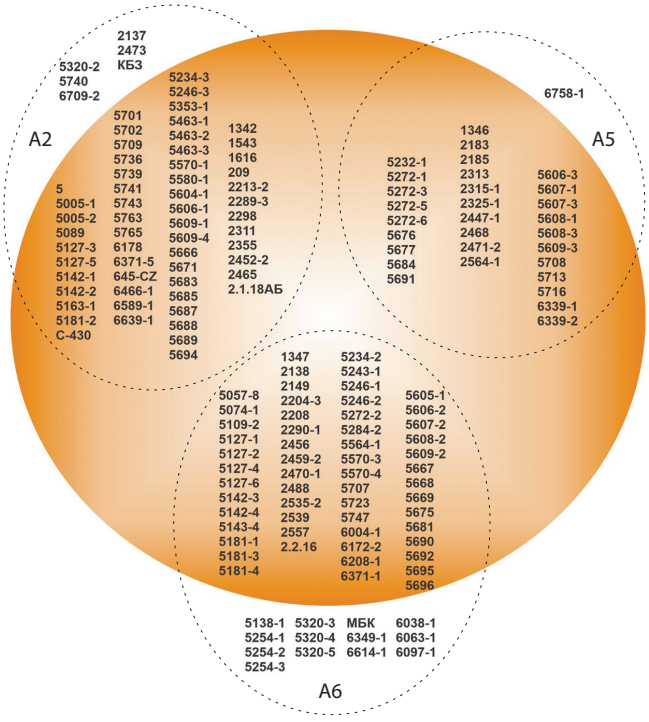


Fig. 2. Distribution of 165 *Zea mays* L. inbred lines (domestic breeding) used in the study according to grain color and 5′-UTR allelic variants of the *LCYE* gene (delimited by ellipses, within which the corresponding allele is indicated – A2, A5, or A6). White-grain accessions are placed in the unpainted part of the ellipses; lines with grain coloring in various shades of yellow, orange and red are located in the colored part.

3 ng of cDNA, cDNA-specific primers (ZmLcyE-F 5′-TTTACGTGCAAATGCAGTCAA-3′; ZmLcyE-R 5′-TGACTCTGAAGCTAGAGAAAG-3′), kit "Reaction mixture for real-time PCR in the presence of SYBR GreenI and ROX" (Sintol, Russia), and thermal cycler CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, USA) were used for the reaction. The reactions were carried out in three technical and two biological replicates under the following conditions: preliminary denaturation (5 min, 95 °C); 40 cycles (15 s, 95 °C; 50 s, 62°C).

qRT-PCR results were statistically processed using GraphPad Prism v.8 (GraphPad Software Inc., USA; https://www.graphpad.com/scientific-software/prism/). Data were expressed as mean with standard deviation (±SD) based on three technical and two biological replicates. The *t*-test was used to assess the significance of differences in gene expression between maize lines ($p < 0.05$ indicates statistical significance of differences).

Results

The study was focused on the characterization of the allelic variability of the *LCYE* gene 5′-UTR sequence in maize inbred lines of domestic selection, as well as the analysis of the level of gene transcripts in the leaf tissue of seedlings of lines that differ in allelic variants of the 5′-UTR of the *LCYE* gene.

To determine allelic variants of the lycopene-ε-cyclase gene, amplification and sequencing of the 5′-UTR region of

[1] Supplementary Material is available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/Suppl_Arkhestova_Engl_27_5.pdf

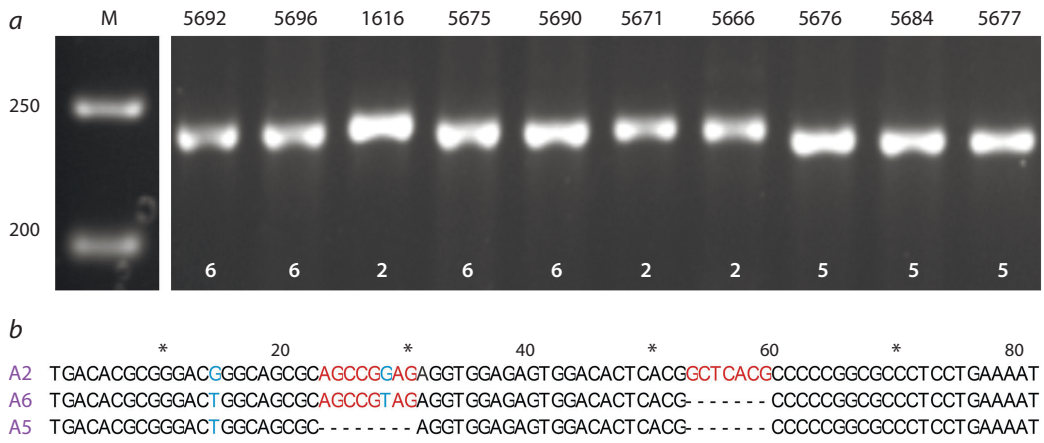


Fig. 3. An example of electrophoretic separation of PCR amplified fragments (*a*) corresponding to the 5′-UTR *LCYE* allelic variants A2 (248 bp), A5 (233 bp), and A6 (240 bp) in a 2.5 % agarose gel (M is the length marker Thermo Fisher GeneRuler 50 bp) and comparative alignment of the variable region of the A2, A5, and A6 alleles (*b*).

Indels in red, SNPs in blue.

the *LCYE* gene was performed (Fig. 3). Expected fragment sizes: alleles A2 (248 bp, according to Harjes et al., 2008), A5 (233 bp, according to Arkhestova et al., 2023) and A4 (993 bp, according to Harjes et al., 2008).

As a result, no A4 variants were found, while alleles A2 and A5 were shown to be present in the analyzed accessions. In addition to these variants described earlier, a new, uncharacterized A6 allele (240 bp) was identified (see Fig. 3). In total, out of the analyzed 165 maize accessions, 64 lines contained the A2 allele, the smallest number of accessions (31) contained the A5 allele, and the largest number of accessions (70) contained the A6 allele (see Fig. 2).

It was determined that, unlike A2, 5′-UTR of the A5 allele contains two deletions (at positions -267–260 and -296–290 from the ATG codon), while the new A6 allele has only one of these deletions (at position -296–290 from the ATG codon) (see Fig. 3). In addition, allele-specific single nucleotide substitutions were found in comparison with A2: G_{251}→T (in the sequence of A5 and A6); G_{265}→T (only for A6) (see Fig. 3). The positions of deletions and substitutions are given in accordance with the *LCYE* gene sequence available in the NCBI database (NCBI Gene ID OK032387.1).

Variants of the A2/A5/A6 alleles were found in accessions with white (59/30/57) and pigmented (5/1/13) grain, respectively. In order to understand whether the 5′-UTR *LCYE* allele (A2, A5, or A6) is associated with the level of *LCYE* gene transcripts in photosynthetic tissue, *LCYE* expression was analyzed in the leaf tissue of 14 lines differing in allelic variants (Fig. 4). Accessions for qRT-PCR were selected based on two assumptions. First, all three types of alleles (A2, A5, and A6) of the *LCYE* gene had to be present in the analysis. Secondly, preference was given to maize accessions that are most interesting for breeders.

Considering the qRT-PCR data, the analyzed accessions were clearly divided into two groups, which significantly differed in the level of *LCYE* gene expression (see Fig. 4). The first group combined four lines with a high expression level (~0.018–0.037), which was ~4.5–370.0 times higher than in ten accessions of the second group (~0.0001–0.004). Among the four lines with high expression, three (6178-1, 6709-2, and 2289-3) carried the A2 allele and one (5677) carried the A5 allele. At the same time, the level of *LCYE* transcripts in line 5677 (A5) was ~1.5–2.0 times lower than in three lines with the A2 allele (see Fig. 4).

In the group of lines with low gene expression, all three allelic variants of 5′-UTR *LCYE* were present (see Fig. 4). This group included all three accessions with the A6 allele taken for analysis (MBK, 6097-1, 5254-3). Against the background of low gene activity in the group, lines 5580-1 and 645CZ (allelic variant A2) were characterized by increased activity of the *LCYE* gene (see Fig. 4).

It should be noted that lines carrying the A2 or A5 alleles were present both in the first and second groups (see Fig. 4).

Thus, using 14 lines representing all three variants of the 5′-UTR *LCYE* allele as an example, it was shown that there was no dependence of the level of *LCYE* gene transcripts on the allele (A2, A5, or A6) in the leaf photosynthetic tissue.

Discussion

Over the past decades, one of the most promising breeding trends has been biofortification (a strategy to improve the nutritional quality of cultivated plants by breeding methods using a number of biotechnologies) aimed at enriching the edible parts of the plant with micronutrients (vitamins, minerals and trace elements) (Medina-Lozano, Díaz, 2022). This approach, combined with molecular methods for identifying parental forms and analyzing hybrid progeny, has made it possible to obtain a large number of high-yielding varieties and hybrids of crops, including maize hybrids with a high content of provitamin A (Pixley et al., 2013; Muthusamy et al., 2014; Liu et al., 2015; Menkir et al., 2017; Prasanna et al., 2020).

In this regard, of interest is the molecular screening of collections of various crops, which makes it possible to determine allelic variants of genes, new alleles, and the linkage of alleles with morphophysiological characteristics (Langridge, Fleury, 2011; Pasala, Pandey, 2020). From a scientific point of view, the totality of the results obtained contributes to a more accurate understanding of the function of specific genes. At

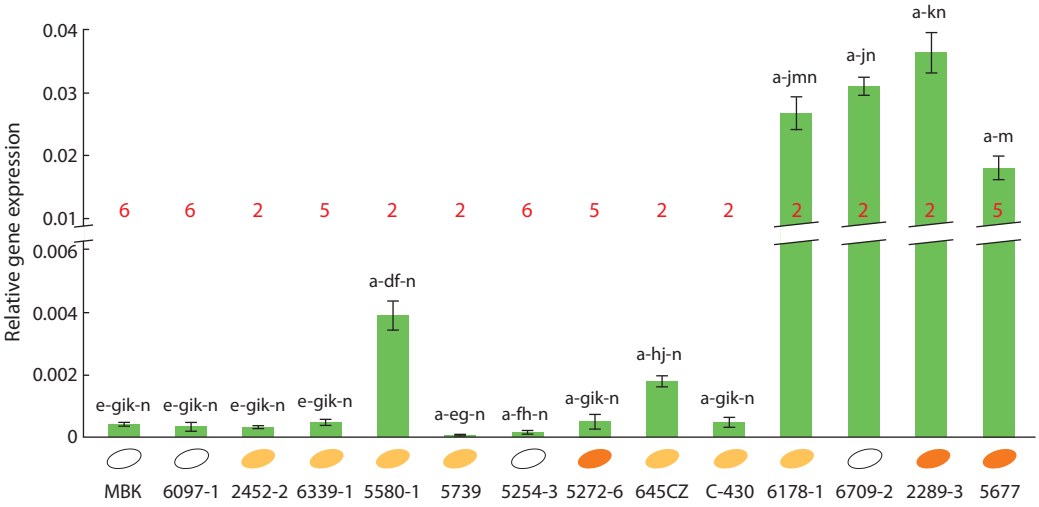


Fig. 4. Relative level of *LCYE* gene expression in the leaf tissue of seedlings of 14 maize inbred lines.

The letters a–n indicate a significant difference ($p < 0.05$) of a particular gene expression value from the values for other accessions (lines 1–14 correspond to letters a–n). The allelic variant (A2 – 2, A5 – 5, or A6 – 6) is indicated in red for each accession. The grain color (white, bright yellow or orange) is shown by a colored ellipse next to the accession name.

the same time, screening of collections is an important stage of breeding, as it allows to assess the representation in the breeding material of a specific allelic variant that determines the desired economically important trait, as well as to identify donors of this trait for introduction into the breeding process (Langridge, Fleury, 2011).

In this work, accessions of 165 maize inbred lines of domestic breeding were characterized by the 5′-UTR allelic variant of the *LCYE* gene. The activity of lycopene-ε-cyclase is considered to be inversely related to the biogenesis of β-carotene and the corresponding β,β-xanthophylls, which, in turn, determines the color of the grain (pale yellow and orange indicate a shift towards β,ε- and β,β-branches (see Fig. 1), respectively) (Harjes et al., 2008; Babu et al., 2013; Zunjare et al., 2018). A decrease in *LCYE* gene activity and, as a result, significant changes in the ratio of β,ε- and β,β-carotenoids are closely associated with mutations in the 5′-UTR region of the gene, namely, with insertions of transposable elements near the translation initiation point (alleles A1 and A4) (Harjes et al., 2008). At the same time, the highest efficiency of provitamin A accumulation in grain is linked to the A4 allele (Babu et al., 2013; Zunjare et al., 2018). Given this, molecular markers have been developed to identify various allelic variants of the 5′-UTR sequence of the *LCYE* gene (Harjes et al., 2008; Babu et al., 2013). Screening of *Z. mays* collections using these markers made it possible to identify donors of the A4 allele and introduce them into breeding programs to obtain maize lines and hybrids with a high content of provitamin A (Harjes et al., 2008; Babu et al., 2013).

Our analysis of 165 inbred lines did not reveal accessions carrying the A4 allele linked to the enhanced accumulation of provitamin A. This indicates that for biofortification for an increased content of provitamin A, sources other than the lines of this collection should be involved. However, in addition to the A2 allele, screening detected two other variants of the 5′-UTR region, the A5 allele (Arkhestova et al., 2023), as well as the previously undescribed A6 allele (see Fig. 3).

Next, we tested the possibility of a relationship between the 5′-UTR *LCYE* allele variant (A2, A5, or A6) and the level of *LCYE* expression. Photosynthetic tissues of seedlings were used for the analysis because data on the correlation of *LCYE* alleles with the content of carotenoids in corn are limited mainly to grain, and also due to the predominant cultivation of corn for silage in Russia, since the presence of such a correlation can serve as the basis for identifying donors of the trait of increased biosynthesis provitamin A in maize photosynthetic tissue (silage). In the case of a clear association of any allele with the level of *LCYE* gene expression, donors of this allele could be used in the breeding of silage corn with a high content of provitamin A.

qRT-PCR was performed on 14 accessions out of 165 lines studied in this work. Among these 14 lines, all three alleles of the 5′-UTR *LCYE* (A2, A5, and A6) were present. Since the result showed no association of the detected 5′-UTR *LCYE* allelic variants with the level of *LCYE* expression in the leaf (see Fig. 4), it can be assumed that there is no such association for the rest of the analyzed collection. The absence of the desired dependence can partly be explained by the fact that the ratio of the amount of synthesized β,ε- and β,β-carotenoids depends on the level of expression of not only *LCYE*, but also the gene of lycopene-β-cyclase *LCYB* (Bai et al., 2009) or other carotenogenesis genes, for example, phytoene synthase genes (*PSY*) (Orlovskaya et al., 2016).

qRT-PCR data showed a clear division of accessions into two groups – with high and low expression of *LCYE*. Considering the known antioxidant role of xanthophylls in plant photoprotection (Jahns, Holzwarth, 2012), it can be assumed that in 10 lines with low *LCYE* expression (see Fig. 4), such protection is carried out mainly by carotenoids of the main xanthophyll cycle (β,β-branch). At the same time, in the

remaining 4 lines with high *LCYE* expression, presumably synthesizing significant amounts of β,ε-carotenoids, photoprotection can actively involve an additional lutein-5,6-epoxide cycle (β,ε-branch).

It is also possible that lines with low *LCYE* expression (see Fig. 4) and a presumed shift of the carotenoid biosynthetic pathway towards the β,β-branch synthesize relatively more phytohormones (strigolactones and abscisic acid) produced by the action of carotenoid-cleaving dioxygenases (Dhar et al., 2020). Thus, ABA is formed by 9-*cis*-epoxycarotenoid dioxygenases NCED from 9-*cis*-violaxanthin and 9-*cis*-neoxanthine (violaxanthin derivatives), while strigolactones are synthesized by cleavage of β-carotene by CCD dioxygenases (Nambara, Marion-Poll, 2005; Cutler et al., 2010; Dhar et al., 2020). ABA plays a crucial role in the adaptability of plants, including *Z. mays*, to various environmental conditions, mediating growth, development, stress response, and nutrient distribution (Huang et al., 2017; Yue et al., 2021). Strigolactones are actively involved in the stress response of plants (López-Ráez et al., 2010). In view of the above, the expected increased synthesis of apocarotenoids may indicate greater adaptability of maize lines with low *LCYE* expression in the vegetative tissue.

Conclusion

In this work, we analyzed variants of the 5′-UTR allele of the lycopene-ε-cyclase *LCYE* gene in the genome of 165 inbred maize lines of Russian selection. As a result, three groups of accessions carrying the A2 (64 lines), A5 (31), or A6 (70) alleles were identified. The shortest of them, the A5 allele, differed by one and two deletions from A6 and A2, respectively. To assess the possible dependence of the *LCYE* mRNA level in leaves on the 5′-UTR allelic variant, gene expression was determined in 14 lines differing in allelic variants. Based on the data obtained, it can be argued that the desired associations are absent. We assume that maize lines with low expression of the *LCYE* gene can serve as a source of traits of increased plant stress resistance and enhanced synthesis of provitamin A in photosynthetic tissue. In this case, the marker will not be the 5′-UTR *LCYE* allelic variant, but the level of expression of the *LCYE* gene. Confirmation of this possibility will require further studies on a larger number of accessions.

References

Arkhestova D.Kh., Kulakova A.V., Khatefov E.B., Shchennikova A.V., Kochieva E.Z. Expression of the lycopene-ε-cyclase *LcyE* gene correlates with the content of β-carotene and chlorophylls in maize vegetative tissues. *Selskokhozyaystvennaya Biologiya = Agricultural Biology.* 2022;57(5):945-953. DOI 10.15389/agrobiology.2022.5.945rus. (in Russian)

Arkhestova D.Kh., Efremov G.I., Appaev S.P., Kochieva E.Z., Shchennikova A.V. A new 5′-UTR *LcyE* allele correlates with increased expression of the lycopene-ε-cyclase gene determining the flow of the β-ε branch of the carotenoid biosynthesis pathway in maize. *Genetika = Genetics.* 2023;59(4):417-424. DOI 10.31857/S0016675823030025. (in Russian)

Babu R., Rojas N.P., Gao S., Yan J., Pixley K. Validation of the effects of molecular marker polymorphisms in *LcyE* and *CrtRB1* on provitamin A concentrations for 26 tropical maize populations. *Theor. Appl. Genet.* 2013;126:389-399. DOI 10.1007/s00122-012-1987-3.

Bai L., Kim E.H., DellaPenna D., Brutnell T.P. Novel lycopene epsilon cyclase activities in maize revealed through perturbation of carotenoid biosynthesis. *Plant J.* 2009;59(4):588-599. DOI 10.1111/j.1365-313X.2009.03899.x.

Cabiddu A., Delgadillo-Puga C., Decandia M., Molle A.G. Extensive ruminant production systems and milk quality with emphasis on unsaturated fatty acids, volatile compounds, antioxidant protection degree and phenol content. *Animals* (*Basel*). 2019;9(10):771. DOI 10.3390/ani9100771.

Cutler S.R., Rodriguez P.L., Finkelstein R.R., Abrams S.R. Abscisic acid: emergence of a core signaling network. *Annu. Rev. Plant Biol.* 2010;61:651-679. DOI 10.1146/annurev-arplant-042809-112122.

Dhar M.K., Mishra S., Bhat A., Chib S., Kaul S. Plant carotenoid cleavage oxygenases: structure-function relationships and role in development and metabolism. *Brief. Funct. Genomics.* 2020;19(1):1-9. DOI 10.1093/bfgp/elz037.

Filyushin M.A., Shchennikova A.V., Kochieva E.Z. Anthocyanin content in the *Capsicum* species fruits correlates with expression levels of flavonoid pathway structural and regulatory genes. *Russ. J. Plant Physiol.* 2023;70:4. DOI 10.1134/S1021443722602324.

Graulet B., Cirié C., Martin B. Contrasted effects of dietary extruded linseed supplementation on carotenoid and liposoluble vitamin status in lactating Holstein or Montbéliarde cows fed hay or corn silage. *J. Dairy Sci.* 2019;102(7):6210-6225. DOI 10.3168/jds.2018-16138.

Harjes C.E., Rocheford T.R., Bai L., Brutnell T.P., Kandianis C.B., Sowinski S.G., Stapleton A.E., Vallabhaneni R., Williams M., Wurtzel E.T., Yan J., Buckler E.S. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science.* 2008;319(5861):330-333. DOI 10.1126/science.1150255.

Huang X., Shi H., Hu Z., Liu A., Amombo E., Chen L., Fu J. ABA is involved in regulation of cold stress response in Bermudagrass. *Front. Plant Sci.* 2017;8:1613. DOI 10.3389/fpls.2017.01613.

Jahns P., Holzwarth A.R. The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochim. Biophys. Acta*. 2012;1817(1):182-193. DOI 10.1016/j.bbabio.2011.04.012.

Kumar S., Stecher G., Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol*. 2016;33(7):1870-1874. DOI 10.1093/molbev/msw054.

Langridge P., Fleury D. Making the most of 'omics' for crop breeding. *Trends Biotechnol.* 2011;29(1):33-40. DOI 10.1016/j.tibtech.2010.09.006.

LaPorte M.F., Vachev M., Fenn M., Diepenbrock C. Simultaneous dissection of grain carotenoid levels and kernel color in biparental maize populations with yellow-to-orange grain. *G3: Genes Genomes Genetics* (*Bethesda*). 2022;12(3):jkac006. DOI 10.1093/g3journal/jkac006.

Li F., Vallabhaneni R., Yu J., Rocheford T., Wurtzel E.T. The maize phytoene synthase gene family: overlapping roles for carotenogenesis in endosperm, photomorphogenesis, and thermal stress tolerance. *Plant Physiol*. 2008;147(3):1334-1346. DOI 10.1104/pp.108.122119.

Liu L., Jeffers D., Zhang Y., Ding M., Chen W., Kang M.S., Fan X. Introgression of the *crtRB1* gene into quality protein maize inbred lines using molecular markers. *Mol. Breed.* 2015;35(8):154. DOI 10.1007/s11032-015-0349-7.

López-Ráez J.A., Kohlen W., Charnikhova T., Mulder P., Undas A.K., Sergeant M.J., Verstappen F., Bugg T.D.H., Thompson A.J., Ruyter-Spira C., Bouwmeester H. Does abscisic acid affect strigolactone biosynthesis? *New Phytol*. 2010;187(2):343-354. DOI 10.1111/j.1469-8137.2010.03291.x.

Medina-Lozano I., Díaz A. Applications of genomic tools in plant breeding: crop biofortification. *Int. J. Mol. Sci*. 2022;23(6):3086. DOI 10.3390/ijms23063086.

Menkir A., Maziya-Dixon B.B., Mengesha W., Rocheford T., Alamu E.O. Accruing genetic gain in pro-vitamin A enrichment from harnessing diverse maize germplasm. *Euphytica.* 2017;213(5):105. DOI 10.1007/s10681-017-1890-8.

Mitani T., Kobayashi K., Ueda K., Kondo S. Regional differences in the fatty acid composition, and vitamin and carotenoid concentra-

tions in farm bulk milk in Hokkaido, Japan. *J. Anim. Sci*. 2021;92(1): e13570. DOI 10.1111/asj.13570.

Muthusamy V., Hossain F., Thirunavukkarasu N., Choudhary M., Saha S., Bhat J.S., Prasanna B.M., Gupta H.S. Development of β-carotene rich maize hybrids through marker-assisted introgression of β-carotene hydroxylase allele. *PLoS One.* 2014;9(12):e11583. DOI 10.1371/journal.pone.0113583.

Nambara E., Marion-Poll A. Abscisic acid biosynthesis and catabolism. *Annu. Rev. Plant Biol.* 2005;56:165-185. DOI 10.1146/annurev.arplant.56.032604.144046.

Orlovskaya O.A., Vakula S.I., Khotyleva L.V., Kilchevsky A.V. Assotiation of total carotenoid level in maize grain (*Zea mays* L.) with polimorphic site InDel1 in *PSY1* gene. *Ekologicheskaya Genetika = Ecological Genetics.* 2016;14(3):28-34. DOI 10.17816/ecogen14328-34. (in Russian)

Pasala R., Pandey B.B. Plant phenomics: high-throughput technology for accelerating genomics. *J. Biosci.* 2020;45:111. DOI 10.1007/s12038-020-00083-w.

Pixley K., Rojas N.P., Babu R., Mutale R., Surles R., Simpungwe E. Biofortification of maize with provitamin A carotenoids. In: Tanumihardjo S.A. (Ed.) Carotenoids and Human Health. Nutrition and Health. Totowa, NJ: Humana Press, 2013;271-292. DOI 10.1007/978-1-62703-203-2_17.

Prasanna B.M., Palacios-Rojas N., Hossain F., Muthusamy V., Menkir A., Dhliwayo T., Ndhlela T., San Vicente F., Nair S.K., Vivek B.S., Zhang X., Olsen M., Fan X. Molecular breeding for nutritionally enriched maize: status and prospects. *Front. Genet.* 2020; 10:1392. DOI 10.3389/fgene.2019.01392.

Rosas-Saavedra C., Stange C. Biosynthesis of carotenoids in plants: enzymes and color. In: Stange C. (Ed.) Carotenoids in Nature. Subcellular Biochemistry. Vol. 79. Cham: Springer, 2016;35-69. DOI 10.1007/ 978-3-319-39126-7_2.

Suwarno W.B., Hannok P., Palacios-Rojas N., Windham G., Crossa J., Pixley K.V. Provitamin A carotenoids in grain reduce aflatoxin contamination of maize while combating vitamin A deficiency. *Front. Plant Sci*. 2019;10:30. DOI 10.3389/fpls.2019.00030.

Trono D. Carotenoids in cereal food crops: composition and retention throughout grain storage and food processing. *Plants* (*Basel*). 2019; 8(12):551. DOI 10.3390/plants8120551.

Yan J., Kandianis B.C., Harjes E.C., Bai L., Kim E.H., Yang X., Skinner D.J., Fu Z., Mitchell S., Li Q., Fernandez M.G., Zaharieva M., Babu R., Fu Y., Palacios N., Li J., Dellapenna D., Brutnell T., Buckler E.S., Warburton M.L., Rocheford T. Rare genetic variation at *Zea mays crtRB1* increases beta carotene in maize grain. *Nat. Genet.* 2010;42(4):322-327. DOI 10.1038/ng.551.

Yue K., Lingling L., Xie J., Coulter J.A., Luo Z. Synthesis and regulation of auxin and abscisic acid in maize. *Plant Signal. Behav.* 2021; 16(7):1891756. DOI 10.1080/15592324.2021.1891756.

Zunjare R.U., Chhabra R., Hossain F., Hossain F., Baveja A., Muthusamy V., Gupta H.S. Molecular characterization of 5′ UTR of the *lycopene epsilon cyclase* (*lcyE*) gene among exotic and indigenous inbreds for its utilization in maize biofortification. *3 Biotech*. 2018; 8(1):75. DOI 10.1007/s13205-018-1100-y.

ORCID ID

D.Kh. Arkhestova orcid.org/0000-0003-1239-3641
A.V. Shchennikova orcid.org/0000-0003-4692-3727
E.Z. Kochieva orcid.org/0000-0002-6091-0765

Acknowledgements. The study was supported by the Federal Scientific and Technical Program for the Development of Agriculture of the Russian Federation.

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

Received February 1, 2023. Revised May 12, 2023. Accepted May 12, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):447-453
DOI 10.18699/VJGB-23-54

Seedling and adult plant resistance to leaf rust in some Bulgarian common wheat lines

V. Ivanova

Agricultural Academy, Dobrudzha Agricultural Institute, General Toshevo, Republic of Bulgaria
vkiryakova@yahoo.com

Abstract. The response of 250 common winter wheat breeding lines was investigated for resistance to the causative agent of *Puccinia triticina* under conditions of an infected field on the territory of Dobrudzha Agricultural Institute – General Toshevo, Bulgaria, during three successive seasons. Twenty lines with different degrees of resistance under field conditions were selected. Multi-pathotype testing was used to study the response of these lines at seedling stage under greenhouse conditions to individual pathotypes of *P. triticina*. Based on the response of the lines at seedling and adult stages, we found out that 20 % of them carried race-specific resistance. One of the lines (99/08-52) reacted with full resistance to the pathotypes used under greenhouse conditions. The reaction demonstrated by this line coincided with the response of isogenic lines carrying the genes *Lr9*, *Lr19*, *Lr22a*, *Lr22b* and *Lr25*. The other three lines (19/06-108, 82/08-43 and 82/08-35) showed a resistant reaction to 6 or 5 of the pathotypes used in the study. Their response partially coincided with the reaction of 5 isogenic lines, and the presence of some of these genes in the above lines is quite possible. Lines carrying this type of resistance are to be subjected to further genetic and breeding investigations to prove the presence of a race-specific gene. Twenty-five percent of the lines combined partial race-specific resistance at seedling stage with the resistance of race non-specific nature at adult stage. Forty percent of all studied lines carried race non-specific resistance, and 15 % of the lines possessed resistance of the "*slow rusting*" type. As a result of the study we carried out, the lines that demonstrated stable resistance to leaf rust can provide sufficient protection of the host and can be included in the breeding programs for developing varieties resistant to *P. triticina*.
Key words: wheat; *Puccinia triticina*; pathotypes; adult and juvenile resistance.

For citation: Ivanova V. Seedling and adult plant resistance to leaf rust in some Bulgarian common wheat lines. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):447-453. DOI 10.18699/VJGB-23-54

Устойчивость проростков и взрослых растений к бурой ржавчине у некоторых линий болгарской мягкой пшеницы

В. Иванова

Сельскохозяйственная академия, Добруджанский сельскохозяйственный институт, Генерал Тошево, Республика Болгария
vkiryakova@yahoo.com

Аннотация. Проверена реакция 250 селекционных линий мягкой озимой пшеницы на устойчивость к возбудителю *Puccinia triticina* в условиях инфекционного поля на территории Добруджанского сельскохозяйственного института (Генерал Тошево, Болгария) в течение трех сезонов. Отобрано 20 линий с разной степенью устойчивости в полевых условиях. Для изучения реакции этих линий на стадии проростков в тепличных условиях на отдельные патотипы *P. triticina* использовали мультипатотипное тестирование. По реакции линий на стадии проростков и взрослых растений установлено, что 20 % из них обладают расоспецифической устойчивостью. Одна из линий (99/08-52) проявила полную устойчивость к патотипам, используемым в тепличных условиях. Ее реакция совпала с реакцией изогенных линий – носителей генов *Lr9*, *Lr19*, *Lr22a*, *Lr22b* и *Lr25*. Еще три линии (19/06-108, 82/08-43 и 82/08-35) показали резистентную реакцию на шесть или пять патотипов, использованных в исследовании. Их реакция частично совпала с реакцией пяти изогенных линий. Вероятно, некоторые из этих генов присутствуют у перечисленных выше линий. Линии-носители данного типа устойчивости подлежат дальнейшим генетико-селекционным исследованиям для подтверждения наличия расоспецифического гена. У 25 % линий частичная расоспецифическая устойчивость на стадии проростков сочеталась с устойчивостью расовой неспецифической природы на взрослой стадии. Из всех изученных линий 40 % обладали расовой неспецифической устойчивостью, а 15 % – устойчивостью типа *slow rusting*. По результатам наших исследований, линии, показавшие стабильную устойчивость к бурой ржавчине, могут быть включены в селекционные программы по созданию сортов, устойчивых к *P. triticina*.
Ключевые слова: пшеница; *Puccinia triticina*; патотипы; ювенильная и возрастная устойчивость.

Introduction

Leaf rust is one of the most widespread diseases on wheat in Bulgaria and one of the most important diseases in those parts of the world, where wheat is the main cereal crop. The development and growing of resistant cultivars is an important, efficient, environmentally friendly and cost-effective method for control of the disease (Bariana, 2003; Bariana et al., 2007; Singh et al., 2016; Volkova et al., 2020; Kokhmetova et al., 2021). In order to avoid the danger of epiphytotic occurrence, it is necessary to have at our disposal a large number of sources – carriers of different genes or types of resistance, which should be properly alternated in the production fields (Donchev et al., 1977). According to Van der Plank (1963), the resistance can be categorized into two classes based on the genetic control and the phenotypic effect – race specific (vertical) and race non-specific (horizontal).

The specific resistance is determined by one or several genes acting independently of one another and is efficient to individual races of the pathogen (Roelfs et al., 1992). Each gene ensures resistance to all races that do not have a respective gene for virulence, but not to races that do not possess such a gene. When this resistance is realized in a widely distributed cultivar, high selection pressure on the pathogen occurs, leading to the formation of new races with new genes for virulence, i. e. this type of resistance quickly loses its efficiency, because the pathogen population evolves (Huerta-Espino et al., 2011, 2014; Lowe et al., 2011; Ellis et al., 2014).

The non-specific (horizontal) resistance ensures protection of the plants against all races of the pathogen and the genes, which determine it, have additive effect. The polygenic nature of this type of resistance is the reason for its durability (Parlevliet, Zadoks, 1977; Singh et al., 2011). It is expressed at adult stage and its mechanism consists in reduction in the amount and rate of the disease (Stubbs et al., 1986; Li et al., 2014). The impact of the qualitative resistance of the host on the evolution of the pathogen populations is less documented in the literature (Volkova et al., 2020). It has been shown that the fungal pathogens may evolve and adapt to qualitative resistance through breeding for higher aggressiveness (Delmas et al., 2016; Frézal et al., 2018).

Especially interesting is the resistance of a non-specific nature – the "*slow rusting*" type, or the retarded development of the pathogen. The cultivars possessing this type of resistance allow the pathogen to sporulate on them, to attack them to a moderate degree, without forcing the pathogen to develop new more aggressive races (Knott, 1989; Kolmer, 2013; Singh et al., 2016). The genes determining this type of resistance are related to such factors as pustule size, infection frequency, latent period, and are most often defined as "*slow rusting genes*" (Caldwell, 1968; Kolmer, 1996; Ellis at al., 2014). Although the genes determining adult stage resistance are considered to determine durable resistance, some authors point out that the occurrence of new and aggressive races of the pathogen may make these genes inefficient (Singh, Rajaram, 1992; Park, McIntosh, 1994; Huerta-Espino, Singh, 1996). This is the reason why it is necessary to search for and develop new sources of resistance (Pathan, Park, 2006; Ivanova, 2015; Ivanova, Chamurliisky, 2017; Ivanova et al., 2019a, b). Hussain et al. (1999) concluded that durable rust resistance mechanism in wheat is achieved through incorporation of partially resistant minor genes, which seems to be more appropriate for sustainable wheat production.

The method based on the "gene for gene" relation is one of the fundamental concepts of the relationship between the plants and the pathogens (Flor, 1956). Based on this hypothesis, Person (1959) developed a method for identification of genes for resistance with the help of testing races of the pathogen, which carry certain virulence. The multipathotype test used for determining the sources of resistance at seedling stage by comparing the response of the tested sources to the reaction of the isogenic lines allows investigating a large number of sources and the obtained information can be used for the development of resistant cultivars. The gene postulation, determined through the multipathotype test, is the most widely applied method worldwide for proving the presence of race specificity and for identification of certain *Lr* genes in different wheat populations (Statler, 1984; Modawi et al., 1985; Singh, Gupta, 1991; Singh, 1993; Singh et al., 1999; Oelke, Kolmer, 2004; Gebrewahid et al., 2017; Yan et al., 2017; Zhang et al., 2019; Wu et al., 2020). The multipathotype test, however, has certain shortcomings. Kadkhodaei et al. (2012) pointed out that the identification of *Lr* genes is rather labor and time consuming. Furthermore, there may be no available pathotypes suitable for identification of the genes for resistance present in the genotypes, or the pathotype may not be able to detect the genes for resistance to rust.

Starting from these premises and estimating the difficulties and disadvantages of the use of the multipathotype test, our investigation, too, could not achieve complete and thorough identification of a gene, but only a suggestion; on the basis of the response of these lines at seedling and adult stages, however, the nature of the resistance was determined, which also provides valuable data that can aid the breeding and improvement work for development of cultivars resistant to the disease.

The aim of this investigation was to study the response of common winter wheat lines both at seeding and adult stages and to use the obtained data on stable resistance present in these lines to aid the breeding for development of cultivars resistant to *Puccinia triticina*.

Materials and methods

In the infection field of Dobrudzha Agricultural Institute – General Toshevo, Bulgaria, the reaction to leaf rust (*P. triticina*) of 250 lines of common winter wheat involved in a competitive varietal trial was studied. From the investigated breeding material, 20 lines were selected, which responded with a certain degree of resistance from moderate to high (MR–VR), and which demonstrated resistant reaction to some of the used pathotypes of *P. triticina* at seedling stage under greenhouse conditions.

Seedling test. The selected 20 lines were tested for resistance to single pathotypes of *P. triticina* and their response was compared to the reaction of a set of 34 differential lines (isogenic lines developed on the basis of cultivar Tatcher and each carrying one of the already identified *Lr* genes) according to the 7 pathotypes used in the study, which possessed different virulence (13763, 33762, 43773, 53723, 53762, 73762, and

73763). The pathotypes used in the test were identified on the basis of the reaction of 15 monogenic lines, *Lr1*, *Lr2a*, *Lr2b*, *Lr2c*, *Lr3*, *Lr9*, *Lr11*, *Lr15*, *Lr17*, *Lr19*, *Lr21*, *Lr23*, *Lr24*, *Lr26* and *Lr28*, coded for by the method of Limpert and Müller (1994). Avirulence/virulence profiles of *P. triticina* pathotypes are present in Table 1.

The inoculated plants were placed in the dark in a moist chamber at temperature 18–20 °C and 100 % relative air humidity. After 24 hours at these conditions, they were transferred to a greenhouse for further growing under controlled conditions: 20–25 °C (day) and 15 °C (night), more than 75 % relative air humidity and 30,000 lx light intensity, for elongation of the photo period – 16 h (day) and 8 h (night).

In order to improve sporulation, the plants were treated with maleic hydrazide 97 % solution (1 g in 3 l water). On the 9–12th day after inoculation, the type of reaction (R) was read according to the scale of Stakman et al. (1962).

Infection types 0, 0;, 1 and 2 were considered expression of a resistant type of reaction (R), while infection types 3, 4 and X were considered susceptible (S) while estimating the disease.

Adult plant test. The investigation was carried out under conditions of a maximum infection created in the field, where the full set of pathotypes identified for the respective year were taken out. The lines were planted manually in 1.5 m wide rows with 0.25 m interspacing, in two replications. Cultivar *Michigan amber* was used as a multiplier and distributor of leaf rust. Spreader rows of *M. amber* were planted perpendicular and adjacent to the test rows. The artificial inoculation with the pathogen was done according to the methodology for working with rusts adopted at the Plant Pathology Laboratory of Dobrudzha Agricultural Institute (Ivanova, 2012). Nine-day old seedlings from the standard susceptible cultivar *M. amber* inoculated with different pathotypes of *P. triticina* were planted in the rows of the spreader cultivar in March and April till the final accumulation of inoculum in June, when the maximum was reached. The type of infection and the attacking rate were read according to the scale of Cobb, modified by Peterson (Peterson et al., 1948) at milk maturity stage. The average coefficient of infection (ACI), or the so called corrected relative attack rate (P_0), was calculated by introducing a coefficient for the respective infection types (R – 0.2; MR – 0.4; M – 0.6; MS – 0.8; S – 1). Depending on the values of ACI, the studied lines were divided into several groups: immune (ACI = 0); very resistant, VR (ACI = 0–5.99); resistant, R (ACI = 6–25.99); moderately resistant, MR (ACI = 26–45.99); moderately susceptible, MS (ACI = 46–65.99); susceptible, S (ACI = 66–100). The lines with susceptible reaction were of no interest to us.

Table 1. Avirulence/virulence profiles of *Puccinia triticina* pathotypes

Race/ Pathotypes	Avirulence/virulence spectrum (effective *Lr* genes) / (ineffective *Lr* genes)
13763	*Lr2a, 2b, 9, 19, 28 / Lr1, 2c, 3, 11, 15, 17, 21, 23, 24, 26*
33762	*Lr2b, 9, 19, 24, 28 / Lr1, 2a, 2c, 3, 11, 15, 17, 21, 23, 26*
43773	*Lr1, 2a, 9, 19, 28 / Lr2b, 2c, 3, 11, 15, 17, 21, 23, 24, 26*
53723	*Lr2a, 9, 19, 23, 28 / Lr1, 2b, 2c, 3, 11, 15, 17, 21, 24, 26*
53762	*Lr2a, 9, 19, 24, 28 / Lr1, 2b, 2c, 3, 11, 15, 17, 21, 23, 26*
73762	*Lr9, 19, 24, 28 / Lr1, 2a, 2b, 2c, 3, 11, 15, 17, 21, 23, 26*
73763	*Lr9, 19, 28 / Lr1, 2a, 2b, 2c, 3, 11, 15, 17, 21, 23, 24, 26*

Results and discussion

The experiment was carried out in three successive vegetative growth seasons. Out of the investigated 250 common winter wheat lines, 20 lines were selected, which responded with high to moderate resistance over the years of study. The lines responding with MR probably carry *slow-rusting* genes. According to Morgunov et al. (2010), some genes with *slow-rusting* effect have a moderately susceptible type of infection but their attack rate does not exceed 50 %. The response of the lines investigated under field conditions is presented in Table 2, and the reaction of the lines at seedling stage to seven separate *P. triticina* pathotypes of different virulence is given in Table 3. The results of the investigation revealed the following.

Line 60/05-49 at seedling stage exhibited a resistant reaction to four phenotypically different pathotypes (see Table 3), and the field evaluation showed that this line had a resistant to very resistant reaction (see Table 2). This allows us to comment that the line is a carrier of partial race-specific resistance in combination with resistance of non-specific nature, but the race-specific resistance has to be checked at a later stage.

Line 15/05-82 demonstrated a susceptible reaction to all pathotypes of the pathogen used under greenhouse conditions, and the field evaluation showed that the line responded with a very resistant to resistant reaction. According to this reaction exhibited at seedling and adult stages, the line can be defined as a carrier of adult or field resistance.

Line 60/05-68, also at seedling stage, responded with a susceptible reaction to all pathotypes used in the study, and the field evaluation showed a resistant to moderately resistant reaction. The response of the line allowed referring it to the group of the carriers of the *slow rusting* type of resistance.

Line 20/05-120 at seedling stage responded with a susceptible reaction to all used pathotypes, and in the field it exhibited resistance of the type (VR–R–MR), but judging by the reaction, this line can be referred to the group of lines carrying resistance of race non-specific nature.

Line 98/05-95 at seedling stage demonstrated a resistant reaction to two pathotypes (33762 and 53762), and the field evaluation was not constant; in 2014, when the attack on cultivar *M. amber* was even higher in comparison to the other two years, the lines responded as moderately susceptible. In 2015 and 2016, the line demonstrated a resistant to moderately resistant reaction. This line carried resistance of race non-specific nature.

Line 223/05-2 responded with a resistant reaction to three pathotypes under greenhouse conditions and with complete resistance at adult stage. The line was a carrier of partial race-specific resistance in combination with race non-specific one.

Line 13/08-87 at seedling stage responded with a resistant reaction to only one pathotype (53762), and its field resistance was of the VR–R type. The line was a carrier of resistance of race non-specific nature.

Table 2. Adult plant resistance

Cultivar/ Lines	2014			2015			2016		
	Final rust severity	ACI	Rating	Final rust severity	ACI	Rating	Final rust severity	ACI	Rating
60/05-49	10/4	12.5	R	0	0	VR	5/4	7.1	R
15/05-82	15/4	18.8	R	0	0	VR	0	0	VR
60/05-68	25/4	31.3	MR	25/4	37.6	MR	10/4	14.3	R
20/05-120	0	0	VR	25/4	37.6	MR	5/4	7.1	R
98/05-95	40/4	50.0	MS	25/4	37.6	MR	5/4	7.1	R
223/05-2	0	0	VR	0	0	VR	0	0	VR
19/06-108	10/4	12.5	R	5/4	7.1	R	5/4	7.1	R
13/08-87	0	0	VR	10/4	14.3	R	0	0	VR
44/08-66	25/4	31.3	MR	25/4	37.6	MR	5/4	7.1	R
19/08-28	0	0	VR	0	0	VR	0	0	VR
18/08-16	5/4	6.3	R	10/4	14.3	R	0	0	VR
14/08-57	10/4	12.5	R	5/4	7.1	R	5/4	7.1	R
44/08-88	40/4	50.0	MS	5/4	7.1	R	5/4	7.1	R
46/08-27	5/4	6.3	R	5/4	7.1	R	5/4	7.1	R
79/08-10	0	0	VR	5/4	7.1	R	0	0	VR
72/08-23	0	0	VR	5/4	7.1	R	5/4	7.1	R
82/08-35	5/4	6.3	R	0	0	VR	5/4	7.1	R
82/08-43	5/4	6.3	R	5/4	7.1	R	10/4	14.3	R
90/08-22	0	0	VR	5/4	7.1	R	0	0	VR
99/08-52	25/4	31.3	MR	5/4	7.1	R	10/4	14.3	R
M. amber	80/4	100	VS	70/4	100	VS	70/4	100	VS

Table 3. Response of common winter wheat lines to 7 pathotypes of *P. triticina* at seedling stage

Cultivar/ Lines	*P. triticina* pathotypes						
	13763	33762	43773	53762	73762	53723	73763
60/05-49	R	R	S	S	R	R	S
15/05-82	S	S	S	S	S	S	S
60/05-68	S	S	S	S	S	S	S
20/05-120	S	S	S	S	S	S	S
98/05-95	S	R	S	R	S	S	S
223/05-2	S	R	R	S	S	R	S
19/06-108	R	R	R	R	R	R	S
13/08-87	S	S	S	R	S	S	S
44/08-66	S	S	R	S	S	S	R
19/08-28	S	R	R	S	S	R	S
18/08-16	S	R	S	S	R	R	S
14/08-57	S	R	S	R	S	S	S
44/08-88	S	S	S	S	S	S	S
46/08-27	S	S	S	S	S	R	S
79/08-10	S	R	S	S	S	R	S
72/08-23	R	R	S	R	S	R	S
82/08-35	S	R	R	R	S	R	R
82/08-43	S	R	R	R	S	R	R
90/08-22	S	S	S	S	S	R	S
99/08-52	R	R	R	R	R	R	R
M. amber	S	S	S	S	S	S	S

Line 44/08-66 responded with a resistant reaction at seedling stage to two pathotypes (43773 and 73763), and during two of the years it demonstrated field resistance of the MR type. Based on the response of the line, it was referred to the group of lines of the *slow rusting* type.

Line 18/08-16 as well as line 19/08-28 responded with a resistant reaction at seedling stage to three of the pathotypes and demonstrated R to VR under field conditions, allowing us to refer them to the group of lines combining partial race-specific resistance with race non-specific one. The combination of these two types of resistance in a single genotype is a good solution for breeding since the host is protected against diseases during the entire vegetative growth season.

Line 14/08-57 demonstrated stable field resistance and a resistant reaction to two of the pathotypes under greenhouse conditions. The line was a carrier of race non-specific resistance.

Line 46/08-27 reacted with stable resistance in the field during the three years of testing, and at seedling stage, it demonstrated a resistant reaction to only one pathotype. The line carries resistance of race non-specific nature.

Line 79/08-10 responded at seedling stage with a resistant reaction to two pathotypes (33762 and 53723), and in the field it demonstrated a resistant to very resistant reaction. The line is a carrier of race non-specific resistance.

Line 72/08-23 showed a resistant to very resistant reaction in the field, and under greenhouse conditions, a resistant reaction to four pathotypes was registered. The line is a carrier of partial race-specific resistance in combination with race non-specific one.

Table 4. Comparative reaction between four breeding lines and the isogenic lines carrying genes *Lr9*, *Lr19*, *Lr22a*, *Lr22b* and *Lr25*, determining resistance to pathotypes of leaf rust

Lr gene/Line	Year	13763	33762	43773	53762	73762	53723	73763
Lr22a	2014	R	R	R	R	R	R	R
Lr22b	2014	R	R	R	R	R	R	R
Lr25	2014	R	R	R	R	R	R	R
Lr9	2015–2016	R	R	R	R	R	R	R
Lr19	2015–2016	R	R	R	R	R	R	R
99/08-52	2014–2016	R	R	R	R	R	R	R
19/06-108	2014–2016	R	R	R	R	R	R	S
82/08-35	2014–2016	S	R	R	R	S	R	R
82/08-43	2014–2016	S	R	R	R	S	R	R

Line 90/08-22 exhibited a resistant to very resistant reaction in the field, while responding with a resistant reaction to only one pathotype at seedling stage. This line is probably a carrier of adult race specific resistance or resistance of race non-specific nature.

Line 99/08-52 responded with a resistant reaction to all 7 pathotypes used in this study, and the field evaluation was within the range of R–MR. The presence of full resistance at seedling stage was a proof that the line possessed race-specific resistance. Its reaction coincided entirely with the reaction of the isogenic lines carrying genes *Lr9*, *Lr19*, *Lr22a*, *Lr22b* and *Lr25* (Table 4).

According to data provided by Ivanova (2020) and Ivanova et al. (2021), in 2014, with 100 % efficiency against the local population of *P. triticina* there were genes *Lr22a*, *Lr22b* and *Lr25*, and in 2015 and 2016, also with 100 % efficiency, genes *Lr9* and *Lr19* were registered. Potentially, any of these genes could be present, determining the resistance of these lines to *P. triticina* (see Table 4). For greater precision, we recommend conducting breeding and genetic studies in order to prove the presence of some of these genes in the above line. Another three lines showed a similar reaction: 19/06-108 responded with susceptibility to pathotype 73763, and lines 82/08-35 and 82/08-43 demonstrated a susceptible reaction to pathotypes 13763 and 73762. To all other pathotypoes, the lines responded with resistance.

Since the reaction of the lines partially coincided with the reaction of the above mentioned isogenic lines, it can be suggested that these breeding lines could also carry some of these genes (see Table 4).

Conclusion

As a result of this study, we identified the following types of resistance in the investigated lines:

- Lines with race-specific resistance, which are to be subjected to breeding and genetic studies to prove the presence of the race-specific gene – four of the lines probably carried this type of resistance: 19/06-108, 99/08-52, 82/08-35 and 82/08-43. They constituted 20 % of all investigated lines.
- Lines combining partial race-specific resistance at seedling stage with resistance of race non-specific nature at adult stage. The combination of race-specific with race non-specific resistance is a good possibility to protect the host against the disease during the entire vegetative growth. The lines that fall in this group are 60/05-49, 223/05-2, 19/08-28, 18/08-16, 72/08-23. They constituted 25 % of the investigated lines.
- Lines-carriers of race non-specific resistance. The non-specific nature of resistance is determined by the fact that at adult age, the host is resistant to all races and in this case the resistance is determined by 4 or 5 small genes with additive effect. Lines 15/05-82, 98/05-95, 14/08-57, 46/08-27, 79/08-10, 90/08-22, 44/08-88 and 13/08-87 fell in this group. They constituted 40 % of all studied lines.
- Lines-carriers of the *slow rusting* type of resistance: lines 20/05-120, 44/08-66 and 60/05-68 belonged to this group and they constituted 15 % of the investigated material. The partial resistance is more durable than the resistance conditioned by single main genes since it is inherited polygenically (Parlevliet, 1985).

The lines studied in this investigation are carriers of certain types of resistance. According to Volkova et al. (2020), the cultivars with race-specific resistance are applied as a mosaic of varieties with subsequent alternation over time and space, and the cultivars that carry non-specific resistance can be used on large areas for a longer period of time in combination with cultivars from different groups, including their own.

In this relation, the studied lines carrying race-specific or race non-specific resistance can be included in the breeding programs for developing resistant cultivars in order to avoid large yield losses caused by the disease.

References

Bariana H.S. Breeding for disease resistance. In: Thomas B., Murphy D.J., Murray B.G. (Eds.) Encyclopedia of Applied Plant Sciences. Harcourt, Acad. Press, 2003;244-253.

Bariana H.S., Brown G.N., Bansal U.K., Miah H., Standen G.E., Lu M. Breeding triple rust resistant wheat cultivars for Australia using conventional and marker-assisted selection technologies. *Aust. J. Agric. Res*. 2007;58(6):576-587. DOI 10.1071/AR07124.

Caldwell R.M. Breeding for general and/or specific plant disease resistance. In: Finlay K.W., Shepherd K.W. (Eds.) Proceeding of the Third International Wheat Genetics Symposium. Canberra (Australia), Australian Academy of Sciences, 1968;263-272.

Delmas C.E.L., Fabre F., Jolivet J., Mazet I.D., Cervera S.R., Delière L., Delmotte F. Adaptation of a plant pathogen to partial host resistance: selection for greater aggressiveness in grapevine downy mildew. *Evol. Appl*. 2016;9(5):709-725. DOI 10.1111/eva.12368.

Donchev N., Todorov I., Malinski K. Immunological characterization of some common wheat cultivars. In: Breeding and Agrotechnical of Wheat and Sunflower. Praha, 1977;81-96. (in Bulgarian)

Ellis J.G., Lagudah E.S., Spielmeyer W., Dodds P.N. Past, present and future of breeding rust-resistance wheat. *Front. Plant Sci*. 2014;5: 641. DOI 10.3389/fpls.2014.00641.

Flor H.H. The complementary genetics system in flax and flax rust. *Adv. Genet*. 1956;8:29-54. DOI 10.1016/S0065-2660(08)60498-8.

Frézal L., Jacqua G., Neema C. Adaptation of a fungal pathogen to host quantitative resistance. *Front. Plant Sci.* 2018;9:1554. DOI 10.3389/fpls.2018.01554.

Gebrewahid T.W., Yao Z.-J., Yan X.-C., Gao P., Li Z.-F. Identification of leaf rust resistance genes in Chinese common wheat cultivars. *Plant Dis*. 2017;101(10):1729-1737. DOI 10.1094/PDIS-02-17-0247-RE.

Huerta-Espino J., Singh R.P. Misconceptions on the durability of some adult plant resistance genes in wheat. In: Proceeding 9th European and Mediterranean Cereal Rust and Powdery Mildew Conf. Lunteren. The Netherlands, 1996;109-111.

Huerta-Espino J., Singh R.P., Germán S., McCallum B.D., Park R.F., Chen W.Q., Bhardwaj S.C., Goyeau H. Global status of wheat leaf rust caused by *Puccinia triticina. Euphytica*. 2011;179:143-160. DOI 10.1007/s10681-011-0361-x.

Huerta-Espino J., Singh R.P., Roelfs A.P. Rusts fungi of wheat. In: Misra J.K., Tewari J.P., Deshmukh S.K., Vágvölgyi C. (Eds.) Fungi from Different Substrates. Florida (USA): CRC Press, 2014;217-259. DOI 10.1201/b17646.

Hussain M.H., Chaudhry A., Rehman J., Anwar S.B. Development of durable rust resistance in wheat. *Pakistan J. Phytopathol*. 1999; 11:130-139.

Ivanova V. Studies on the resistance of common wheat and other species to the cause agent of brown rust *Puccinia triticina* Erikss. PhD thesis. General-Toshevo: Dobrudzha Agricultural Institute, 2012. (in Bulgarian)

Ivanova V. New common winter wheat lines with resistance to leaf rust (*P. triticina*). In: VI International Agricultural Symposium "Agrosym 2015", 15–18 October, 2015. Jahorina (Bosnia and Herzegovina), 2015;169-175.

Ivanova V. Monitoring on the pathogen population of *P. triticina* in Bulgaria during 2013–2014. *Int. J. Innov. Approaches Agric. Res*. 2020;4(3):353-365. DOI 10.29329/ijiaar.2020.274.8.

Ivanova V., Chamurliisky P. Resistance of common winter wheat cultivars from different geographic origin to *P. triticina*. In: 52nd Croatian and 12th International Symposium on Agriculture, 12–17 February. Dubrovnik, 2017;216-221.

Ivanova V., Mihova G., Ivanova B. New common winter wheat lines with resistance to leaf rust (*P. triticina*). *Int. J. Innov. Approaches Agric. Res*. 2019a;3(3):420-428. DOI 10.29329/ijiaar.2019.206.8.

Ivanova V., Mihova G., Nikolova D., Velikova S. Screening of wheat genotypes for resistance to brown rust (*P. triticina*). *Field Crop Studies*. 2019b;XII(2):67-82. (in Bulgarian)

Ivanova V., Velikova S., Nikolova D. Physiologic specialization of *Puccinia triticina* Erikss. and effectiveness of *Lr* genes in Bulgaria. *Bulg. J. Agric. Sci*. 2021;27(3):555-561.

Kadkhodaei M., Dadkhodaie A., Assad M.T., Heidari B., Mostowfizadeh-Chalamfarsa R. Identification of the leaf rust resistance genes *Lr9*, *Lr26*, *Lr28*, *Lr34* and *Lr35* in a collection of Iranian wheat genotypes using STS and SCAR markers. *J. Crop Sci. Biotech*. 2012; 15(4):267-274. DOI 10.1007/s12892-012-0035-9.

Knott D.R. The Wheat Rusts – Breeding for Resistance. Berlin; Heidelberg: Springer-Verlag, 1989. DOI 10.1007/978-3-642-83641-1.

Kokhmetova A., Rsaliyev S., Atishova M., Kumarbayeva M., Malysheva A., Keishilov Zh., Zhanuzak D., Bolatbekova A. Evaluation of wheat germplasm for resistance to leaf rust (*Puccinia triticina*) and identification of the sources of *Lr* resistance genes using molecular markers. *Plants*. 2021;10(7):1484. DOI 10.3390/plants.10071484.

Kolmer J.A. Genetics of resistance to wheat leaf rust. *Annu. Rev. Phytopathol*. 1996;34:435-455. DOI 10.1146/annurev.phyto.34.1.435.

Kolmer J.A. Leaf rust of wheat: pathogen biology, variation and host resistance. *Forests*. 2013;4(1):70-84. DOI 10.3390/f4010070.

Li Z., Lan C., He Z., Singh R.P., Rosewarne G.M., Chen X., Xia X. Overview and application of QTL for adult plant resistance to leaf rust and powdery mildew in wheat. *Crop Sci*. 2014;54(5):1907-1925. DOI 10.2135/cropsci2014.02.0162.

Limpert E., Müller K. Designation of pathotypes of plant pathogens. *J. Phytopathol*. 1994;140(4):346-358. DOI 10.1111/j.1439-0434.1994.tb00617.x.

Lowe I., Cantu D., Dubcovsky J. Durable resistance to the wheat rusts: integrating systems biology and traditional phenotype-based research methods to guide the deployment of resistance genes. *Euphytica*. 2011;179(1):69-79. DOI 10.1007/s10681-010-0311-z.

Modawi R.S., Browder L.L., Heyne E.G. Use of infection type data to identify genes for low reaction to *Puccinia recondita* in several winter wheat cultivars. *Crop Sci*. 1985;25(1):9-13. DOI 10.2135/cropsci1985.0011183X002500010003x.

Morgunov A., Ablova I., Babayants O., Babayants L., Bespalova L., Khudokormov Zh., Litvinenko N., Shamanin V., Synkov V. Genetic protection of wheat from rusts and development of resistant varieties in Russia and Ukraine. In: BGRI-2010 Technical Workshop, 30–31 May 2010, St. Petersburg, Russia. New York, 2010; 1-20.

Oelke L.M., Kolmer J.A. Characterization of leaf rust resistance in hard red spring wheat cultivars. *Plant Dis*. 2004;88(10):1127-1133. DOI 10.1094/PDIS.2004.88.10.1127.

Park R.F., McIntosh R.A. Adult plant resistance to *Puccinia recondita* f. sp. *tritici* in wheat. *N.Z.J. Crop Hortic. Sci*. 1994;22(2):151-158. DOI 10.1080/01140671.1994.9513819.

Parlevliet J.E. Resistance of the non-race-specific type. In: Bushnell W.R., Roelfs A.P. (Eds.) The Cereal Rusts. Vol. II. Diseases, Distribution, Epidemiology, and Control. Orlando, Florida: Acad. Press, 1985;501-525. DOI 10.1016/B978-0-12-148402-6.50024-9.

Parlevliet J.E., Zadoks J.C. The integrated concept of diseases resistance; a new view including horizontal and vertical resistance in plants. *Euphytica*. 1977;26:5-21. DOI 10.1007/BF00032062.

Pathan A., Park R. Evaluation of seedling and adult plant resistance to leaf rust in European wheat cultivars. *Euphytica*. 2006;149:327-342. DOI 10.1007/S10681-005-9081-4.

Person C. Gene for gene relationships in host: parasite systems. *Can. J. Bot.* 1959;37(5):1101-1130. DOI 10.1139/b59-087.

Peterson R.F., Campbell A.B., Hannah A.E. A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. *Can. J. Res*. 1948;26c(5):496-500. DOI 10.1139/cjr48c-033.

Roelfs A.P., Singh R.P., Saari E.E. Rust Diseases of Wheat, Concepts and Methods of Disease Management. Mexico: CIMMYT, 1992.

Singh R.P. Resistance to leaf rust in 26 Mexican wheat cultivars. *Crop Sci*. 1993;33(3):633-637. DOI 10.2135/cropsci1993.0011183X003300030041x.

Singh R.P., Gupta A.K. Genes for leaf rust in Indian and Pakistani wheats tested with Mexican pathotypes of *Puccinia recondita* f. sp. *tritici*. *Euphytica*. 1991;57:27-36. DOI 10.1007/BF00040475.

Singh R.P., Rajaram S. Genetics of adult plant resistance to leaf rust in "Frontana" and three CIMMYT wheats. *Genome*. 1992;35(1):24-31. DOI 10.1139/g92-004.

Singh R.P., Chen W.Q., He Z.H. Leaf rust resistance of spring facultative and winter wheat cultivars from China. *Plant Dis*. 1999;83(7): 644-651. DOI 10.1094/PDIS.1999.83.7.644.

Singh R.P., Huerta E.J., Bhavani S., Herrera F.S.A., Singh D., Singh P.K., Velu G., Mason R.E., Jin Y., Njau P., Crossa J. Race non-specific resistance to rust diseases in CIMMYT spring wheats. *Euphytica.* 2011;179:175-186. DOI 10.1007/s10681-010-0322-9.

Singh R.P., Singh P.K., Rutkoski J., Hodson D.P., He X., Jorgensen L.N., Hovmoller M.S., Huerta-Espino J. Disease impact on wheat yield potential and prospects of genetic control. *Annu. Rev. Phytopathol*. 2016;54:303-322. DOI 10.1146/annurev-phyto-080615-095835.

Stakman E.C., Stewart D.M., Loegering W.Q. Identification of Physiologic Races of *P. graminis* var. *tritici.* Washington: US Department of Agriculture, USA, 1962.

Statler G.D. Probable genes for leaf rust resistance in several hard red spring wheats. *Crop Sci*. 1984;24(5):883-886. DOI 10.2135/cropsci1984.0011183X002400050013x.

Stubbs R.W., Prescott J.M., Saari E.E., Dubin H.J. Cereal Disease Methodology Manual. Mexico: CIMMYT Publications, 1986.

Van der Plank J.E. Plant Diseases: Epidemics and Control. Acad. Press, 1963. DOI 10.1016/C2013-0-11642-X.

Volkova G.V., Kudinova O.A., Vaganova O.F. Selection of *Puccinia triticina* virulence genes on winter wheat varieties with different types of resistance. *Agrarian Bull. Urals*. 2020;199(8):25-33. DOI 10.32417/1997-4868-2020-199-8-25-33. (in Russian)

Wu H., Kang Z.H., Li X., Li Y.Y., Li Y., Wang S., Liu D.Q. Identification of wheat leaf rust resistance genes in Chinese wheat cultivars and the improved germplasms. *Plant Dis*. 2020;104(10):2669-2680. DOI 10.1094/PDIS-12-19-2619-RE.

Yan X.C., Li Z.F., Yang H.L., Zhang H.H., Gebrewahid T.W., Yao Z.J., Liu D.Q., Zhou Y. Analysis of wheat leaf rust resistance genes in 30 important wheat cultivars. *Sci. Agric. Sin*. 2017;50(2):272-285. DOI 10.3864/j.issn.0578-1752.2017.02.007.

Zhang P.P., Gebrewahid T.W., Zhou Y., Li O.L., Li Z.F., Liu D.Q. Seedling and adult plant resistance to leaf rust in 46 Chinese bread wheat landraces and 39 wheat lines with known *Lr* genes. *J. Integr. Agric*. 2019;18(5):1014-1023. DOI 10.1016/S2095-3119(19)62575-X.

ORCID ID

V. Ivanova orcid.org/0000-0003-3745-5001

Conflict of interest. The author declares no conflict of interest.

Received October 27, 2022. Revised January 15, 2023. Accepted January 16, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):454-462
DOI 10.18699/VJGB-23-55

Original Russian text https://vavilovj-icg.ru/

Comprehensive agrobiological assessment and analysis of genetic relationships of promising walnut varieties of the Nikitsky Botanical Gardens

Yu.V. Plugatar[1], I.I. Suprun[2] ✉, S.Yu. Khokhlov[1], I.V. Stepanov[2], E.A. Al-Nakib[2]

[1] The Order of the Red Banner of Labour Nikitsky Botanical Gardens – National Scientific Center of the Russian Academy of Sciences, Yalta, Republic of Crimea, Russia
[2] North Caucasian Federal Scientific Center of Horticulture, Viticulture, Wine-making, the Functional Scientific Center of "Breeding and Nursery", Krasnodar, Russia
✉ supruni@mail.ru

Abstract. Walnut is an important horticultural crop, the production of which ranks second among all nut crops. Despite the significant demand in the domestic market in Russia, the industrial production of walnut fruits in Russia is currently underdeveloped. At the same time, there is a need to update the assortment with new highly productive varieties adapted to local agro-climatic conditions and having high quality nuts that are competitive at the world level. An important issue for the successful implementation of breeding programs is a comprehensive study of the gene pool. In this regard, within the framework of the study, the task was to evaluate promising varieties from the collection of the walnut gene pool of the Nikitsky Botanical Gardens and analyze genetic relationships based on microsatellite genotyping. On the basis of the performed phenotypic assessment, the study sample, which included 31 varieties, was divided into several groups according to the main phenotypic traits, such as frost and drought resistance, the start of the growing season, the ripening period, the weight and type of flowering, the weight of the fruit, and the thickness of the endocarp. Varieties with economically valuable traits that can be recommended as promising as initial parental forms in breeding work for resistance to abiotic stress factors have been identified, as well as varieties with increased productivity and large fruit sizes. Based on the analysis of eight SSR markers (WGA001, WGA376, WGA069, WGA276, WGA009, WGA202, WGA089 and WGA054), an analysis of the level of genetic diversity was performed and genetic relationships were established in the studied sample of varieties. Six (for WGA089) to eleven (for WGA276) alleles per locus have been identified. A total of 70 alleles were identified for the eight DNA markers used, with an average value of 8.75. Analysis of SSR genotyping data using Bayesian analysis established the presence of two main groups of genotypes. Taking into account the fact that all the studied varieties are selections from local seed populations in different regions of the Crimean Peninsula, the revealed level of polymorphism may indirectly reflect the level of genetic diversity of the local walnut populations. Furthermore, the presence of two genetically distant groups indicates the presence of two independently formed pools of the autochthonous gene pool of the species *Juglans regia* L. on the Crimean Peninsula
Key words: walnut; SSR markers; perspective cultivars; collection; genetic diversity; phenotypic evaluation.

For citation: Plugatar Yu.V., Suprun I.I., Khokhlov S.Yu., Stepanov I.V., Al-Nakib E.A. Comprehensive agrobiological assessment and analysis of genetic relationships of promising walnut varieties of the Nikitsky Botanical Gardens. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):454-462. DOI 10.18699/VJGB-23-55

Комплексная агробиологическая оценка и анализ генетических взаимосвязей перспективных сортов ореха грецкого Никитского ботанического сада

Ю.В. Плугатарь[1], И.И. Супрун[2] ✉, С.Ю. Хохлов[1], И.В. Степанов[2], Е.А. Аль-Накиб[2]

[1] Ордена Трудового Красного Знамени Никитский ботанический сад – Национальный научный центр Российской академии наук, Ялта, Республика Крым, Россия
[2] Северо-Кавказский федеральный научный центр садоводства, виноградарства, виноделия, Краснодар, Россия
✉ supruni@mail.ru

Аннотация. Орех грецкий – важная садовая культура, которая по объему производства занимает второе место среди всех орехоплодных. Несмотря на значительную потребность на внутреннем рынке, промышленное производство плодов ореха грецкого в России в настоящее время развито недостаточно. При этом существует необходимость обновления сортимента новыми высокопродуктивными сортами, адаптированными к местным агроклиматическим условиям и обладающими высоким качеством плодов, конкурентоспособными на мировом уровне. Важным вопросом для успешной реализации селекционных программ является комплексное изуче-

ние генофонда. В связи с этим в рамках исследования была поставлена задача оценки по комплексу признаков перспективных сортов из коллекции генофонда ореха грецкого Никитского ботанического сада и анализа генетических взаимосвязей на основе микросателлитного генотипирования. По результатам выполненной фенотипической оценки изучаемая выборка, включающая 31 сорт, была разделена на несколько групп по основным хозяйственно-биологическим характеристикам, таким как морозо- и засухоустойчивость, срок начала вегетации, сроки созревания, тип цветения, масса плода, толщина эндокарпа. Выделены сорта с хозяйственно ценными признаками, которые можно рекомендовать как перспективные в качестве исходных родительских форм в селекционной работе на устойчивость к абиотическим стресс-факторам, а также сорта с повышенной урожайностью и обладающие крупным размером плодов. На основании анализа восьми SSR-маркеров (WGA001, WGA376, WGA069, WGA276, WGA009, WGA202, WGA089 и WGA054) оценен уровень генетического разнообразия и определены генетические взаимосвязи в изученной выборке сортов. Выявлено наличие от шести (WGA089) до одиннадцати (WGA276) аллелей на локус. Суммарно по восьми использованным ДНК-маркерам было идентифицировано 70 аллелей при среднем значении 8.75. Анализ данных SSR-генотипирования в программе Structure 2.3.4 установил наличие двух основных групп генотипов. С учетом того, что все изученные сорта представляют собой отборы из местных семенных популяций в разных районах Крымского полуострова, выявленный уровень полиморфизма может опосредованно отражать уровень генетического разнообразия местного генофонда ореха грецкого. При этом наличие двух генетически обособленных групп, вероятно, свидетельствует о существовании двух независимо сформировавшихся пулов автохтонного генофонда вида *Juglans regia* L. на Крымском полуострове.

Ключевые слова: орех грецкий; SSR-маркеры; перспективные сорта; коллекция; генетическое разнообразие; фенотипическая оценка.

Introduction

Walnut is one of the most important nut crops, which is second only to almonds in terms of production. The world leaders in the production of walnuts are China, Iran, the USA and Turkey (Vahdati et al., 2019). Commercial production of walnut fruits in Russia is currently not developed, however, there are positive trends in the establishment of industrial orchards. At the same time, there is a need to update the assortment with new highly productive varieties adapted to local agro-climatic conditions and having high quality fruits.

Obviously, a comprehensive study of the walnut gene pool is an important issue for increasing the efficiency of using genetic resources in solving breeding problems to create new generation varieties, as well as for preserving and replenishing collections. At the same time, the assessment of the level of genetic diversity, elucidation of the degree of genetic similarity, as well as DNA certification of collection samples occupy an important place. One of the most popular methods for assessing the genetic diversity of the walnut is the analysis of microsatellite loci of the genome (Vahdati et al., 2019). SSR markers currently widely used for the analysis of walnut polymorphism were developed both directly for this species (Dangl et al., 2005; Topcu et al., 2015; Ikhsana et al., 2016) and for the species *Juglans nigra* L. (Woeste et al., 2002). Subsequently, these markers were effectively used to study the interspecific diversity of *J. regia* due to the high level of cross-reproducibility within the *Juglans* genus. At the same time, it is worth highlighting the SSR markers marked as "WGA", which are the most commonly used (Bernard et al., 2018b). With the use of SSR markers for walnut, a significant amount of research has been carried out aimed at analyzing the genetic structure of gene pool collections, including varieties, promising breeding forms, as well as selections from local populations of interest for breeding.

The most large-scale studies of the genetic diversity of collections of genetic resources include the work performed by a team of authors from the INRA Research Center, in which, using 13 SSR markers, genotyping of 217 accessions of walnut and 36 accessions of other species of the genus *Juglans* from the INRA collection was carried out. Based on the SSR genotyping data, the presence of the main two groups of the greatest genetic similarity was established, which for the most part corresponded to the ecological and geographical origin of the varieties. The data obtained made it possible to form a core collection of fifty samples, reflecting the genetic polymorphism of the entire sample (Bernard et al., 2018a). It is noteworthy that a comparative analysis of the effectiveness of using 13 SSR markers in the above work (Bernard et al., 2018a) and 364,275 SNP markers – data obtained using the Axiom™ *J. regia* 700K SNP genotyping array SNP chip, showed a close level of information content of the two approaches used in assessing the genetic structure of collections (Bernard et al., 2018a, 2020a). A comparable study of a sample of 189 varieties and breeding forms, representative of gene collections from 25 regions in 14 countries of the world, made it possible to establish the presence of two main groups, including accessions from: (1) Europe and North Africa and (2) Greece and the Middle East (Ebrahimi et al., 2016). Along with such large-scale studies, a wide range of studies was performed using microsatellites on gene pool collections, as well as local populations in different regions of the world: Europe (Pollegioni et al., 2011; Ebrahimi et al., 2017b; Vischi et al., 2017; Cseke et al., 2022), East Asia (Gunn et al., 2010; Wang et al., 2015; Zhou et al., 2017), Central and South Asia (Pollegioni et al., 2014; Roor et al., 2017; Shah et al., 2018; Gaisberger et al., 2020; Magige et al., 2022), the Middle East region (Ebrahimi et al., 2011; Shamlu et al., 2018; Orhan et al., 2020; Davoodi et al., 2021; Guney et al., 2021), North America (Dangl et al., 2005; Aradhya et al., 2010; Ebrahimi et al., 2017a).

A number of studies are known in which, along with molecular genetic analysis of polymorphism based on

SSR markers, a comprehensive assessment of the phenotypic variability of samples was performed (Ebrahimi et al., 2011), or an assessment of individual groups of traits, such as fruit characteristics (Chen et al., 2014). This made it possible both to compare the efficiency of using different approaches to determine the groups of the greatest genetic similarity (Pop et al., 2013), and to identify selectively valuable accessions at the first stage and subsequently to evaluate the heterogeneity of the selected groups of accessions based on microsatellite analysis data (Karimi et al., 2014; Davoodi et al., 2021).

Despite the ongoing breeding work on walnuts in the south of Russia (Lugovskoi, Murzinova, 2010; Khokhlov, Baskakova, 2015; Suprun et al., 2016; Lugovskoy, Balapanov, 2018) and the availability of research results on the study of collections of varieties using molecular genetic methods (Balapanov et al., 2019), one should still note the limitations of studies aimed at analyzing the level of genetic diversity and identifying the genetic structure of the gene pool in the South of Russia, including the Crimea and the North Caucasus. The Nikitsky Botanical Gardens (NBG-NSC) is one of the leading scientific organizations in the Russian Federation that performs breeding work on walnuts. The collection of genetic resources of the Nikitsky botanical walnut is represented by 76 accessions. The basis is made up of varieties of selection NBG-NSC (86 %). Among the introduced genotypes, 10 % of the total collection falls on varieties from Moldova and 3 % each on accessions from Ukraine, Europe, the USA, and Tajikistan (Khokhlov, Baskakova, 2015). It is obvious that a comprehensive phenotypic assessment, the identification of groups of the most valuable genotypes, characterized by the presence of several breeding-valuable traits at the same time, as well as the analysis of genetic relationships of valuable varieties and forms, will improve the efficiency of the breeding use of the gene pool in order to create new adaptive varieties with increased productivity potential and high fruit quality.

In the presented work, we set the task of assessing perspective varieties of walnut from the collection of the gene pool of the NBG-NSC according to economically valuable traits, identifying groups of varieties with a complex of important characteristics and analyzing their genetic relationships using microsatellite DNA markers.

Materials and methods

Phenotypic assessment was carried out in the collection plantations of the laboratory of steppe horticulture (LSH) of the NBG-NSC in 2014–2022. 31 samples of walnut selection from the Nikitsky Botanical Gardens were chosen as the object of observation (Khokhlov, 2012). The LSH territory is located 25 km north of Simferopol, in the village of Novy Sad (45°08′50″ N, 33°59′55″ E), Republic of Crimea, Russia. In the system of agro-climatic zoning of the peninsula, it belongs to the central plain-steppe region, characterized by an arid climate with a moderately hot growing season and mild, unstable winters (Antyufeev et al., 2002). Also, in the genotyping work, the Chandler variety of the USA selection was used. The relief of the area on which the collection garden is located is flat and slightly wavy; the soil of the plot is southern carbonate low-humus heavy loamy chernozem on red-brown Pliocene clays. The average annual air temperature is +10.5 °C, the average January does not exceed –1.0 °C, and the average July, +21.9 °C. Walnut plants are planted according to the scheme 12×12 m, peach is used as a compactor. The aisles are kept under black fallow. The age of the trees is 30 years.

Determination of the degree of frost resistance of varieties was carried out according to the method developed in the Nikitsky Gardens (Rihter, Yadrov, 1981) and the method of Lapin and Ryabova (1982). The assessment of drought resistance of walnut plants was carried out in accordance with methodological recommendations (Eremeev, Lishchuk, 1974; Kushnirenko et al., 1975; Il'nitskiy, 2005).

A modified CTAB method was used for DNA extraction (Rogers, Bendich, 1985). Genotyping of walnut varieties was carried out using 8 SSR markers: WGA001, WGA376, WGA069, WGA276, WGA009, WGA202, WGA089, WGA054 (Woeste et al., 2002; Dangl et al., 2005). PCR was carried out under the following conditions: the concentration of PCR reagents of the mixture: Buffer 1X, dNTP – 0.24 mM, Taq 1U, SSR primers (forward and reverse) – 0.16 μM each, DNA – 40–50 ng. The following PCR parameters were used: 3 min initial denaturation at 94 °C; the next 35 cycles: 20 sec denaturation at 94 °C, 30 sec primer annealing at 58 °C, 40 sec elongation at 72 °C; final elongation for 10 min at 72 °C. The size of the reaction products was analyzed on a Nanofor 05 automatic genetic analyzer.

The data were processed using the GeneMarker V3.0.1 program. The following genetic parameters were calculated in the Microsoft Exel GenAlEx 6.503 macro: Na – number of alleles, Na (average) – average number of alleles, Ne – effective number of alleles, I – Shannon diversity index, Ho – observed heterozygosity, He – expected heterozygosity, F – fixation index (Peakall, Smouse, 2012). The PCoA plot with the genetic similarity coefficient Dice was built using the Past 2.17 program (Hammer et al., 2001) based on a binary matrix. Cluster analysis was carried out using the Structure 2.3.4 program. The optimal value of clusters for analysis was established in the Structure Harvester online program (Evanno et al., 2005).

Research results

Phenotypic evaluation

Based on the comprehensive phenotypic assessment, the studied varieties were combined into several groups according to the main economic and biological characteristics.

By the degree of frost resistance: varieties ***with high frost resistance***, in which 60–100 % of generative and vegetative buds were preserved without damage – 'Arkad', 'Burlyuk', 'Orionid', 'Skaberi', 'Yuzhnoberezhniy', 'Pozdnotsvetushchiy'; ***moderate frost resistance*** (from 40 to 60 %) – 'Bospor', 'Al'minskiy', 'Konkursniy', 'Pamyati Pasenkova', 'Dolinniy', 'Zolotopolyanskiy', 'Krymskiy Skoroplodniy', 'Zhemchuzhniy', 'Sokoliniy', 'Novikov', 'Bulganak', 'Gurzufskiy', 'Sladkoyaderniy', 'Pioner Kryma', 'Bel'bekskiy Ranniy', 'Partizanskiy', 'Dzerzhinskiy', 'Bel'bekskiy', 'Komsomolets', 'Bomba Chkalovskaya', 'Kollektivniy'; ***low frost resistance*** (less than 40 %) – 'Bubenchik', 'Kacha', 'Malosadoviy', 'Podlesniy'.

By the degree of drought resistance: ***with high stability*** – Arkad', 'Burlyuk, 'Orionid', 'Bel'bekskiy Ranniy', 'Zhemchuzhniy'; ***with stability above average*** – 'Al'minskiy', 'Bospor', 'Konkursniy', 'Pamyati

Pasenkova', 'Zolotopolyanskiy', 'Krymskiy Skoroplodniy', 'Pozdnotsvetushchiy', 'Sokoliniy', 'Yuzhnoberezhniy', 'Novikov', 'Bulganak', 'Gurzufskiy', 'Sladkoyaderniy', 'Dolinniy', 'Pioner Kryma', 'Kollektivniy', 'Partizanskiy', 'Dzerzhinskiy', 'Bel'bekskiy', 'Komsomolets', 'Malosadoviy', 'Podlesniy', 'Skaberi', 'Bomba Chkalovskaya'; ***with stability below average*** – 'Bubenchik', 'Kacha'.

By maturity: ***early*** – Arkad', 'Bulganak', 'Dolinniy', 'Komsomolets', 'Krymskiy Skoroplodniy', 'Zhemchuzhniy', 'Orionid', 'Yuzhnoberezhniy'; ***middle*** – 'Al'minskiy', 'Novikov', 'Gurzufskiy', 'Sladkoyaderniy', 'Zolotopolyanskiy', 'Pamyati Pasenkova', 'Burlyuk', 'Pozdnotsvetushchiy', 'Sokoliniy', 'Dzerzhinskiy', 'Bospor', 'Bubenchik', 'Kollektivniy', 'Kacha', 'Partizanskiy', 'Podlesniy', 'Pioner Kryma', 'Bel'bekskiy Ranniy', 'Bomba Chkalovskaya', 'Skaberi', 'Kollektivniy'; ***late*** – Konkursniy', 'Malosadoviy'.

By type of flowering: ***protogeny*** (male inflorescences bloom first) – 'Al'minskiy', 'Novikov', 'Bulganak', 'Gurzufskiy', 'Sladkoyaderniy', 'Dolinniy', 'Pioner Kryma', 'Bel'bekskiy Ranniy', 'Bomba Chkalovskaya', 'Skaberi'; ***protandria*** (female flowers bloom first) – 'Bubenchik', 'Kollektivniy', 'Kacha', 'Partizanskiy', 'Konkursniy', 'Dzerzhinskiy', 'Bel'bekskiy', 'Komsomolets', 'Malosadoviy', 'Podlesniy'; ***homogamy*** (simultaneous flowering of male inflorescences and female flowers) – 'Zolotopolyanskiy', 'Arkad', 'Krymskiy Skoroplodniy', 'Zhemchuzhniy', 'Pamyati Pasenkova', 'Burlyuk', 'Pozdnotsvetushchiy', 'Sokoliniy', 'Yuzhnoberezhniy', 'Bospor', 'Orionid'.

By fruit weight: ***large-fruited*** (more than 12 g, belong to the variety *J. regia* L. var. *macrocarpa* DC. or *J. regia* f. *maxima*) – 'Bomba Chkalovskaya', 'Bulganak', 'Dolinniy', 'Pioner Kryma', 'Bel'bekskiy Ranniy', 'Skaberi', 'Komsomolets', 'Malosadoviy', 'Podlesniy', 'Arkad', 'Krymskiy Skoroplodniy', 'Burlyuk', 'Pozdnotsvetushchiy', 'Sokoliniy', 'Bospor', 'Orionid'; ***medium-fruited*** (from 6 to 12 g) – 'Al'minskiy', 'Zolotopolyanskiy', 'Pamyati Pasenkova', 'Yuzhnoberezhniy', 'Kollektivniy', 'Kacha', 'Novikov', 'Partizanskiy', 'Konkursniy', 'Dzerzhinskiy', 'Bel'bekskiy', 'Gurzufskiy', 'Sladkoyaderniy', 'Zhemchuzhniy'; ***small-fruited*** (less than 6 g) – 'Bubenchik'. In all varieties, with the exception of 'Bomba Chkalovskaya', the shape of the fruit is oval-round or ovoid.

According to the thickness of the endocarp: ***thin-shelled*** (from 1.0 to 1.5 mm, belong to the variety *J. regia* L. var. *tenera* DC.) – 'Zolotopolyanskiy', 'Yuzhnoberezhniy'; ***standard shell*** (from 1.5 to 2 mm, *J. regia* f. *semidura* DC.) – 'Bomba Chkalovskaya', 'Bulganak', 'Dolinniy', 'Pioner Kryma', 'Bel'bekskiy Ranniy', 'Skaberi', 'Partizanskiy', 'Komsomolets', 'Malosadoviy', 'Podlesniy', 'Arkad', 'Krymskiy Skoroplodniy', 'Burlyuk', 'Pozdnotsvetushchiy', 'Sokoliniy', 'Bospor', 'Orionid', 'Bubenchik', 'Kollektivniy', 'Al'minskiy', 'Novikov', 'Gurzufskiy', 'Sladkoyaderniy', 'Pamyati Pasenkova', 'Zhemchuzhniy'; ***hard-shelled*** (more than 2.0 mm, belong to the variety *J. regia* L. var. *dura* DC.) – 'Kacha', 'Partizanskiy', 'Konkursniy', 'Dzerzhinskiy'.

By the beginning of the growing season: ***early*** – 'Dolinniy', 'Komsomolets', 'Arkad', 'Zhemchuzhniy', 'Bel'bekskiy Ranniy'; ***medium*** – 'Al'minskiy', 'Novikov', 'Bulganak', 'Gurzufskiy', 'Sladkoyaderniy', 'Pioner Kryma', 'Bomba Chkalovskaya', 'Skaberi', 'Bubenchik', 'Kollektivniy', 'Kacha, 'Partizanskiy', 'Dzerzhinskiy', 'Bel'bekskiy', 'Malosadoviy', 'Podlesniy', 'Zolotopolyanskiy', 'Pamyati Pasenkova', 'Burlyuk', 'Sokoliniy', 'Bospor', 'Orionid', 'Yuzhnoberezhniy', 'Krymskiy Skoroplodniy'; ***late*** – 'Pozdnotsvetushchiy', 'Konkursniy'.

The results of a long-term study of the walnut gene pool make it possible to identify varieties with economically valuable traits that can be recommended as initial parental forms in breeding work: for tolerance to abiotic stresses during winter-spring period – 'Arkad', 'Burlyuk', 'Orionid', 'Skaberi', 'Yuzhnoberezhniy', 'Pozdnotsvetushchiy'; for increased and high drought resistance – 'Arkad', 'Burlyuk', 'Orionid', 'Bel'bekskiy Ranniy', 'Zhemchuzhniy'. For introduction into production, varieties with complex resistance to adverse climatic conditions are recommended – 'Burlyuk', 'Bospor', 'Arkad', 'Al'minskiy', 'Pamyati Pasenkova', 'Orionid', 'Konkursniy', as well as those characterized by high yield and large fruits – 'Bulganak', 'Dolinniy', 'Pioner Kryma', 'Bel'bekskiy Ranniy', 'Skaberi', 'Partizanskiy', 'Komsomolets', 'Malosadoviy', 'Podlesniy', 'Arkad', 'Krymskiy Skoroplodniy', 'Burlyuk', 'Pozdnotsvetushchiy', 'Sokoliniy', 'Bospor', 'Orionid'.

Analysis of genetic diversity

In order to establish genetic relationships within the studied sample of varieties, analyze the level of genetic diversity and identify the groups of the closest genetic relationship, an analysis of the polymorphism of microsatellite loci was performed.

As a result of microsatellite genotyping, DNA profiles specific for all studied varieties were obtained. Six (WGA089) to eleven (WGA276) alleles per locus have been identified. A total of 70 alleles were identified for the eight DNA markers used, with an average value of 8.75. Analysis of the level of genetic polymorphism included the indicators presented in Table 1.

The value of the Ne index varied from 2.873 (WGA089) to 6.450 (WGA276). At the same time, in the group of markers with the same number of identified alleles (9 alleles per locus): WGA001, WGA376, WGA009, and WGA202, this indicator varied from 3.269 (WGA202) to 5.071 (WGA376), which may be due to variation in allele frequencies. The lowest (1.337) and highest (2.060) values of the diversity index I were found in the least polymorphic marker (WGA089) and the most polymorphic marker (WGA276), respectively. At the same time, the highest value of the observed heterozygosity was found for the WGA001 marker, and the expected heterozygosity, for the most polymorphic WGA276 marker.

Based on the genotyping data of 32 walnut varieties for eight SSR markers, an analysis was carried out using the Structure 2.3.4 program. The range of analyzed clusters was from 2 to 7. Based on the results of the analysis in the Structure Harvester online program, the optimal cluster value was calculated equal to 2. The results obtained with a value of K = 2 are shown in Fig. 1.

According to the predominance of the first or second clusters, the studied Crimean varieties can be conditionally divided into two groups. The first group (the predominance of cluster 1): 'Bulganak', 'Arkad', 'Dolinniy', 'Partizanskiy',

Table 1. Level of polymorphism of SSR markers

Locus	Na	Ne	I	Ho	He
WGA001	9.000	4.854	1.781	0.710	0.794
WGA376	9.000	5.071	1.829	0.677	0.803
WGA069	10.000	6.160	2.022	0.677	0.838
WGA276	11.000	6.450	2.060	0.645	0.845
WGA009	9.000	4.215	1.750	0.419	0.763
WGA202	9.000	3.269	1.487	0.484	0.694
WGA089	6.000	2.873	1.337	0.484	0.652
WGA054	7.000	4.107	1.579	0.581	0.757
Average	8.750	4.625	1.731	0.585	0.768

Note. Na is the number of identified alleles; Ne is the number of effective alleles; I – diversity index; Ho – observed heterozygosity; He – expected heterozygosity.

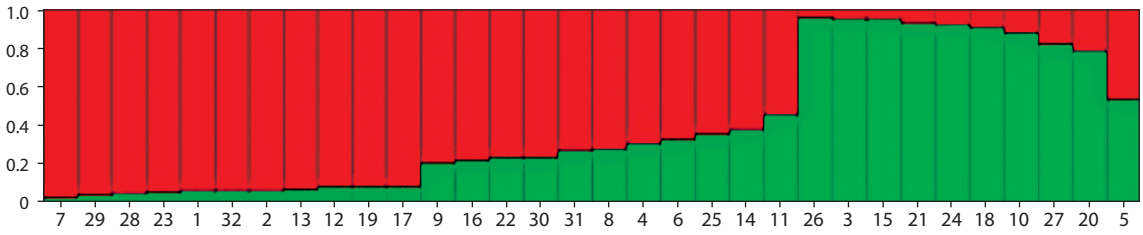


Fig. 1. Graph constructed in the Structure program based on genotyping data of eight SSR markers of 32 walnut varieties.

Varieties: 1 – 'Al'minskiy', 2 – 'Novikov', 3 – 'Bulganak', 4 – 'Bubenchik', 5 – 'Arkad', 6 – 'Gurzufskiy', 7 – 'Sladkoyaderniy', 8 – 'Zolotopolyanskiy', 9 – 'Krymskiy Skoroplodniy', 10 – 'Dolinniy', 11 – 'Kollektivniy', 12 – 'Zhemchuzhniy', 13 – 'Konkursniy', 14 – 'Kacha', 15 – 'Partizanskiy', 16 – 'Pamyati Pasenkova', 17 – 'Pozdnotsvetushchiy', 18 – 'Pioner Kryma', 19 – 'Dzerzhinskiy', 20 – 'Sokoliniy', 21 – 'Bel'bekskiy', 22 – 'Bospor', 23 – 'Komsomolets', 24 – 'Bel'bekskiy Ranniy', 25 – 'Bomba Chkalovskaya', 26 – 'Malosadoviy', 27 – 'Podlesniy', 28 – 'Skaberi', 29 – 'Yuzhnoberezniy', 30 – 'Burlyuk', 31 – 'Orionid', 32 – 'Chandler'.

'Pioner Kryma', 'Sokoliniy', 'Bel'bekskiy', 'Bel'bekskiy Ranniy', 'Malosadoviy', 'Podlesniy'. The second group (the predominance of cluster 2): 'Al'minskiy', 'Novikov', 'Bubenchik', 'Gurzufskiy', 'Sladkoyaderniy', 'Zolotopolyanskiy', 'Krymskiy Skoroplodniy', 'Kollektivniy', 'Zhemchuzhniy', 'Konkursniy', 'Kacha', 'Pamyati Pasenkova', 'Pozdnotsvetushchiy', 'Dzerzhinskiy', 'Bospor', 'Komsomolets', 'Bomba Chkalovskaya', 'Skaberi', 'Yuzhnoberezniy', 'Burlyuk', 'Orionid'. Variety 'Chandler' was assigned to the second group of varieties. It should be noted that some varieties of the second group have a significant contribution from the first cluster (from 0.185 to 0.481), on the other hand, among the varieties assigned to the first group, two varieties have a significant contribution from the second cluster (0.215 and 0.448).

For a detailed analysis of the genetic relationship of the studied walnut genotypes, an analysis was carried out by the method of principal coordinates (PCoA) in the PAST 2.17 program (Fig. 2).

The distribution of varieties on the PCoA plot largely reflects the grouping of varieties obtained in the Structure program. Varieties of the first group are concentrated at the bottom of the graph. In turn, the varieties of the second group are distributed in the middle and upper parts of the graph. In the arrangement of varieties of the first group, subgroups can be distinguished: (1) varieties 'Bel'bekskiy Ranniy', 'Bel'bekskiy' and 'Podlesniy', (2) 'Partizanskiy', 'Malosadoviy', 'Pioner Kryma', (3) 'Sokoliniy', 'Bulganak'. Two varieties from the first group were not included in one of the subgroups: variety 'Arkad' occupies an intermediate position between varieties of the first and second groups, variety 'Dolinniy' is equidistant from other varieties included in the first group.

Varieties of the second group are distributed on the graph less orderly and do not form clear structures, however, it is worth noting that the varieties 'Kollektivniy', 'Zolotopolyanskiy', 'Bomba Chkalovskaya' and 'Yuzhnoberezniy' occupy an intermediate position between the varieties of the first and second groups. Variety 'Chandler' on the graph of principal coordinates is spatially close to variety 'Al'minskiy'.

Discussion

The SSR markers used in our work were previously widely used to solve various problems in walnut genetics and breeding, including DNA fingerprinting and analysis of the genetic diversity of cultivar collections, breeding-promising forms, and interspecific hybrids (Woeste et al., 2002; Pollegioni et al., 2009; Ebrahimi et al., 2016; Vahdati et al., 2019), study of trait-related collections (Ebrahimi et al., 2017a), clarification of the

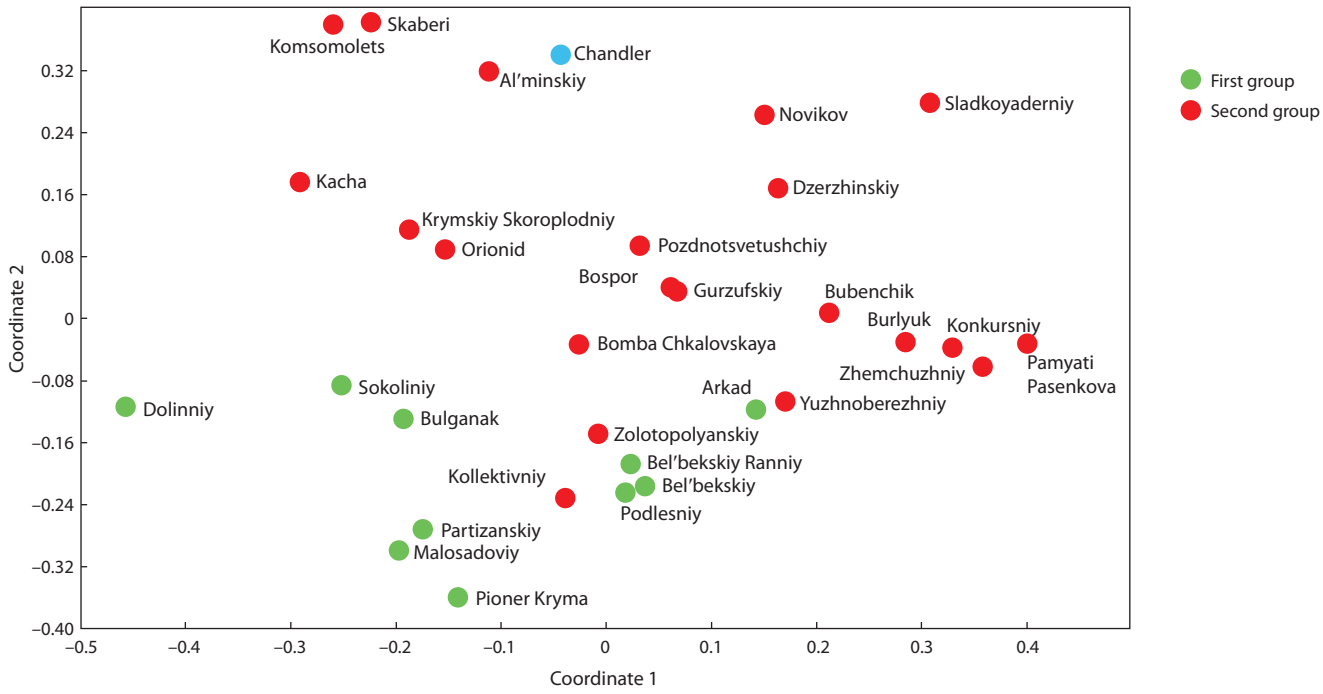


Fig. 2. Estimation of genetic relationship by the method of principal coordinates of walnut varieties according to SSR genotyping data.

issues of gene pool formation within its natural habitats, as well as distribution pathways in the process of domestication (Pollegioni et al., 2014, 2015, 2017).

Comparison of the average values of indicators characterizing polymorphism, identified by the results of our work, and in studies conducted on collections of walnut varieties from other regions of the world, makes it possible to compare the heterogeneity of the studied samples of varieties with that studied in this work. Thus, in a study of a collection of 35 varieties of Chinese breeding using 10 SSR markers, the average Na and Ne values were 9.4 and 4.67, respectively, while the values of expected (He) and observed (Ho) heterozygosity were 0.77 and 0.62, respectively (Chen et al., 2014). In the work of M. Aradhya et al. (Aradhya et al., 2010), when analyzing the genetic polymorphism of a collection of 236 varieties from different walnut growing regions using 15 microsatellite markers, the average number of identified alleles per locus was 11, while the average He and Ho values were 0.699 and 0.536, respectively. It is worth noting that in this study, markers WGA001, WGA202, WGA009, and WGA069 showed higher polymorphism (12, 19, 11, and 13 alleles per locus, respectively), while WGA089 was one of the least polymorphic (8 alleles) (Aradhya et al., 2010). We obtained similar results in terms of the level of allelic polymorphism of markers (see Table 1).

In an analysis of a sample of 189 varieties representing the gene pool of 14 countries, an average of 11.5 alleles per locus was identified, while the average values of observed and expected heterozygosity were 0.62 and 0.73 (Ebrahimi et al., 2016). In this work, the markers WGA001, WGA202, and WGA276, as well as in our study, were included in the group of more polymorphic ones, and the WGA068 marker showed a lower level of polymorphism. In the work of Turkish researchers who performed genotyping of 30 elite breeding forms (candidates for varieties) using 21 SSR markers, an average of 6.15 alleles per locus were identified, while the observed and expected heterozygosity was 0.64 and 0.62 (Bozhuyuk et al., 2020).

Considering the works aimed at studying the genetic diversity of natural populations and promising for breeding forms selected from them, one can also speak of a comparable level of polymorphism. For example, in a study by F. Shamlu et al. (2018), when assessing polymorphism and genetic relationships in a sample of 39 walnut accessions selected in natural populations in northeastern Iran, the average number of identified alleles per locus was 7.9, and the number of effective alleles was 3.91. Expected and observed heterozygosity values were 0.74 and 0.93, respectively. At the same time, the diversity index was lower than the indicator we identified, 1.34 (Shamlu et al., 2018).

In studies devoted to the analysis of the structure of natural populations, elucidation of the ways in which the gene pool spreads, and the formation of its local pools, the indicators of the number of alleles varied. In a study of the genetic diversity of local populations in the Eastern Alps of Italy (Vischi et al., 2017), the average number of alleles per locus was 4.7 in a sample of 13 markers (WGA class) and 2.7 in a sample of seven EST-SSR markers, which is a rather low figure, especially given the sample size of about 200 samples. This can be explained by the possible isolation of the studied population, selected in mountainous areas, as well as in the flat area limited by them (Vischi et al., 2017). In studies of natural walnut populations in Southwestern Tibet, when analyzing a sample of 86 genotypes selected in five geographical locations, the Na indicator was 9.92, but the Ne value was 3.95, which may be due to an uneven distribution of allele frequencies (Wang et al., 2015). In a study aimed at studying the ways of formation and distribution of the walnut gene

Table 2. Comparison of groups of walnut varieties by genetic characteristics

Group	Na	Ne	I	Ho	He	F
First	4.125	2.581	1.082	0.588	0.579	–0.009
Second	8.375	5.063	1.783	0.583	0.786	0.258

pool from the centers of its origin in Eurasia, as a result of SSR genotyping of a sample of about 2000 genotypes using 14 SSR markers, an average allele number per locus was found to be 14.21 (Pollegioni et al., 2017).

In general, considering the work on the study of the genetic polymorphism of the walnut gene pool, both cultural forms (varieties, hybrids, selections from local populations) and local populations, including natural ones in regions related to the centers of origin of the species *J. regia* L., we can make a conclusion about the high level of polymorphism of the sample of varieties studied by us. Given the fact that all the varieties studied are selections from local seed populations in different regions of the Crimean Peninsula, this level of polymorphism may indirectly reflect the level of genetic diversity of the local walnut gene pool. This is confirmed to a certain extent by the results obtained by us in the course of cluster analysis. Based on the data obtained, it can be concluded that the autochthonous gene pool of the walnut probably comes from two hypothetical populations. The division of the studied sample of varieties into two groups based on the results of Bayesian analysis is consistent with the high level of polymorphism, since the presence of two genetically distinct groups may contribute to a higher level of genetic heterogeneity, including allelic polymorphism of DNA markers.

In the work, the groups identified during clustering were compared according to a number of population genetic parameters (Table 2).

The value of the average number of alleles per locus in the second group is two times higher than this indicator in the first group; such indicators of genetic diversity as the effective number of alleles and the Shannon diversity index also reflect a greater allelic polymorphism of microsatellite markers in the second group of varieties. The observed heterozygosity in the groups has a similar value of 0.583 and 0.588, in turn, the expected heterozygosity is higher in the second group. The fixation index has a low positive value in the second group of varieties; in the varieties of the first group, the parameter tends to zero.

At the genetic level, the differences between the groups, in addition to the specific allelic composition of SSR loci, characteristic of each sample of varieties, are also expressed in the degree of allelic diversity. The second group of varieties significantly exceeds the first group of varieties in terms of a number of genetic parameters that reflect the degree of allelic diversity. It can be assumed that the first group of varieties is represented by the most isolated part of the autochthonous gene pool of the Crimean Peninsula. In turn, the genetic diversity of the second group of varieties was influenced by the genetic contribution of the introduced gene plasma brought into the region from outside. An additional confirmation of this assumption is the assignment of the 'Chandler' variety, US selection, to the second group. The value of the fixation index in the first group of varieties is typical for populations in a state of panmixia and the absence of genetic barriers that increase the number of observed homozygotes. The data obtained in the Structure 2.3.4 program indicate a genetic relationship between the groups, expressed in the presence of samples with a comparable contribution of two clusters. Since the studied varieties are forms selected in the walnut population that exists within the borders of the Crimean Peninsula, it can be assumed that the high values of the observed heterozygosity reflect the significant size of the walnut gene pool, which contributes to panmixia. In general, the first group of walnut varieties is characterized by high yields and large fruits (more than 12 g).

Conclusion

Based on a comprehensive phenotypic assessment of samples from the collection of the walnut gene pool of the Nikitsky Botanical Gardens, groups of cultivars with a complex of economically valuable traits were identified. When performing microsatellite genotyping, a high level of genetic diversity and the presence of two genetically distinct groups of varieties were established. One of the groups includes predominantly large-nut cultivars with increased productivity potential, which actualizes the use of them as breeding material, and their genetic remoteness from the rest of the gene pool can increase the effect of heterosis during hybridization.

References

Antyufeev V.V., Vazhov V.I., Ryabov V.A. Reference Book on the Climate of the Steppe Department of the Nikitsky Botanical Garden. Yalta, 2022. (in Russian)

Aradhya M., Woeste K., Velasco D. Genetic diversity, structure and differentiation in cultivated walnut (*Juglans regia* L.). *Acta Hortic*. 2010;861:127-132. DOI 10.17660/ActaHortic.2010.861.16.

Balapanov I., Suprun I., Stepanov I., Tokmakov S., Lugovskoy A. Comparative analysis Crimean, Moldavian and Kuban Persian walnut collections genetic variability by SSR-markers. *Sci. Hortic.* 2019; 253:322-326. DOI 10.1016/j.scienta.2019.04.014.

Bernard A., Barreneche T., Lheureux F., Dirlewanger E. Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. *PLoS One.* 2018a;13(11): 0208021. DOI 10.1371/journal. pone.0208021.

Bernard A., Lheureux F., Dirlewanger E. Walnut: past and future of genetic improvement. *Tree Genet. Genomes.* 2018b;14:1. DOI 10.1007/s11295-017-1214-0.

Bernard A., Barreneche T., Donkpegan A., Lheureux F., Dirlewanger E. Comparison of structure analyses and core collections for the management of walnut genetic resources. *Tree Genet. Genomes.* 2020a; 16:76. DOI 10.1007/s11295-020-01469-5.

Bernard A., Marrano A., Donkpegan A., Brown P.J., Leslie C.A., Neale D.B., Lheureux F., Dirlewanger E. Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian walnut (*Juglans regia* L.). *BMC Genomics.* 2020b;21(1):20. DOI 10.21203/rs.2.18573/v1.

Bozhuyuk M.R., Ercisli S., Orhan E., Koc A. Determination of the genetic diversity of walnut (*Juglans regia* L.) cultivar candidates from northeastern Turkey using SSR markers. *Mitt. Klosterneubg.* 2020; 70(4):269-277.

Chen L.N., Ma Q.G., Chen Y.K., Wang B.Q., Pei D. Identification of major walnut cultivars grown in China based on nut phenotypes and SSR markers. *Sci. Hortic.* 2014;168:240-248. DOI 10.1016/j.scienta.2014.02.004.

Cseke K., Bujdosó G., Báder M., Mertl T., Benke A., Kámpel J.D. Genetic identification of hybrid walnuts (*Juglans × intermedia* Carr.) in Hungary, the hidden potential for future breeding. *Sustainability.* 2022;14(8):4782. DOI 10.3390/su14084782.

Dangl G.S., Woeste K., Aradhya M.K., Koehmstedt A., Simon C., Potter D., Leslie C.A., McGranahan G. Characterization of 14 microsatellite markers for genetic analysis and cultivars identification of walnut. *J. Am. Soc. Hortic. Sci*. 2005;130(3):348-354. DOI 10.21273/JASHS.130.3.348.

Davoodi F., Rezaei M., Heidari P., Hokmabadi H., Lawson S. Identification and DNA fingerprinting of some superior Persian walnut genotypes in Iran. *Erwerbs-Obstbau.* 2021;63:393-402. DOI 10.1007/s10341-021-00597-z.

Ebrahimi A., Fatahi R., Zamani Z. Analysis of genetic diversity among some Persian walnut genotypes (*Juglans regia* L.) using morphological traits and SSRs markers. *Sci. Hortic.* 2011;130(1):146-151. DOI 10.1016/j.scienta.2011.06.028.

Ebrahimi A., Zarei A., Lawson S., Woeste K.E., Smulders M.J.M. Genetic diversity and genetic structure of Persian walnut (*Juglans regia*) accessions from 14 European, African, and Asian countries using SSR markers. *Tree Genet. Genomes.* 2016;12:114. DOI 10.1007/s11295-016-1075-y.

Ebrahimi A., Zarei A., Fardadonbeh M.Z., Lawson S. Evaluation of genetic variability among "Early Mature" *Juglans regia* using microsatellite markers and morphological traits. *PeerJ.* 2017a;5:e3834. DOI 10.7717/peerj.3834.

Ebrahimi A., Zarei A.K., McKennac J.R., Bujdoso G., Woeste K.E. Genetic diversity of Persian walnut (*Juglans regia*) in the cold-temperate zone of the United States and Europe. *Sci. Hortic.* 2017b;220: 36-41. DOI 10.1016/j.scienta.2017.03.030.

Eremeev G.N., Lishchuk A.I. Guidelines for Selecting Drought-resistant Cultivars and Stocks of Fruit Plants. Yalta, 1974. (in Russian)

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.

Gaisberger H., Legay S., Loo A.C.J., Azimov R., Aaliev S., Bobokalonov F., Mukhsimov N., Kettle C., Vinceti B. Diversity under threat: connecting genetic diversity and threat mapping to set conservation priorities for *Juglans regia* L. populations in Central Asia. *Front. Ecol. Evol.* 2020;8:171. DOI 10.3389/fevo.2020.00171.

Guney M., Kafkas S., Keles H., Zarifikhosroshahi M., Gundesli M.A., Ercisli S., Necas T., Bujdoso G. Genetic diversity among some walnut (*Juglans regia* L.) genotypes by SSR markers. *Sustainability.* 2021;13(12):6830. DOI 10.3390/su13126830.

Gunn B.F., Aradhya M., Salick J.M., Miller A.J., Yongping Y., Lin L., Xian H. Genetic variation in walnuts (*Juglans regia* and *J. sigillata*; Juglandaceae): species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. *Am. J. Bot.* 2010; 97(4):660-671. DOI 10.3732/ajb.0900114.

Hammer Ø., Harper D.A.T., Ryan P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 2001;4(1):1-9.

Ikhsana A.S., Topc H., Sütyemez M., Kafkas S. Novel 307 polymorphic SSR markers from BAC-end sequences in walnut (*Juglans regia* L.): effects of motif types and repeat lengths on polymorphism and genetic diversity. *Sci. Hortic.* 2016;213:1-4. DOI 10.1016/j.scienta.2016.10.006.

Il'nitskiy O.A. Fundamentals of Phytomonitoring (Monitoring of Physiological Processes in Plants). Kherson, 2015. (in Russian)

Karimi R., Ershadi A., Ehtesham Nia A., Sharifani M., Rasouli M., Ebrahimi A., Vahdati K. Morphological and molecular evaluation of Persian walnut populations in Northern and Western regions of Iran. *J. Nuts*. 2014;5(2):21-31. DOI 10.22034/JON.2014.515686.

Khokhlov S.Yu. Study of varietals diversity of walnuts in the Crimea and the perspective of its use in selection. *Byulleten Gosudarstvennogo Nikitskogo Botanicheskogo Sada = Bulletin of the State Nikitsky Botanical Gardens*. 2012;105:57-61. (in Russian)

Khokhlov S.Yu., Baskakova V.L. Collection of circassian walnut. *Nauchnye Zapiski Prirodnogo Zapovednika "Mys Mart'yan" = Proceedings of the Cape Martyan Nature Reserve.* 2015;6:235-238. (in Russian)

Kushnirenko M.D., Kurchatova G.P., Kryukova E.V. Methods for Assessing the Drought Resistance of Fruit Plants. Chisinau, 1975. (in Russian)

Lapin N.I., Ryabova N.V. Some issues in the practice of woody plant introduction in botanical gardens. In: Study of Woody Plants During Introduction. Moscow: Nauka Publ., 1982;5-29. (in Russian)

Lugovskoy A.P., Balapanov I.M. Prospective persian walnut cultivars for North Caucasus zone and their biological characteristics. *Plodovodstvo i Vinogradarstvo Yuga Rossii = Fruit Growing and Viticulture of the South of Russia*. 2018;51(3):98-110. DOI 10.30679/2219-5335-2018-3-51-98-110. (in Russian)

Lugovskoi A.P., Murzinova D.G. Improving the sistem of walnut groving in Northern Caucasus. *Plodovodstvo i Vinogradarstvo Yuga Rossii = Fruit Growing and Viticulture of the South of Russia*. 2010; 6(5):15-23. (in Russian)

Magige E.A., Fan P.-Z., Wambulwa M.C., Milne R., Wu Z.-Y., Luo Y.-H., Khan R., Wu H.-Y., Qi H.-L., Zhu G.-F., Maity D., Khan I., Gao L.-M., Liu J. Genetic diversity and structure of Persian walnut (*Juglans regia* L.) in Pakistan: implications for conservation. *Plants.* 2022;11(13):1652. DOI 10.3390/plants11131652.

Orhan E., Eyduran S.P., Poljuha D., Akin M., Weber T., Ercisli S. Genetic diversity detection of seed-propagated walnut (*Juglans regia* L.) germplasm from Eastern Anatolia using SSR markers. *Folia Hort.* 2020;32(1):1-10. DOI 10.2478/fhort-2020-0004.

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.

Pollegioni P., Woeste K., Major A., Scarascia G., Malvolti M. Characterization of *Juglans nigra* (L.), *Juglans regia* (L.) and *Juglans × intermedia* (Carr.) by SSR markers: a case study in Italy. *Silvae Genet.* 2009;58(1):68-78. DOI 10.1515/sg-2009-0009.

Pollegioni P., Woeste K., Olimpieri I., Marandola D., Cannata F., Malvolti M.E. Long-term human impacts on genetic structure of Italian walnut inferred by SSR marker. *Tree Genet. Genomes.* 2011;7:707-723. DOI 10.1007/s11295-011-0368-4.

Pollegioni P., Woeste K.E., Chiocchini F., Olimpieri I., Tortolano V., Clark J., Hemery G.E., Mapelli S., Malvolti M.E. Landscape genetics of Persian walnut (*Juglans regia* L.) across its Asian range. *Tree Genet. Genomes.* 2014;10:1027-1043. DOI 10.1007/s11295-014-0740-2.

Pollegioni P., Woeste K.E., Chiocchini F., Del Lungo S., Olimpieri I., Tortolano V., Clark J., Hemery G.E., Mapelli S., Malvolti M.E. Ancient humans influenced the current spatial genetic structure of common walnut populations in Asia. *PLoS One.* 2015;10(9):e0135980. DOI 10.1371/journal. pone.0135980.

Pollegioni P., Woeste K., Chiocchini F., Del Lungo S., Ciolfi M., Olimpieri I., Tortolano V., Clark J., Hemery G.E., Mapelli S., Malvolti M. Rethinking the history of common walnut (*Juglans regia* L.) in Europe: its origins and human interactions. *PLoS One.* 2017;12(3):e0172541. DOI 10.1371/journal.pone.0172541.

Pop I.F., Vicol A.C., Botu M., Raica P.A., Vahdati K., Pamfil D. Relationships of walnut cultivars in a germplasm collection: comparative analysis of phenotypic and molecular data. *Sci. Hortic.* 2013;153: 124-135. DOI 10.1016/j.scienta.2013.02.013.

Rihter A.A., Yadrov A.A. Walnut. Moscow, 1985. (in Russian)

Rogers S.O., Bendich A.J. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant. Mol. Biol.* 1985;5:69-76. DOI 10.1007/BF00020088.

Roor D.W., Konrad H., Mamadjanov D., Geburek T. Population differentiation in common walnut (*Juglans regia* L.) across major parts of its native range – insights from molecular and morphometric data. *J. Hered.* 2017;108(4):391-404. DOI 10.1093/jhered/esw122.

Shah U.N., Mir J.I., Ahmed N., Fazili K.M. Assessment of germplasm diversity and genetic relationships among walnut (*Juglans regia* L.) genotypes through microsatellite markers. *J. Saudi Soc. Agric. Sci.* 2018;17(4):339-350. DOI 10.1016/j.jssas.2016.07.005.

Shamlu F., Rezaei M., Lawson S., Ebrahimi A., Biabani A., Khan-Ahmadi A. Genetic diversity of superior Persian walnut genotypes in Azadshahr, Iran. *Physiol. Mol. Biol. Plants.* 2018;24(5):939-949. DOI 10.1007/s12298-018-0573-9.

Suprun I.I., Lugovskoy A.P., Balapanov I.M. Introduction of new forms and updating the walnut gene pool as the basis of improvement of crop's assortment in the South of Russia. *Plodovodstvo i Vinogradarstvo Yuga Rossii = Fruit Growing and Viticulture of the South of Russia*. 2016;39(3):26-41. (in Russian)

Topcu H., Ikhsana A.S., Sütyemez M., Güneya N.M., Kafkas S. Development of 185 polymorphic simple sequence repeat (SSR) markers from walnut (*Juglans regia* L.). *Sci. Hortic.* 2015;194:160-167. DOI 10.1016/j.scienta.2015.08.014.

Vahdati K., Arab M.M., Sarikhani S., Sadat-Hosseini M., Leslie C.A., Brown P.J. Advances in Persian walnut (*Juglans regia* L.) breeding strategies. In: Al-Khayri J., Jain S., Johnson D. (Eds.) Advances in Plant Breeding Strategies: Nut and Beverage Crops. Cham: Springer, 2019;401-472. DOI 10.1007/978-3-030-23112-5_11.

Vischi M., Chiabà C., Raranciuc S., Poggetti L., Messina R., Ermacora P., Cipriani G., Paffetti D., Vettori C., Testolin R. Genetic diversity of walnut (*Juglans regia* L.) in the Eastern Italian Alps. *Forests.* 2017;8(3):81. DOI 10.3390/f8030081.

Wang H., Pan G., Ma Q., Zhang J., Pei D. The genetic diversity and introgression of *Juglans regia* and *Juglans sigillata* in Tibet as revealed by SSR markers. *Tree Genet. Genomes.* 2015;11:804. DOI 10.1007/s11295-014-0804-3.

Woeste K., Burns R., Rhodes O., Michler C. Thirty polymorphic nuclear microsatellite loci from black walnut. *J. Hered.* 2002;93:58-60. DOI 10.1093/jhered/93.1.58.

Zhou H., Zhao P., Woeste K., Zhang S. Gene flow among wild and cultivated common walnut (*Juglans regia*) trees in the Qinling Mountains revealed by microsatellite markers. *J. For. Res.* 2021;32(5): 2189-2201. DOI 10.1007/s11676-020-01254-z.

ORCID ID

I.I. Suprun orcid.org/0000-0003-0355-8395
I.V. Stepanov orcid.org/0000-0002-6251-300X

Acknowledgements. The work was carried out with the support from the Russian Scientific Foundation and Kuban Scientific Foundation (Project No. 22-16-20061, https://rscf.ru/en/project/22-16-20061/).

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

Received December 13, 2022. Revised April 19, 2023. Accepted April 20, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):463-470
DOI 10.18699/VJGB-23-56

Original Russian text https://vavilovj-icg.ru/

Candidate genes for domestication and resistance to cold climate according to whole genome sequencing data of Russian cattle and sheep breeds

N.S. Yudin[1], D.M. Larkin[2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Royal Veterinary College, University of London, London, United Kingdom
dmlarkin@gmail.com

Abstract. It is known that different species of animals, when living in the same environmental conditions, can form similar phenotypes. The study of the convergent evolution of several species under the influence of the same environmental factor makes it possible to identify common mechanisms of genetic adaptation. Local cattle and sheep breeds have been formed over thousands of years under the influence of domestication, as well as selection aimed at adaptation to the local environment and meeting human needs. Previously, we identified a number of candidate genes in genome regions potentially selected during domestication and adaptation to the climatic conditions of Russia, in local breeds of cattle and sheep using whole genome genotyping data. However, these data are of low resolution and do not reveal most nucleotide substitutions. The aim of the work was to create, using the whole genome sequencing data, a list of genes associated with domestication, selection and adaptation in Russian cattle and sheep breeds, as well as to identify candidate genes and metabolic pathways for selection for cold adaptation. We used our original data on the search for signatures of selection in the genomes of Russian cattle (Yakut, Kholmogory, Buryat, Wagyu) and sheep (Baikal, Tuva) breeds. We used the HapFLK, DCMS, FST and PBS methods to identify DNA regions with signatures of selection. The number of candidate genes in potentially selective regions was 946 in cattle and 151 in sheep. We showed that the studied Russian cattle and sheep breeds have at least 10 genes in common, apparently involved in the processes of adaptation/selection, including adaptation to a cold climate, including the *ASTN2*, *PM20D1*, *TMEM176A*, and *GLIS1* genes. Based on the intersection with the list of selected genes in at least two Arctic/Antarctic mammal species, 20 and 8 genes, have been identified in cattle and sheep, respectively, that are potentially involved in cold adaptation. Among them, the most promising for further research are the *ASPH*, *NCKAP5L*, *SERPINF1,* and *SND1* genes. Gene ontology analysis indicated the existence of possible common biochemical pathways for adaptation to cold in domestic and wild mammals associated with cytoskeleton disassembly and apoptosis.
Key words: signatures of selection; adaptation; cold; cattle; sheep; local breed; Russia; whole genome sequencing.

For citation: Yudin N.S., Larkin D.M. Candidate genes for domestication and resistance to cold climate according to whole genome sequencing data of Russian cattle and sheep breeds. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):463-470. DOI 10.18699/VJGB-23-56

Гены-кандидаты доместикации и устойчивости к холоду по данным полногеномного секвенирования российских пород крупного рогатого скота и овец

Н.С. Юдин[1], Д.М. Ларкин[2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Королевский Ветеринарный Колледж, Университет Лондона, Лондон, Великобритания
dmlarkin@gmail.com

Аннотация. Известно, что различные виды животных при обитании в одинаковых условиях среды могут сформировать сходные фенотипы. Изучение конвергентной эволюции нескольких видов под действием одного и того же средового фактора позволяет выявить у них общие механизмы генетической адаптации. Местные породы крупного и мелкого рогатого скота формировались на протяжении тысяч лет под воздействием доместикации, а также отбора, направленного на адаптацию к факторам местной среды обитания и удовлетворение потребностей человека. Ранее нами был выявлен ряд генов-кандидатов в участках генома, подвергшихся отбору в ходе доместикации и адаптации к климатическим условиям России, включая низкие зимние температуры, у местных пород крупного рогатого скота (КРС) и овец с использованием данных полногеномного генотипирования. Однако эти данные обладают низким разрешением и не позволяют выявить большинство нуклеотидных замен. Целью работы было создание по данным полногеномного секвенирования списка

генов, связанных с адаптацией российских пород КРС и овец, а также идентификация генов-кандидатов и метаболических путей для проведения селекции на адаптацию к холоду. Использованы опубликованные нами данные по поиску следов отбора в геномах российских или разводимых в России пород КРС (якутская, холмогорская, бурятская, вагю) и овец (забайкальская, тувинская). Количество генов-кандидатов в районах, потенциально подвергавшихся селекции, составило 946 у КРС и 151 у овец. Нами показано, что изученные российские породы КРС и овец имеют не менее 10 общих генов под отбором, по-видимому, участвующих в процессах адаптации/селекции, в том числе адаптации к холодному климату, включая гены *ASTN2*, *PM20D1*, *TMEM176A*, *GLIS1*. На основании пересечения со списком генов, подвергавшихся отбору по крайней мере у двух видов арктических/антарктических млекопитающих, у КРС и овец выявлено 20 и 8 генов соответственно, которые потенциально вовлечены в адаптацию к холоду. Среди них наиболее перспективными для дальнейших исследований являются *ASPH*, *NCKAP5L*, *SERPINF1* и *SND1*. Анализ генных онтологий указывает на существование возможных общих биохимических путей адаптации к холоду у домашних и диких млекопитающих, связанных с разборкой цитоскелета и апоптозом.
Ключевые слова: признаки селекции; адаптация; холод; крупный рогатый скот; овца; местная порода; Россия; полногеномное секвенирование

Introduction

The impact of extreme environmental factors can lead to either the extinction of a species or its adaptation to new environmental conditions (Nevo, 2011). It is known that different animal species, inhabiting similar conditions, can develop similar phenotypes using similar biochemical pathways (Storz, 2016). Studying convergent evolution of several species under the influence of the same environmental factor allows for the identification of common genetic adaptation mechanisms (Romashov et al., 2022).

For example, the same non-synonymous mutation in the rhodopsin gene independently arose and was subjected to selection in at least 20 species of fish in response to changes in water light conditions (Hill et al., 2019). The convergent amino acid substitutions also occurred in the prestin gene of whales and bats during the evolution of echolocation (Liu Y. et al., 2010). The non-synonymous substitution His207Arg in the melanocortin 1 receptor gene is associated with light feather coloration in red-footed boobies and ruffs (Lamichhaney et al., 2016). Our recent work on the northernmost cattle, the Yakut cattle from Siberia, revealed the phenomenon of convergent nucleotide evolution among domestic breeds and wild species living in similar harsh conditions and/or exhibiting similar phenotypes. We found the same amino acid substitution in the NRAP protein in Yakut cattle and 16 species of cold-adapted, hibernating or deep-diving mammals, which was absent in all other breeds of cattle and other Bovinae species in the "1000 Bull Genomes" dataset (Buggiotti et al., 2021). According to our data, this amino acid substitution presumably arose 500–800 years ago and is almost fixed in the modern Yakut cattle population.

It is known that the domestication of animals of different species is accompanied by a number of similar morphophysiological and behavioral changes (Belyaev, 1979; Wilkins et al., 2014). For example, one of the typical morphological features of domestication is the disruption of melanin synthesis, as well as a slowing down of melanocyte development, leading to the appearance of white spots on the body, up to the emergence of a uniform white color (Prasolova, Trut, 1993). Such phenotypic parallelism is observed in cattle, horses, pigs, dogs, cats, minks, chickens, pigeons, etc. (Larkin, Yudin, 2016). Indeed, when studying the genomes of populations of domestic animals, strong selection signals have been found in melanin metabolic pathway genes (*KIT*, *KITLG*, *MITF*, *PAX3*) (Cieslak et al., 2011).

Local breeds of cattle and sheep have been formed over thousands of years under the influence of domestication, as well as natural and artificial selection directed towards adaptation to the factors of the local environment and meeting human needs (Moiseeva et al., 2006; Kantanen et al., 2015). Studying the genomes of local breeds of cattle allows for the identification of genetic mechanisms of adaptation, including to low temperatures of the surrounding environment (Yudin et al., 2021). Earlier, we identified a number of candidate genes in genome regions that were potentially subject to selection during domestication and adaptation to harsh climatic conditions in Russia, in local breeds of cattle (*Bos taurus*) and sheep (*Ovis aries*) using data from whole-genome genotyping on standard SNP arrays (Yurchenko et al., 2018, 2019). Based on these results, we also identified 31 common candidate genes related to adaptation to the environment, including cold climate, in animals of the studied breeds (Yudin, Larkin, 2019). For example, the *NEB* gene, probably associated with heat production through shivering thermogenesis, was identified by us in genome regions subject to positive selection both in native Russian breeds of cattle and sheep, as well as in the genomes of the mammoth, polar bear, and minke whale.

However, whole-genome genotyping data have low resolution and do not allow the detection of most nucleotide substitutions in the genomes of different agricultural animal species. The aim of this study was to create a list of common genes associated with environmental adaptation in Russian breeds of cattle and sheep, as well as to identify promising genetic variants/candidate genes/metabolic pathways for further experiments, marker-assisted and genomic selection aimed at cold adaptation in agricultural animals, using whole-genome sequencing data. Previously, we analyzed selection signatures in the DNA samples from Yakut, Kholmogory, and Buryat cattle using GeneSeek Genomic Profiler High-Density SNP array containing approximately 139,000 SNPs (Yurchenko et al., 2018), and from Baikal and Tuva sheep using Ovine Infinium HD SNP BeadChip (Yurchenko et al., 2019).

Materials and methods

In the study, we used our own published data on the search for selection signatures using whole-genome sequencing in the genomes of Russian or bred in Russia cattle breeds (Yakut, Kholmogory, Buryat, Wagyu) (Buggiotti et al., 2021; Igoshin et al., 2023) and sheep (Baikal, Tuva) (Sweet-Jones et al., 2021). High-throughput sequencing was performed in paired-end mode (150 bp + 150 bp) on the Illumina platform at Novogene Co., Ltd. (Hong Kong, China) for 20 animals per breed. The average coverage depth was at least 11x for cattle and 15x for sheep.

To identify regions potentially under selection pressure in the genomes of Buryat and Wagyu cattle, we used four complementary methods (Igoshin et al., 2023). The hapFLK method is based on statistics that consider haplotype structure in populations (Fariello et al., 2013). The DCMS method combines five whole-genome statistics: Fisher's fixation index (FST), haplotype homozygosity (H1), modified haplotype homozygosity (H12), Tajima's D index (D), and nucleotide diversity index (Pi) (Ma et al., 2015). The FST method identifies genome regions subject to selection by identifying DNA segments with high allelic frequency variability between compared populations (Porto-Neto et al., 2013). The PBS statistic uses pairwise FST values between three populations to quantitatively assess sequence differentiation (Yi et al., 2010). It is considered that genes with high differentiation between sequences may potentially be under positive selection. Candidate gene lists for further analysis of Buryat and Wagyu cattle were compiled by combining lists obtained by different methods. Potential selection regions in the genomes of Yakut and Kholmogory cattle breeds were identified using hapFLK statistics (Buggiotti et al., 2021). For the search for selection signatures in the genomes of Baikal and Tuva sheep, a computational pipeline based on the DCMS method was used (Yurchenko et al., 2019).

Gene identifiers in the Ensembl database were converted into gene symbols using the db2db tool (http://biodbnet.abcc.ncifcrf.gov/db/db2dbRes.php?input=inputType&outputs[]=outputType&idList=value(s)). Intersections between gene lists were analyzed using the Venn program (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Biological functions of the shared genes that were under selection in Russian cattle and sheep breeds, and Arctic/Antarctic mammals were analyzed using the DAVID web tool (Huang et al., 2009). We identified enriched GO terms from the category of biological processes (GOTERM_BP_ALL) associated with four or more genes, compared to the control list of all human genes. We used a significance threshold criterion, characterized by the statistical significance of the observed number of genes with a specific GO term compared to the expected number of genes from the control list, and accepted $p < 0.05$ as the threshold value.

Results

The number of candidate genes in regions potentially subjected to selection was 946 for four Russian cattle breeds (*List_Cattle*, Suppl. Material 1)[1] and 151 for two Russian sheep breeds (*List_Sheep*, Suppl. Material 2) (see the Table). The difference in the number of candidate genes between species is likely due to differences in the number of breeds included in the analysis, as well as the number of statistical methods used to detect signatures of selection (four statistics for Buryat and Wagyu cattle, one for the other breeds). Analysis of the intersection of the lists showed that 10 genes could potentially have been under selection in both species (see the Figure, *Cattle_Sheep* list, Suppl. Material 3).

Previously, by intersecting the lists of genes potentially subjected to selection in six Arctic and Antarctic mammal species, we compiled a list of genes that may be involved in cold adaptation (Yudin et al., 2017). The list contained 416 genes that were likely under selection in at least two mammal species (*List_Mammals*, Suppl. Material 4). To identify common genes that may be associated with adaptation to cold climate in Russian cattle and sheep breeds, we compared the lists of *List_Mammals*, *List_Cattle*, and *List_Sheep*. As a result, we found 20 (*Cattle_Mammals*, Suppl. Material 5) and 8 (*Sheep_Mammals*, Suppl. Material 6) genes that were potentially under selection in at least two wild mammal species adapted to cold climate as well as in cattle and sheep, respectively (see the Figure).

To test the hypothesis that these lists were enriched in functional categories of genes related to cold adaptation, we performed gene ontology (GO) analysis on a list of 38 genes obtained by merging the *Cattle_Sheep*, *Cattle_Mammals*, and *Sheep_Mammals* lists (*Cattle_Sheep_Mammals*, Suppl. Material 7). As a result, we found significant enrichment in eight GO terms that were associated with 4 or more genes (Suppl. Material 8).

Discussion

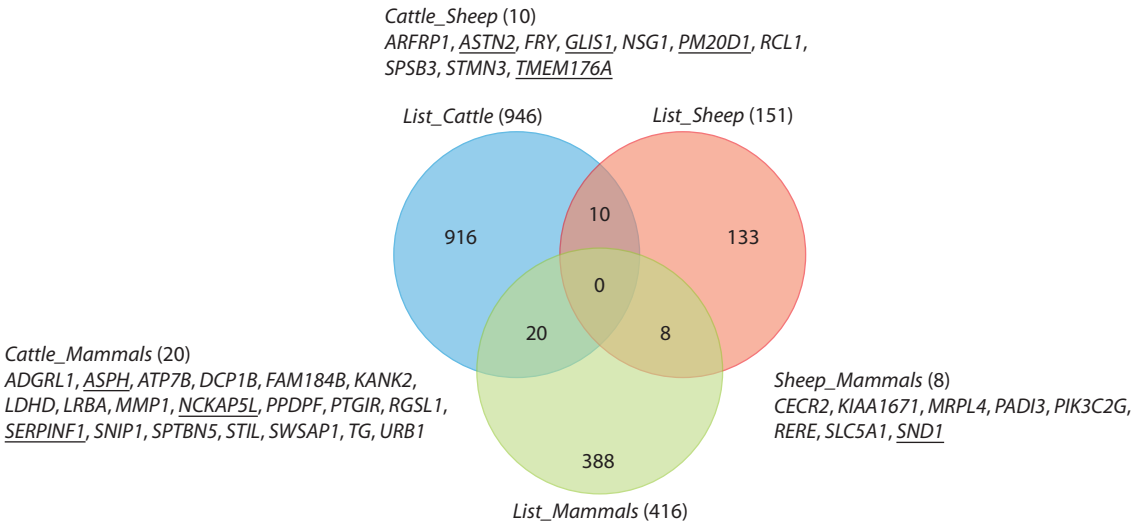
Our study aimed to identify common candidate genes in the genomes of domestic cattle and sheep breeds in Russia that may have undergone selection and played a role in adaptation to extreme climates, as well as to identify promising genetic variants/candidate genes/metabolic pathways for further cold adaptation research. We identified a total of 10 genes that potentially could have been under selection simultaneously in both Russian cattle and sheep breeds (*Cattle_Sheep* list, see Suppl. Material 3). These genes were likely subjected to selection during domestication and/or subsequent selection for economically important traits, as well as during adaptation to cold climates.

According to the theory of D.K. Belyaev, numerous morphophysiological transformations in domestic animals are caused by destabilizing selection for the absence of aggressive behavior towards humans (Belyaev, 1979). Indeed, we have previously shown that a list of 1262 common genes that underwent selection in Russian cattle and sheep breeds through whole-genome genotyping was enriched in genes predominantly expressed in the brain (Yudin, Larkin, 2019). Several common genes identified in our study (see Suppl. Material 3) are expressed in nervous tissue and are involved in normal neuron function. For example, the protein astrotactin 2 (ASTN2) modulates synaptic activity in neurons by regulating the expression of synaptic proteins in post-migratory neurons via endocytosis (Behesti et al., 2018). Genetic variants in the *ASTN2* gene are associated with Alzheimer's disease (Wang et al., 2015), schizophrenia (Autism Spectrum Disorders Working

[1] Supplementary Materials 1–8 are available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/Suppl_Yudin_Engl_27_5.pdf

Lists of potentially selected candidate genes

Name of the list in the Supplement	Description	Number of genes	Reference
List_Cattle (Suppl. Material 1)	List of genes potentially subjected to selection in the genomes of four Russian cattle breeds	946	Buggiotti et al., 2021; Igoshin et al., 2023
List_Sheep (Suppl. Material 2)	List of genes potentially subjected to selection in the genomes of two Russian sheep breeds	151	Sweet-Jones et al., 2021
Cattle_Sheep (Suppl. Material 3)	List of genes potentially subjected to selection in the genomes of Russian cattle and sheep breeds	10	–
List_Mammals (Suppl. Material 4)	List of genes potentially subjected to selection in the genomes of at least two species of Arctic/Antarctic mammals	416	Yudin et al., 2017
Cattle_Mammals (Suppl. Material 5)	List of genes potentially subjected to selection in the genomes of at least one Russian cattle breed, as well as at least two species of Arctic/Antarctic mammals	20	–
Sheep_Mammals (Suppl. Material 6)	List of genes potentially subjected to selection in the genomes of at least one Russian sheep breed, as well as at least two species of Arctic/Antarctic mammals	8	–
Cattle_Sheep_ Mammals (Suppl. Material 7)	List of genes potentially subjected to selection in the genomes of Russian cattle, sheep, as well as Arctic/Antarctic mammals – combination of the *Cattle_Sheep*, *Cattle_Mammals*, and *Sheep_Mammals* lists	38	–



Venn diagram showing overlaps between lists of genes potentially subjected to selection in the genomes of Russian cattle (*List_Cattle*) and sheep (*List_Sheep*) breeds, and at least two species of Arctic/Antarctic mammals (*List_Mammals*).

The number of genes in each list is indicated in parentheses. The most promising cold adaptation candidate genes (based on their biological role) are underlined.

Group of The Psychiatric Genomics Consortium, 2017), autism (Lionel et al., 2014), and other psychiatric disorders. The gene encoding the protein containing a peptidase domain M20 1 (*PM20D1*) is associated with Alzheimer's disease (Sanchez-Mut et al., 2018) and Parkinson's disease (Rudakou et al., 2021). The transmembrane protein *TMEM176A* gene is associated with schizophrenia (Kos et al., 2017).

At the same time, deletion in the *ASTN2* gene results in a reversal of normal orientation of hair follicles in adult mice (parallel to "from head to tail") to the opposite direction (parallel to "from tail to head") (Chang et al., 2015). In humans, the *ASTN2* gene is associated with the level of triglycerides in the blood (Jiao et al., 2015) and the development of obesity (Burt et al., 2021). Signatures of selection in this gene have been found in ethnic groups of southern Ethiopia, who have lived in high-altitude conditions for over a thousand years (Scheinfeldt et al., 2012). Interestingly, adaptive introgression of a large number of ancient Neanderthal alleles has been identified in the *ASTN2* gene in the population of South Asia (Racimo et al., 2017). The biochemical pathway of PM20D1

modulates the accumulation of brown fat and thus participates in the process of heat production through non-shivering thermogenesis (Gao et al., 2018). The pro-adipogenic factor GLIS1 may play a critical role in the differentiation of mesodermal cells during fetal development and affect fat distribution in the tail of sheep (Luo et al., 2021). SNP polymorphism in the *NSG1* gene is associated with the fat content in milk of Holstein cows (Lee et al., 2016).

We investigated promising genetic variants and candidate genes for cold adaptation by intersecting the lists *List_Cattle* and *List_Sheep* with the *List_Mammals* of 416 genes that were positively selected in at least two species of Arctic/Antarctic mammals (Yudin et al., 2017). When all three lists of common genes and genetic variants were intersected, none were found (see the Figure), but we identified 20 and 8 genes (see Suppl. Materials 5 and 6), respectively, that were potentially subjected to selection during adaptation to the climate of both Arctic mammals and cattle or sheep, respectively.

Thus, genetic variants in the *ASPH* gene, which encodes a protein that regulates the process of excitation–contraction in muscles, are associated with heat stroke and malignant hyperthermia in humans (Endo et al., 2022). According to whole-genome association analysis, single nucleotide polymorphisms in this gene are associated with intramuscular fat distribution in beef cattle (Ramayo-Caldas et al., 2014). Genetic variants in the *NCKAP5L* (Chen et al., 2013) and *SERPINF1* (Böhm et al., 2012) genes are associated with the development of obesity in humans. The human gene *FAM184B* is associated with body composition and fatty acid profile (Yuan et al., 2021). The protein product of the *PADI3* gene controls hair shape on the human scalp (Liu F. et al., 2018). The mRNA expression of the gene of the protein containing the staphylococcal nuclease domain 1 (*SND1*) in the New Zealand alpine stick insect significantly increases in response to cold exposure (Dunning et al., 2013). In mammals, SND1 plays an important role in regulating lipid metabolism through the activation of the SREBP2 protein (Navarro-Imaz et al., 2020).

The gene ontology terms identified by the DAVID program when analyzing the list of potentially selected genes in Russian cattle, sheep, as well as Arctic/Antarctic mammals (*Cattle_Sheep_Mammals* list) can be divided into three groups: (1) terms related to the disassembly of cell parts and protein complexes ("disassembly of cell components", "disassembly of protein complex", "disassembly of macromolecular complex", etc.); (2) terms related to DNA disintegration ("hydrolysis of phosphodiester bonds in nucleic acids"); (3) uninformative terms of the top hierarchy describing general biological processes ("biological process occurring at the level of a multicellular organism") (see Suppl. Material 8).

Enrichment of gene ontology terms related to the disassembly of cell parts, proteins, and DNA may be the result of natural selection for genes encoding cytoskeletal proteins and/or participating in programmed cell death (apoptosis). Studies on hibernating mammals have shown that their cells respond to low temperatures by disassembling the cytoskeleton and delaying apoptosis (Van Breukelen, Martin, 2002). It is believed that cytoskeletal disassembly may be the cause of protein synthesis suppression in mammalian cells during cold stress (Al-Fageeh, Smales, 2006). Hypothermia causes disassembly of microtubules by activating p38 MAP kinase in human retinal cells (Thanuja et al., 2021). In *in vivo* and *in vitro* experiments, it has been shown that microtubules in peripheral axons of *Xenopus* are sensitive to cold, and their density varies depending on the season (Alvarez, Fadić, 1992). It has been shown that cold stress induces apoptosis of neurons in the hippocampus of mice (Xu et al., 2019).

In our study, the term "disassembly of cellular components" was associated with seven genes (see Suppl. Material 8). Among them, the gene *SPTBN5* encodes one of the spectrin family proteins, which are common components of the cytoskeleton, interacting with elements of the cell scaffold and plasma membrane, providing proper localization of major membrane proteins, signal transmission into the cell, and other processes (Beijer, Züchner, 2022). The protein NCKAP5L, interacting with the protein CDK5RAP2, regulates microtubule stability in HeLa cells (Mori et al., 2015). The protein stathmin-3, encoded by the *STMN3* gene, regulates the rapid reorganization of the cytoskeleton in response to environmental factors by affecting the balance of microtubule assembly and disassembly (Nair et al., 2014). With the term "hydrolysis of phosphodiester bonds of nucleic acids", genes *RCL1*, *CECR2*, *SND1*, and *DCP1B* were associated (see Suppl. Material 8). It has been shown that the SND1 protein suppresses apoptosis in hepatocellular carcinoma cells by interacting with the long non-coding RNA UCA1 (Cui et al., 2018). The CECR2 protein is localized in the DNA condensation regions of apoptotic human liver cells and interacts with the chromatin-associated protein TAFII30 (Liu L. et al., 2002).

Conclusion

Thus, using whole-genome sequencing data, we have shown that the studied Russian cattle and sheep breeds have at least 10 common genes, presumably involved in adaptation/selection processes, including adaptation to cold climate, such as *ASTN2*, *PM20D1*, *TMEM176A*, *GLIS1*. Based on the overlap with the list of genes subjected to selection in at least two species of Arctic/Antarctic mammals, 20 and 8 genes potentially involved in adaptation to cold were identified in cattle and sheep, respectively. Among them, the most promising for further research are the genes *ASPH*, *NCKAP5L*, *SERPINF1*, and *SND1*. Gene ontology analysis indicates the existence of possible common biochemical pathways for adaptation to cold in domestic and wild mammals, related to cytoskeleton disassembly and apoptosis.

References

Al-Fageeh M.B., Smales C.M. Control and regulation of the cellular responses to cold shock: the responses in yeast and mammalian systems. *Biochem. J.* 2006;397(2):247-259. DOI 10.1042/BJ20060166.

Alvarez J., Fadić R. Assembly and disassembly of axonal microtubules of the toad *Xenopus laevis* under the effect of temperature. *J. Exp. Zool.* 1992;264(3):261-266. DOI 10.1002/jez.1402640305.

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism*. 2017;8:21. DOI 10.1186/s13229-017-0137-9.

Behesti H., Fore T.R., Wu P., Horn Z., Leppert M., Hull C., Hatten M.E. ASTN2 modulates synaptic strength by trafficking and degradation of surface proteins. *Proc. Natl. Acad. Sci. USA*. 2018;115(41):E9717-E9726. DOI 10.1073/pnas.1809382115.

Beijer D., Züchner S.L. Commentary: *SPTBN5*, encoding the βV-spectrin protein, leads to a syndrome of intellectual disability, developmental delay, and seizures. *Front. Mol. Neurosci.* 2022;15:1011856. DOI 10.3389/fnmol.2022.1011856.

Belyaev D.K. The Wilhelmine E. Key 1978 invitational lecture. Destabilizing selection as a factor in domestication. *J. Hered.* 1979;70(5):301-308. DOI 10.1093/oxfordjournals.jhered.a109263.

Böhm A., Ordelheide A.M., Machann J., Heni M., Ketterer C., Machicao F., Schick F., Stefan N., Fritsche A., Häring H.U., Staiger H. Common genetic variation in the *SERPINF1* locus determines overall adiposity, obesity-related insulin resistance, and circulating leptin levels. *PLoS One*. 2012;7(3):e34035. DOI 10.1371/journal.pone.0034035.

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.

Burt O., Johnston K.J.A., Graham N., Cullen B., Lyall D.M., Lyall L.M., Pell J.P., Ward J., Smith D.J., Strawbridge R.J. Genetic variation in the ASTN2 locus in cardiovascular, metabolic and psychiatric traits: evidence for pleiotropy rather than shared biology. *Genes* (*Basel*). 2021;12(8):1194. DOI 10.3390/genes12081194.

Chang H., Cahill H., Smallwood P.M., Wang Y., Nathans J. Identification of Astrotactin2 as a genetic modifier that regulates the global orientation of mammalian hair follicles. *PLoS Genet*. 2015;11(9):e1005532. DOI 10.1371/journal.pgen.1005532.

Chen J., Meng Y., Zhou J., Zhuo M., Ling F., Zhang Y., Du H., Wang X. Identifying candidate genes for Type 2 Diabetes Mellitus and obesity through gene expression profiling in multiple tissues or cells. *J. Diabetes Res*. 2013;2013:970435. DOI 10.1155/2013/970435.

Cieslak M., Reissmann M., Hofreiter M., Ludwig A. Colours of domestication. *Biol. Rev. Camb. Philos. Soc.* 2011;86(4):885-899. DOI 10.1111/j.1469-185X.2011.00177.x.

Cui X., Zhao C., Yao X., Qian B., Su C., Ren Y., Yao Z., Gao X., Yang J. SND1 acts as an anti-apoptotic factor via regulating the expression of lncRNA *UCA1* in hepatocellular carcinoma. *RNA Biol.* 2018;15(10):1364-1375. DOI 10.1080/15476286.2018.1534525.

Dunning L.T., Dennis A.B., Park D., Sinclair B.J., Newcomb R.D., Buckley T.R. Identification of cold-responsive genes in a New Zealand alpine stick insect using RNA-Seq. *Comp. Biochem. Physiol. Part D Genomics Proteomics*. 2013;8(1):24-31. DOI 10.1016/j.cbd.2012.10.005.

Endo Y., Groom L., Celik A., Kraeva N., Lee C.S., Jung S.Y., Gardner L., Shaw M.A., Hamilton S.L., Hopkins P.M., Dirksen R.T., Riazi S., Dowling J.J. Variants in *ASPH* cause exertional heat illness and are associated with malignant hyperthermia susceptibility. *Nat. Commun*. 2022;13(1):3403. DOI 10.1038/s41467-022-31088-8.

Fariello M.I., Boitard S., Naya H., SanCristobal M., Servin B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics.* 2013;193(3):929-941. DOI 10.1534/genetics.112.147231.

Gao Y., Qimuge N.R., Qin J., Cai R., Li X., Chu G.Y., Pang W.J., Yang G.S. Acute and chronic cold exposure differentially affects the browning of porcine white adipose tissue. *Animal*. 2018;12(7):1435-1441. DOI 10.1017/S1751731117002981.

Hill J., Enbody E.D., Pettersson M.E., Sprehn C.G., Bekkevold D., Folkvord A., Laikre L., Kleinau G., Scheerer P., Andersson L. Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proc. Natl. Acad. Sci. USA*. 2019;116(37):18473-18478. DOI 10.1073/pnas.1908332116.

Huang D.W., Sherman B.T., Lempicki R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2009;4(1):44-57. DOI 10.1038/nprot.2008.211.

Igoshin A.V., Romashov G.A., Yurchenko A.A., Yudin N.S., Larkin D.M. Resequencing commercial and locally adapted Turano-Mongolian cattle breeds reveals differences in selective pressures. *Animal Genetics*. 2023 (Submitted).

Jiao H., Wang K., Yang F., Grant S.F., Hakonarson H., Price R.A., Li W.D. Pathway-based genome-wide association studies for plasma triglycerides in obese females and normal-weight controls. *PLoS One*. 2015;10(8):e0134923. DOI 10.1371/journal.pone.0134923.

Kantanen J., Lovendahl P., Strandberg E., Eythorsdottir E., Li M.H., Kettunen-Praebel A., Berg P., Meuwissen T. Utilization of farm animal genetic resources in a changing agro-ecological environment in the Nordic countries. *Front. Genet*. 2015;6:52. DOI 10.3389/fgene.2015.00052.

Kos M.Z., Carless M.A., Peralta J., Curran J.E., Quillen E.E., Almeida M., Blackburn A., Blondell L., Roalf D.R., Pogue-Geile M.F., Gur R.C., Göring H.H.H., Nimgaonkar V.L., Gur R.E., Almasy L. Exome sequences of multiplex, multigenerational families reveal schizophrenia risk loci with potential implications for neurocognitive performance. *Am. J. Med. Genet. B Neuropsychiatr. Genet*. 2017;174(8):817-827. DOI 10.1002/ajmg.b.32597.

Lamichhaney S., Fan G., Widemo F., Gunnarsson U., Thalmann D.S., Hoeppner M.P., Kerje S., Gustafson U., Shi C., Zhang H., Chen W., Liang X., Huang L., Wang J., Liang E., Wu Q., Lee S.M., Xu X., Höglund J., Liu X., Andersson L. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet*. 2016;48(1):84-88. DOI 10.1038/ng.3430.

Larkin D.M., Yudin N.S. The genomes and history of domestic animals. *Mol. Genet. Microbiol. Virol.* 2016;31(4):197-202. DOI 10.3103/S0891416816040054.

Lee Y.S., Shin D., Lee W., Taye M., Cho K., Park K.D., Kim H. The prediction of the expected current selection coefficient of single nucleotide polymorphism associated with Holstein milk yield, fat and protein contents. *Asian-Australas. J. Anim. Sci*. 2016;29(1):36-42. DOI 10.5713/ajas.15.0476.

Lionel A.C., Tammimies K., Vaags A.K., Rosenfeld J.A., Ahn J.W., Merico D., Noor A., Runke C.K., Pillalamarri V.K., Carter M.T., … Ogilvie C.M., Shen Y., Hodge J.C., Talkowski M.E., Stavropoulos D.J., Marshall C.R., Scherer S.W. Disruption of the *ASTN2/TRIM32* locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. *Hum. Mol. Genet.* 2014;23(10):2752-2768. DOI 10.1093/hmg/ddt669.

Liu F., Chen Y., Zhu G., Hysi P.G., Wu S., Adhikari K., Breslin K., Pospiech E., Hamer M.A., Peng F., Muralidharan C., Acuna-Alonzo V., Canizales-Quinteros S., Bedoya G., Gallo C., Poletti G., Rothhammer F., Bortolini M.C., Gonzalez-Jose R., Zeng C., Xu S., Jin L., Uitterlinden A.G., Ikram M.A., van Duijn C.M., Nijsten T., Walsh S., Branicki W., Wang S., Ruiz-Linares A., Spector T.D., Martin N.G., Medland S.E., Kayser M. Meta-analysis of genome-wide association studies identifies 8 novel loci involved in shape variation of human head hair. *Hum. Mol. Genet*. 2018;27(3):559-575. DOI 10.1093/hmg/ddx416.

Liu L., Amy V., Liu G., McKeehan W.L. Novel complex integrating mitochondria and the microtubular cytoskeleton with chro-

mosome remodeling and tumor suppressor RASSF1 deduced by in silico homology analysis, interaction cloning in yeast, and colocalization in cultured cells. *In Vitro Cell. Dev. Biol. Anim.* 2002;38(10):582-594. DOI 10.1290/1543-706x(2002)38<582:ncimat>2.0.co;2.

Liu Y., Cotton J.A., Shen B., Han X., Rossiter S.J., Zhang S. Convergent sequence evolution between echolocating bats and dolphins. *Curr. Biol.* 2010;20(2):R53-R54. DOI 10.1016/j.cub.2009.11.058.

Luo R., Zhang X., Wang L., Zhang L., Li G., Zheng Z. *GLIS1*, a potential candidate gene affect fat deposition in sheep tail. *Mol. Biol. Rep*. 2021;48(5):4925-4931. DOI 10.1007/s11033-021-06468-w.

Ma Y., Ding X., Qanbari S., Weigend S., Zhang Q., Simianer H. Properties of different selection signature statistics and a new strategy for combining them. *Heredity* (*Edinb*). 2015;115(5):426-436. DOI 10.1038/hdy.2015.42.

Moiseeva I.G., Ukhanov S.V., Stolpovsky Yu.A., Sulimova G.E., Kashtanov S.N. Gene Pools of Farm Animals. Livestock Genetic Resources in Russia. Moscow: Nauka Publ., 2006. (in Russian)

Mori Y., Inoue Y., Tanaka S., Doda S., Yamanaka S., Fukuchi H., Terada Y. Cep169, a novel microtubule plus-end-tracking centrosomal protein, binds to CDK5RAP2 and regulates microtubule stability. *PLoS One.* 2015;10(10):e0140968. DOI 10.1371/journal.pone.0140968.

Nair S., Bora-Singhal N., Perumal D., Chellappan S. Nicotine-mediated invasion and migration of non-small cell lung carcinoma cells by modulating *STMN3* and *GSPT1* genes in an ID1-dependent manner. *Mol. Cancer.* 2014;13:173. DOI 10.1186/1476-4598-13-173.

Navarro-Imaz H., Ochoa B., García-Arcos I., Martínez M.J., Chico Y., Fresnedo O., Rueda Y. Molecular and cellular insights into the role of SND1 in lipid metabolism. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids*. 2020;1865(5):158589. DOI 10.1016/j.bbalip.2019.158589.

Nevo E. Evolution under environmental stress at macro- and micro-scales. *Genome Biol. Evol.* 2011;3:1039-1052. DOI 10.1093/gbe/evr052.

Porto-Neto L.R., Lee S.H., Lee H.K., Gondro C. Detection of signatures of selection using F_{st}. In: Gondro C., van der Werf J., Hayes B. (Eds.) Genome-Wide Association Studies and Genomic Prediction. Methods in Molecular Biology. Totowa, NJ: Humana Press, 2013;1019:423-436. DOI 10.1007/978-1-62703-447-0_19.

Prasolova L.A., Trut L.N. The effect of the *Star* gene on the rate of melanoblast migration in silver fox embryos. *Doklady RAN = Proceedings of the Russian Academy of Sciences*. 1993;329(6):787-789. (in Russian)

Racimo F., Marnetto D., Huerta-Sánchez E. Signatures of archaic adaptive introgression in present-day human populations. *Mol. Biol. Evol*. 2017;34(2):296-317. DOI 10.1093/molbev/msw216.

Ramayo-Caldas Y., Fortes M.R., Hudson N.J., Porto-Neto L.R., Bolormaa S., Barendse W., Kelly M., Moore S.S., Goddard M.E., Lehnert S.A., Reverter A. A marker-derived gene network reveals the regulatory role of *PPARGC1A*, *HNF4G*, and *FOXP3* in intramuscular fat deposition of beef cattle. *J. Anim. Sci*. 2014;92(7):2832-2845. DOI 10.2527/jas.2013-7484.

Romashov G., Yurchenko A., Yudin N., Larkin D. How does a harsh environment make wild and domestic animals evolve similarly? In: Veerkamp R.F., de Haas Y. (Eds.) Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP). Rotterdam, The Netherlands, 3–8 July 2022. Wageningen: Wageningen Acad. Publishers, 2022;1078-1081. DOI 10.3920/978-90-8686-940-4_255.

Rudakou U., Yu E., Krohn L., Ruskey J.A., Asayesh F., Dauvilliers Y., Spiegelman D., Greenbaum L., Fahn S., Waters C.H., Dupré N., Rouleau G.A., Hassin-Baer S., Fon E.A., Alcalay R.N., Gan-Or Z. Targeted sequencing of Parkinson's disease loci genes highlights *SYT11*, *FGF20* and other associations. *Brain.* 2021;144(2):462-472. DOI 10.1093/brain/awaa401.

Sanchez-Mut J.V., Heyn H., Silva B.A., Dixsaut L., Garcia-Esparcia P., Vidal E., Sayols S., Glauser L., Monteagudo-Sánchez A., Perez-Tur J., Ferrer I., Monk D., Schneider B., Esteller M., Gräff J. *PM20D1* is a quantitative trait locus associated with Alzheimer's disease. *Nat. Med.* 2018;24(5):598-603. DOI 10.1038/s41591-018-0013-y.

Scheinfeldt L.B., Soi S., Thompson S., Ranciaro A., Woldemeskel D., Beggs W., Lambert C., Jarvis J.P., Abate D., Belay G., Tishkoff S.A. Genetic adaptation to high altitude in the Ethiopian highlands. *Genome Biol*. 2012;13(1):R1. DOI 10.1186/gb-2012-13-1-r1.

Storz J.F. Causes of molecular convergence and parallelism in protein evolution. *Nat. Rev. Genet.* 2016;17(4):239-250. DOI 10.1038/nrg.2016.11.

Sweet-Jones J., Yurchenko A.A., Igoshin A.V., Yudin N.S., Swain M.T., Larkin D.M. Resequencing and signatures of selection scan in two Siberian native sheep breeds point to candidate genetic variants for adaptation and economically important traits. *Anim. Genet.* 2021;52(1):126-131. DOI 10.1111/age.13015.

Thanuja M.Y., Suma B.S., Dinesh D., Ranganath S.H., Srinivas S.P. Microtubule stabilization protects hypothermia-induced damage to the cytoskeleton and barrier integrity of the corneal endothelial cells. *J. Ocul. Pharmacol. Ther.* 2021;37(7):399-411. DOI 10.1089/jop.2021.0036.

Van Breukelen F., Martin S.L. Invited review: molecular adaptations in mammalian hibernators: unique adaptations or generalized responses? *J. Appl. Physiol.* 2002;92(6):2640-2647. DOI 10.1152/japplphysiol.01007.2001.

Wang K.S., Tonarelli S., Luo X., Wang L., Su B., Zuo L., Mao C., Rubin L., Briones D., Xu C. Polymorphisms within *ASTN2* gene are associated with age at onset of Alzheimer's disease. *J. Neural. Transm.* (*Vienna*). 2015;122(5):701-708. DOI 10.1007/s00702-014-1306-z.

Wilkins A.S., Wrangham R.W., Fitch W.T. The "domestication syndrome" in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*. 2014;197(3):795-808. DOI 10.1534/genetics.114.165423.

Xu B., Lang L.M., Li S.Z., Guo J.R., Wang J.F., Wang D., Zhang L.P., Yang H.M., Lian S. Cortisol excess-mediated mitochondrial damage induced hippocampal neuronal apoptosis in mice following cold exposure. *Cells.* 2019;8(6):612. DOI 10.3390/cells8060612.

Yi X., Liang Y., Huerta-Sanchez E., Jin X., Cuo Z.X., Pool J.E., Xu X., Jiang H., Vinckenbosch N., Korneliussen T.S., … Zhang X., Li R., Li S., Yang H., Nielsen R., Wang Jun, Wang Jian. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science.* 2010;329(5987):75-78. DOI 10.1126/science.1190371.

Yuan Z., Sunduimijid B., Xiang R., Behrendt R., Knight M.I., Mason B.A., Reich C.M., Prowse-Wilkins C., Vander Jagt C.J., Chamberlain A.J., MacLeod I.M., Li F., Yue X., Daetwyler H.D. Expression quantitative trait loci in sheep liver and muscle contribute to variations in meat traits. *Genet. Sel. Evol*. 2021;53(1):8. DOI 10.1186/s12711-021-00602-9.

Yudin N.S., Larkin D.M. Shared signatures of selection related to adaptation and acclimation in local cattle and sheep breeds from Russia. *Russ. J. Genet*. 2019;55(8):1008-1014. DOI 10.1134/S1022795419070159.

Yudin N.S., Larkin D.M., Ignatieva E.V. A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. *BMC Genet.* 2017;18(Suppl.1):111. DOI 10.1186/s12863-017-0580-9.

Yudin N.S., Yurchenko A.A., Larkin D.M. Signatures of selection and candidate genes for adaptation to extreme environmental factors in the genomes of Turano-Mongolian cattle breeds. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(2):190-201. DOI 10.18699/VJ21.023. (in Russian)

Yurchenko A.A., Daetwyler H.D., Yudin N., Schnabel R.D., Vander Jagt C.J., Soloshenko V., Lhasaranov B., Popov R., Taylor J.F., Larkin D.M. Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Sci. Rep.* 2018;8(1):12984. DOI 10.1038/s41598-018-31304-w.

Yurchenko A.A., Deniskova T.E., Yudin N.S., Dotsev A.V., Khamiruev T.N., Selionova M.I., Egorov S.V., Reyer H., Wimmers K., Brem G., Zinovieva N.A., Larkin D.M. High-density genotyping reveals signatures of selection related to acclimation and economically important traits in 15 local sheep breeds from Russia. *BMC Genomics.* 2019;20(Suppl.3):294. DOI 10.1186/s12864-019-5537-0.

ORCID ID
Yudin N.S. orcid.org/0000-0002-1947-5554
Larkin D.M. orcid.org/0000-0001-7859-6201

Acknowledgements. This work was funded by the Russian Scientific Foundation (RSF) grant No. 19-76-20026.

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

Received December 2, 2022. Revised February 19, 2023. Accepted February 20, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):471-479
DOI 10.18699/VJGB-23-57

Identification of homozygosity-rich regions in the Holstein genome

M.G. Smaragdov

Russian Research Institute of Farm Animal Genetics and Breeding – Branch of the L.K. Ernst Federal Science Center for Animal Husbandry, St. Petersburg, Pushkin, Russia
mik7252@yandex.ru

Abstract. In this study, 371 Holstein cows from six herds and 26 Holstein bulls, which were used in these herds, were genotyped by the Illumina BovineSNP50 array. For runs of homozygosity (ROH) identification, consecutive and sliding runs were performed by the detectRUNS and Plink software. The missing calls did not significantly affect the ROH data. The mean number of ROH identified by consecutive runs was 95.4 ± 2.7, and that by sliding runs was 86.0 ± 2.6 in cows, while this number for Holstein bulls was lower 58.9 ± 1.9. The length of the ROH segments varied from 1 Mb to over 16 Mb, with the largest number of ROH having a length of 1–2 Mb. Of the 29 chromosomes, BTA 14, BTA 16, and BTA 7 were the most covered by ROH. The mean coefficient of inbreeding across the herds was 0.111 ± 0.003 and 0.104 ± 0.004 based on consecutive and sliding runs, respectively, and 0.078 ± 0.005 for bulls based on consecutive runs. These values do not exceed those for Holstein cattle in North America. The results of this study confirmed the more accurate identification of ROH by consecutive runs, and also that the number of allowed heterozygous SNPs may have a significant effect on ROH data.
Key words: ROH; SNP; inbreeding; cattle.

For citation: Smaragdov M.G. Identification of homozygosity-rich regions in the Holstein genome. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):471-479. DOI 10.18699/VJGB-23-57

Идентификация гомозиготно обогащенных участков в геноме голштинов

М.Г. Смарагдов

Всероссийский научно-исследовательский институт генетики и разведения сельскохозяйственных животных – филиал Федерального исследовательского центра животноводства – ВИЖ им. академика Л.К. Эрнста (ВНИИГРЖ), Санкт-Петербург, Пушкин, Россия
mik7252@yandex.ru

Аннотация. В настоящем исследовании 371 корова голштинской породы из шести стад и 26 быков голштинской породы, которые использовались в этих стадах, были генотипированы с помощью чипа Illumina BovineSNP50. Для идентификации гомозиготных последовательностей (ROH) выполнялись последовательные и скользящие сканирования с помощью программ detectRUNS и Plink. Пропущенные SNP генотипы не оказали существенного влияния на данные ROH. Среднее количество ROH, выявленное у коров при последовательных сканированиях, составило 95.4 ± 2.7, при скользящих сканированиях – 86.0 ± 2.6, тогда как у быков голштинской породы оно было меньше – 58.9 ± 1.9. Длина сегментов ROH варьировала от 1 до 16 Мб и более, при этом основное количество ROH имело длину 1–2 Мб. Из 29 хромосом наиболее насыщены ROH оказались BTA 14, BTA 16 и BTA 7. Средний коэффициент инбридинга по стадам составил 0.111 ± 0.003 и 0.104 ± 0.004 на основе последовательных и скользящих сканирований соответственно, а для быков на основе последовательных сканирований – 0.078 ± 0.005. Эти значения не превышали показатели для голштинского скота в Северной Америке. Результаты исследования подтвердили более точную идентификацию ROH последовательными сканированиями, а также то, что количество разрешенных гетерозиготных SNP в ROH может оказывать существенное влияние на данные ROH.
Ключевые слова: ROH; SNP; инбридинг; крупный рогатый скот.

Introduction

Inbreeding in dairy cattle is an inevitable phenomenon of artificial selection. Traditionally, the inbreeding coefficient is calculated based on ancestry (Meuwissen, Luo, 1992). With the advent of SNPs arrays (Matukumalli et al., 2009), it became possible to investigate autozygosity at a previously unattainable level (Peripolli et al., 2016). In fact, due to the runs of homozygosity (ROH) approach, animal genome analysis for long continuous homozygous stretches is still ongoing. The primary cause of autozygosity in livestock measured by ROH is assumed to be inbreeding (Peripolli et al., 2016) or consanguineous marriage in humans (Ceballos et al., 2018b). For identifying ROH, software based either on identity by descent (IBD) GERMLINE (Gusev et al., 2009), or on Hidden Markov Model (HMM) Beagle (Browning S., Browning B., 2010) and BCFtools (Narasimhan et al., 2016) has been elaborated. In addition, software based on scanning by SNPs window Plink (Purcell et al., 2007), overlapping sliding window SNP101 (Forutan et al., 2018), or both consecutive and sliding runs detectRUNS (Biscarini et al., 2018), as well the software based on other scripts (Howard et al., 2015; Kim et al., 2015), cgaTON (Zhang L. et al., 2013) have been

provided. The commercial software SNP & Variation Suite (Golden Helix SNP & Variation Suite) is also widely used.

It has been shown that software based on HMM and IBD is inferior to other software mentioned above (Howrigan et al., 2011). The main challenge facing scientists is the lack of consistent criteria among studies regarding a threshold value of each parameter analyzed to determine ROH (Peripolli et al., 2016). The most crucial parameters that are used in any software are the number of heterozygous or missing SNP calls allowed in ROH. There is an inconsistency between thresholds that should be applied in studies. Some authors disallowed any number of heterozygous SNPs in ROH (Ferencakovic et al., 2011; Purfield et al., 2012; Bjelland et al., 2013; Marras et al., 2014), others allowed one, two and more heterozygous SNPs depending on the length of the ROH segments (Ferenčaković et al., 2013; Karimi, 2013; Zavarez et al., 2015; Zhang Q. et al., 2015a; Mastrangelo et al., 2016; Ceballos et al., 2018a; Addo et al., 2019; Zinovieva et al., 2020). Anyway, M. Ferenčaković et al. (2013) suggested that allowing a certain amount of genotype errors in a long ROH could minimize the underestimation of these segments. Although S. Mastrangelo et al. (2016) showed different values of the inbreeding coefficient, if heterozygous genotypes were allowed.

There are relatively few studies assessing which set of these parameters is optimal for detecting ROH, in order to better understand their effect on identified autozygosity. M. Ferenčaković et al. (2013) have shown that SNP array density and genotyping errors introduce patterns of bias in the assessment of autozygosity. These authors observed that allowing heterozygous SNPs in ROH can lead to the merging of adjacent ROH segments which resulted in biased estimates of the ROH number. Based on simulation data, D. Howrigan et al. (2011) recommended disallowing existence of any heterozygous SNPs in ROH. Summarizing, there is currently no consensus on a reasonable number of heterozygous SNPs in ROH to avoid bias in the ROH data.

When planning this study, special attention was paid to assessing the impact of allowed missing SNPs and heterozygous SNPs in ROH runs on the results using commonly applied consecutive and sliding runs. Another goal of the study was to evaluate the distribution of ROH in the chromosomes, and the effect of allowed heterozygous SNPs on inbreeding scores.

The following main objectives of the study were: (i) to assess the number and length of ROH segments in the cows and bulls genome, as well as their proportion in the chromosomes, (ii) to calculate the inbreeding coefficient, (iii) to assess the data bias resulting from an allowance of missing and heterozygous SNPs in ROH, (iv) to use the sliding windows and consecutive runs to obtain ROH data.

Materials and methods

Animal resources and SNPs genotyping. Data and genotypes were obtained from Committee on Agro-Industrial Complex of the Leningrad region. This study analyzed Holstein cows born from 2010 to 2013 in six herds located in the Leningrad region (Russia). More information on breeding our local Holstein cattle can be found in the article (Kudinov et al., 2022).

Animals for genotyping were selected by farmers with regard to the pedigree structure of the herd. The sampled animals accounted for 8–15 % of the total number of dairy cows in herds. Altogether, 371 cows from six herds and 26 bulls from the Netherlands, North America, Germany and Canada used in these herds were genotyped by BovineSNP50 v. 2.0 array (Illumina, USA). Quality control was carried out by Plink. (i) SNPs calls with a quality score of less than 0.7 were removed. (ii) Only autosomal chromosomes were considered. (iii) 5 % of missed SNPs and 1 % MAF were allowed, which resulted in 48,108 SNPs for cows and 43,441 for bulls. Total genotyping rate was > 0.99.

Identification of ROH. The ROH segments were identified using detectRUNS (Biscarini et al., 2018) implemented in the R environment (http://www.r-project.org/index.html), and Plink tool (Purcell et al., 2007). The parameters applied to define ROH by detectRUNS for consecutive runs method were: (i) the minimum number of SNPs required to define segments as ROH, 15 and 20, (ii) the number of missing calls allowed in a ROH segment, 0–4, (iii) the number of heterozygous calls allowed in a ROH segment, 0–2, (iv) the minimum length of ROH segments, 250 Kb, (v) the maximum gap between ROH segments, 1 Mb.

For sliding window method in detectRUNS the parameters and thresholds were: (i) window size 15 and 20 SNPs, (ii) the threshold 0.05, (iii) the minimum number of SNPs required to define segments as ROH, 15 and 20, (iv) the number of missing calls allowed in a ROH segment, 0–4, (v) the number of heterozygous calls allowed in a ROH segment, 0–2, (vi) the minimum length of ROH, 250 Kb, (vii) the maximum gap between ROH segments, 1 Mb, (viii) the minimum allowed density of SNPs, 1 SNP per 1 Mb.

The parameters applied to define ROH by Plink were (i) the sliding window, 20 SNPs, (ii) the proportion of homozygous overlapping windows, 0.05, (iii) the minimum number of SNPs in ROH, 20, (iv) the density was one SNP per 60 Kb, (v) the number of missing SNPs was zero, (vi) the number of heterozygous SNPs was zero.

Inbreeding coefficients (F_{ROH}) were calculated as the sum of the animal's ROH lengths divided by the total length of the autosomes covered by SNPs (2508.706681 Mb).

Results

Impact of missing SNPs on ROH data. Primarily, the effect of missing SNPs allowed in ROH on the data was evaluated by consecutive and sliding runs. No impact on ROH data was found for either method if one to four missing SNP calls were allowed in ROH. Therefore, to further evaluate the ROH results, this value was set to zero.

Effect of heterozygous SNPs on ROH data based on consecutive runs. To evaluate the number of ROH segments in the cow genome, 15 SNPs (Suppl. Material 1)[1] and 20 SNPs (Table 1) consecutive runs were carried out. When ROH segments were not interrupted by heterozygous SNPs, the mean number of ROH was 1.9 times greater at 15 SNPs runs ($p \leq 0.03$). In fact, the average number of ROH across the herds was 182.1 ± 3.4 at 15 SNPs runs compared to 95.4 ± 2.7 at 20 SNPs runs. To avoid overestimation of the autozygous ROH due to short ROH segments, 20 SNPs runs were used further.

[1] Supplementary Materials 1–4 are available in the online version of the paper: https://vavilovj-icg.ru/download/pict-2023-27/appx17.pdf

Table 1. Estimated mean ROH number (± SE) in the herds based on 20 SNPs consecutive runs (detectRUNS)

ROH number	Herd						
	1	2	3	4	5	6	Mean
Zero heterozygous SNPs in ROH							
The mean number of ROH	99.6 ± 6.5	90.5 ± 1.0	91.5 ± 1.4	100.4 ± 14.7	93.0 ± 1.3	91.4 ± 1.5	95.4 ± 2.7
Maximum	360	112	148	757	111	125	
Minimum	2	66	65	48	75	73	
One heterozygous SNP in ROH							
The mean number of ROH	161.9 ± 10.2	145.4 ± 1.3	146.5 ± 1.6	155.8 ± 12.1	149.2 ± 1.5	148.8 ± 1.6	151.3 ± 2.7
Maximum	565	179	195	692	175	174	
Minimum	7	117	109	82	125	126	
Two heterozygous SNPs in ROH							
The mean number of ROH	262.2 ± 13.1	243.1 ± 1.6	244.7 ± 1.8	253.7 ± 7.4	245.5 ± 1.0	248.2 ± 1.8	249.6 ± 2.6
Maximum	761	277	281	564	270	283	
Minimum	21	211	215	145	214	207	

Table 2. Estimated mean ROH number (± SE) across the herds based on sliding runs (detectRUNS)

ROH number	Herd						
	1	2	3	4	5	6	Mean
Zero heterozygous SNPs in ROH (20 SNPs sliding runs)							
The mean number of ROH	91 ± 6	82 ± 1	83 ± 1	92 ± 1	85 ± 1	83 ± 1	86.0 ± 2.6
Maximum	336	194	138	731	106	111	
Minimum	1	62	59	44	67	64	
One heterozygous SNP in ROH (20 SNPs sliding runs)							
The mean number of ROH	1579 ± 9	142 ± 1	143 ± 1	146 ± 5	147 ± 1	145 ± 1	146.7 ± 1.7
Maximum	513	174	175	335	175	169	
Minimum	5	117	114	82	118	120	
Two heterozygous SNPs in ROH (20 SNPs sliding runs)							
The mean number of ROH	268 ± 10	255 ± 1	254 ± 2	256 ± 4	255 ± 2	257 ± 2	257.5 ± 1.9
Maximum	631	284	287	294	285	288	
Minimum	21	225	220	109	237	222	
Zero heterozygous SNPs (15 SNPs sliding runs)							
The mean number of ROH	190.7 ± 11.2	175.5 ± 1.5	178.2 ± 1.7	191.6 ± 16.7	178.6 ± 1.7	179.8 ± 2.0	182.1 ± 3.4
Maximum	635	207	243	934	214	217	
Minimum	12	143	150	105	151	147	

For an adequate understanding of the results, it is necessary to define the term ROH further used. ROH is a contiguous homozygous SNP sequence uninterrupted by heterozygous SNPs, except for the allowed number of heterozygous SNP. Descriptive data statistics are given in Table 1. The mean number of ROH varied across herds. However, the differences between them are insignificant (*t*-test). It should be noted that there is considerable variation in the ROH number among the fourth herd cows. This result was due to a large number of ROH in one cow (757 ROH segment). The exclusion of this cow resulted in the mean ROH of 85.9 ± 2.1 in the fourth herd. However, this did not lead to significant differences between the herds (*t*-test). The effect of allowed heterozygous SNPs on the number and length of the ROH segments was assessed when their values ranged from 0 to 2. Initially, the average number of ROH increased more than 1.6-fold from 95.4 ± 2.7 to 151.3 ± 2.7 when one heterozygous SNP in ROH was allowed (see Table 1). Then the mean increased to 249.6 ± 2.6 with an increase in the number of allowed heterozygous SNPs in ROH to two. Thus, the allowance of heterozygous SNPs leads to a significant ($p \leq 0.02$) increase in the number of ROH.

The length of the ROH segments has been classified into five categories (1–2 Mb, 2–4, 4–8, 8–16, and >16 Mb). The most abundant in the number of ROH was the 1–2 Mb class (Suppl. Material 2). The largest proportion of the ROH number had the same class, up to two allowed heterozygous SNPs. A particularly noticeable increase in the number of ROH in this class occurred with the use of 15 SNP runs (see Suppl. Material 2). These data indicate the presence in the genome of cows of a large number of short (less than 1 Mb) ROH segments, which are more effectively detected when scanning for 15 SNPs.

ROH identification based on sliding runs. As for 15 SNPs and 20 SNPs, the sliding runs were used (Table 2). Interest-

Table 3. Rank of the cows chromosomes by their ROH coverage

BTA*	14	16	7	26	8	13	1	17	4	20	19	6	22	21	24
Consecutive runs**	1.329	1.292	1.236	1.191	1.187	1.119	1.116	1.054	1.048	1.024	1.020	1.011	0.980	0.961	0.960
Sliding runs	1.388	1.295	1.218	1.170	1.169	1.161	1.086	1.050	1.030	0.980	1.010	0.999	1.011	0.965	0.955
BTA	3	11	10	12	2	5	9	25	29	23	15	18	27	28	
Consecutive runs	0.955	0.923	0.916	0.914	0.914	0.911	0.896	0.876	0.844	0.806	0.789	0.758	0.749	0.676	
Sliding runs	0.963	0.909	0.931	0.918	0.898	0.911	0.902	0.920	0.865	0.851	0.800	0.743	0.774	0.680	

* *Bos taurus* autosome.
** The rank values were ranged from maximum to minimum only for consecutive runs.

ingly, the data for 15 SNPs runs identified by both consecutive and sliding runs were largely the same (see Suppl. Material 1 *vs.* Table 2 (15 SNPs)), while for 20 SNPs consecutive and 20 SNPs sliding runs, the data differ somewhat, but insignificantly (*t*-test) (see Tables 1 and 2). To obtain a comparable result with consecutive runs, 20 SNPs window was used further. Descriptive statistic for 20 SNPs sliding data is given in Table 2. The mean number of ROH between herds was insignificant (*t*-test). But, similar to consecutive runs after exclusion of the most deviated cow (it included 731 ROH segments) among the fourth herd, the mean number of ROH became 77.6±2.0. However, this value still did not significantly differ from those for other herds (*t*-test). The average number of ROH increased by 1.7 times, from 86.0±2.6 to 146.7±1.7, when one heterozygous SNP was allowed in ROH, then by 3 times when two heterozygous SNPs were allowed. The observed increase in the number of ROH was significant $p \leq 0.02$ (*t*-test).

The length of the ROH segments for sliding runs has been classified into the same five categories (1–2, 2–4, 4–8, 8–16, and >16 Mb) as it has been carried out for consecutive runs (Suppl. Material 3). The most numerous in the number of ROH has occurred in the same class of 1–2 Mb, in which a considerable increase in ROH segments was observed with an increase in the number of allowed heterozygous SNPs in ROH. This indicates the proximity of numerous ROH segments shorter than 1 Mb in the cow genome.

ROH identification based on Plink. Plink software is widely used in ROH studies. Therefore, it is necessary to compare the data obtained by Plink and detectRUNS. The mean number of ROH obtained with Plink was 74.9±1.9 and this value was no different from the value calculated by detectRUNS based on sliding runs 86.0±2.6 (*t*-test). The fact that Plink identified fewer ROH segments mainly in the 1–2 Mb class than detectRUNS detected (see Suppl. Materials 2 and 3) indicates Plink's lesser ability to identify short segments less than 1 Mb. Thus, the data obtained for the shortest ROH length class can be highly dependent on the software and parameters used.

Comparative analysis of consecutive and sliding runs. Comparative analysis of the consecutive and sliding data led to the following conclusions. When heterozygous SNPs were disallowed, the consecutive runs showed a slightly bigger mean number of ROH than sliding runs (94.4±2.7 *vs.* 86.0±2.6) and even bigger for sliding windows (Plink) 74.9±1.9, but the difference between them was insignificant (*t*-test). The fewer SNPs were used in consecutive runs, the more 1–2 Mb ROH segments there were (Table 3). Summarizing the comparative analysis of the applied methods, one can come to the conclusion that there are some differences in the results obtained by these methods.

Distribution of ROH in the cow chromosomes. To evaluate the chromosomes with the largest number of ROH segments taking into account their length, the following calculation was carried out. For each chromosome, the proportion of ROH in it was divided by the share of its size in the cattle genome. The rank chromosome calculation is shown in Suppl. Material 4. For both runs, the list of chromosomes ranked in this way is shown in Table 3. Out of 29 chromosomes, the top chromosomes covered with ROH were BTA 14, BTA 16 and BTA 7, not BTA 1 (seventh position in the list), BTA 2 (20th position in the list) and BTA 3 (16th position in the list). Thus, the number of ROH in the chromosomes was not proportional to their length. Spearman's correlation between consecutive and sliding runs data in Table 3 was $r = 1.0$ ($p \leq 2.0E{-}07$). Whether this fact is a result of drift or/and selection requires further study.

Inbreeding. To assess the level of inbreeding in the herds, the mean inbreeding coefficient was calculated across all herds (Tables 4 and 5). When heterozygous SNPs were disallowed, the mean inbreeding coefficients across the herds amounted to 0.111±0.003 and 0.104±0.004 for consecutive and sliding runs, respectively, and the difference between them was insignificant (*t*-test). The mean inbreeding coefficient estimated by Plink was 0.105±0.004, which is consistent with those for sliding runs. A greater variability in inbreeding occurred for the fourth herd. This result is mainly associated with a highly inbred cow in this herd. Exclusion of this cow results in the average inbreeding coefficient of 0.096±0.005 and 0.089±0.005 for consecutive and sliding runs. It should be noted that in this herd the cows were inseminated only from the Netherlands bulls, while in other herds the bulls' semen from North America, Germany, Canada, and the Netherlands was used. The proportion of the bulls from these countries used in the herds was published in the article (Smaragdov et al., 2018). After excluding the highly inbred cow, the average inbreeding coefficient in the fourth herd decreased compared to other herds. This result indicates the correct selection of the bulls even if their semen was imported from the same country. The fourth herd deviated significantly from the other herds

Table 4. Estimated average inbreeding coefficient (± SE) in the herds based on 20 SNPs consecutive runs (detectRUNS)

Parameter	Herd						
	1	2	3	4	5	6	Mean
				Zero heterozygous SNPs in ROH			
Inbreeding coefficient	0.117 ± 0.004	0.112 ± 0.002	0.105 ± 0.002	0.111 ± 0.016	0.116 ± 0.003	0.105 ± 0.003	0.111 ± 0.003
Maximum	0.227	0.153	0.160	0.779	0.166	0.158	
Minimum	0.0006	0.060	0.068	0.047	0.075	0.071	
				One heterozygous SNP in ROH			
Inbreeding coefficient	0.153 ± 0.006	0.143 ± 0.002	0.143 ± 0.002	0.147 ± 0.018	0.147 ± 0.002	0.137 ± 0.003	0.145 ± 0.003
Maximum	0.367	0.184	0.184	0.908	0.191	0.192	
Minimum	0.003	0.091	0.091	0.072	0.104	0.102	
				Two heterozygous SNPs in ROH			
Inbreeding coefficient	0.208 ± 0.008	0.195 ± 0.002	0.189 ± 0.002	0.198 ± 0.017	0.199 ± 0.002	0.190 ± 0.003	0.196 ± 0.003
Maximum	0.511	0.235	0.229	0.948	0.240	0.246	
Minimum	0.010	0.139	0.148	0.102	0.158	0.155	

Table 5. Estimated average inbreeding coefficient (± SE) in herds based on 20 SNPs sliding runs (detectRUNS)

Parameter	Herd						
	1	2	3	4	5	6	Mean
				Zero heterozygous SNPs in ROH			
Inbreeding coefficient	0.110 ± 0.004	0.105 ± 0.002	0.098 ± 0.002	0.103 ± 0.015	0.109 ± 0.003	0.098 ± 0.003	0.104 ± 0.004
Maximum	0.204	0.146	0.153	0.739	0.158	0.151	
Minimum	0.0003	0.055	0.063	0.040	0.070	0.065	
				One heterozygous SNP in ROH			
Inbreeding coefficient	0.154 ± 0.007	0.143 ± 0.002	0.137 ± 0.002	0.146 ± 0.018	0.148 ± 0.002	0.138 ± 0.003	0.148 ± 0.003
Maximum	0.392	0.186	0.184	0.938	0.254	0.194	
Minimum	0.002	0.092	0.094	0.070	0.173	0.104	
				Two heterozygous SNPs in ROH			
Inbreeding coefficient	0.227 ± 0.010	0.213 ± 0.002	0.207 ± 0.002	0.217 ± 0.018	0.217 ± 0.002	0.208 ± 0.003	0.215 ± 0.015
Maximum	0.583	0.254	0.244	0.980	0.254	0.257	
Minimum	0.010	0.157	0.167	0.108	0.173	0.173	

when variability was measured by the Wright's fixation index or PCA (Smaragdov, Kudinov, 2020). When one heterozygous SNP was allowed in ROH, then the mean inbreeding coefficient across all herds was 0.145 ± 0.003 and 0.148 ± 0.003 based on consecutive and sliding runs. Thus, the allowance of even one heterozygous SNP resulted in an increase in the inbreeding coefficient ($p \leq 0.06$). Therefore, to assess inbreeding in the herds, heterozygous SNPs should be disallowed in ROH due to sizable bias.

Confirmation of results obtained on cows with data on bulls. To validate the results obtained on the cows, the bulls that have been used two generations ago in the same herds were analyzed for ROH. The mean number of the ROH segments for the bulls, 58.9 ± 1.9, turned out to be significantly less than for the cows, 95.4 ± 2.7 ($p \leq 0.05$) (Tables 1 and 6). The mean inbreeding coefficient for the bulls was 0.078 ± 0.005 and did not differ significantly from the cows (*t*-test). The coefficient of inbreeding did not significantly increase when one heterozygous SNP was allowed (*t*-test) (Table 7).

Table 6. Estimated mean ROH number (± SE) on bulls based on 20 SNPs consecutive runs (detectRUNS)

ROH number	N*		
	0	1	2
The mean number of ROH	58.9 ± 1.9	93.6 ± 1.7	153.5 ± 1.8
Maximum	85	112	172
Minimum	44	79	128

* The number of allowed heterozygous SNPs in ROH.

Table 7. Estimated inbreeding coefficient (± SE) on bulls based on 20 SNPs consecutive runs (detectRUNS)

N*	0	1	2
Inbreeding coefficient	0.078 ± 0.005	0.098 ± 0.005	0.133 ± 0.004

* The number of allowed heterozygous SNPs in ROH.

Discussion

Over the past decade, the runs of homozygosity approach has been widely used both in humans (Ceballos et al., 2018b) and farm animals (Peripolli et al., 2016). A distinctive feature of ROH studies is the variety of software and threshold criteria used in them. The most widely applied software tools for identifying ROH segments are either sliding window or consecutive runs. We preferred detectRUNS, where both approaches have been implemented (Biscarini et al., 2018).

The consecutive runs resulted in the average number of ROH 94.4±2.7, while sliding runs, 86.0±2.6. These values for North American (Forutan et al., 2018), Italian (Marras et al., 2014), European Holstein (Zinovieva et al., 2020), and Polish Holstein Black-and-White variety (Szmatoła et al., 2019) are 82.3±9.8 (SD), 81.7±9.7 (SD), 74.6±2.3 (SE), and 53.3±7.3 (SD) respectively. The first three values do not differ significantly from ours, while the value for Polish cattle differs considerably. It should be noted that the allowance of even one heterozygous SNP in ROH significantly increases the number of ROH by 55.9 and 60.7 points for consecutive and sliding runs, respectively (see Tables 1 and 2). A limited number of studies have analyzed the effect of allowed heterozygous SNPs on ROH data. D. Howrigan et al. (2011) recommended disallowing the use of any heterozygous SNPs in ROH, while M. Ferenčaković et al. (2013) suggested that the number of allowed heterozygous SNPs should be determined separately for each ROH length of interest and for each SNPs density. Moreover, the allowance of heterozygous SNPs in ROH leads to a sizable bias in the inbreeding coefficient (Mastrangelo et al., 2016). My results confirm this conclusion.

The relative frequency of the ROH number in different length classes obtained from the cows data for consecutive runs were 61.4 % (1–2 Mb), 19.8 % (2–4 Mb), 11.3 % (4–8 Mb), 5.5 % (8–16 Mb) and 1.9 % (longer than 16 Mb), while for sliding runs these values were 60, 19.8, 12.1, 5.8, and 2.1 %. Thus, the largest number of ROH was identified in the shortest 1–2 Mb class. Plink-running of the cows genome revealed the following ROH frequencies in five categories 52 % (1–2 Mb), 25 % (2–4 Mb), 14 % (4–8 Mb), 7 % (8–16 Mb) and 2.5 % (longer than 16 Mb), the distribution of which is slightly different from those defined by detectRUNS. For North American Holstein animals, these values were 43.5, 23.9, 17.7, 10.5, and 4.7 % (Forutan et al., 2018). The corresponding values for Italian Holstein bulls were 56.9, 20.8, 11.9, 7.2, and 3.7 % (Marras et al., 2014) and Polish Holstein, 23, 19, 9.8, 4.4, and 1.3 % (Szmatoła et al., 2019). Thus, when we used detectRUNS to scan the genome of our local Holstein cows, we obtained an abundant number of short ROH as a result of haplotypes reflecting the ancient relationship within breeding animals. But, when we used Plink, the values were similar to the American and Italian data. It should be noted that the authors of the article (Szmatoła et al., 2019) used the cgaTOH software and their data differ considerably from other data. Whether this result was due to the cgaTOH software (minimal number of 30 consecutive homozygous SNPs in ROH) or/and selection requires further analysis. Estimation of the true number of short ROH is important, since 0.1–3 Mb ROH segments have the more number of deleterious variants than segments longer than 3 Mb (Zhang Q. et al., 2015b). For evaluation of the genomic estimated breeding value (GEBV), short ROH is essential for genomic construction of ROH-based relationship matrix (G_{ROH}) (Luan et al., 2014).

According to my data, the largest number of ROH falls into the 1–2 Mb class. As the number of allowed heterozygous SNPs in ROH increases, the number of ROH segments in the shortest 1–2 Mb class increases as well (see Suppl. Materials 2 and 3). This fact indicates a close location of a large number of short, less than 1 Mb, ROH segments.

The same conclusion was reached in a study of ten sequenced (WGS) breeds of cattle (Mulim et al., 2022). Then, the results of the animals ROH genome scanning can substantially depend not only on the selection but also on the genotyping method and the software used to identify short ROH segments. This fact should be taken into account in the comparative analysis of the ROH data.

Estimated by detectRUNS, the mean inbreeding coefficient for six herds was 0.111±0.003 and 0.104±0.004 for consecutive and sliding runs, respectively, and for bulls, 0.078±0.005 for consecutive runs. It was equal to 0.105±0.004 based on the sliding window runs evaluated by Plink. It should be noted that cows from six herds did not differ in the mean inbreeding coefficient (see Tables 4 and 5), while according to Principal Components Analysis, the fourth herd differed significantly from all other herds (Smaragdov, Kudinov, 2020). Therefore, this difference is not due to inbreeding.

The accurate knowledge of inbreeding in the herds that occurred several decades in the past is necessary both for calculating the inbreeding trend and for evaluating selection strategies. To solve this problem, high-density arrays or whole genome sequencing (WGS) should be used. Comparison of 50k and HD panels provides evidence that the data from the 50k panel lead to imprecise determination of short ROH segments (Ferenčaković et al., 2013). However, it has been shown that ROH detection based on high-density or 50k array data might give the estimates of current inbreeding most similar to ROH values obtained from the sequence data (Zhang Q. et al., 2015a). M. Bhati et al. (2020) provided comprehensive WGS data for Braunvich cattle. Medium-sized ROH (0.1–2 Mb) were the most frequent class (50.46 %) and made the largest contribution (75 %) to total genomic inbreeding, while short, 50–100 Kb, ROH occurred almost as frequently (49.17 %) as medium-sized ROH, they contributed only 19.52 % to total genomic inbreeding. These findings provide an accurate estimate of short ROH in the cattle genome and their contribution to total inbreeding. The average F_{ROH} estimated from the WGS data was 0.14 in Braunvich cattle. This value is less than WGS F_{ROH} in Holstein, 0.18 (Bhati et al., 2020). Summarizing, the 50k panel cannot accurately capture ancient inbreeding that occurred a few decades in the past. The inbreeding coefficient of American Holstein measured with ROH in 2011 was 0.12 and after applying genomic selection, it increased to 0.15 in 2018 (Forutan et al., 2018). For European (Zinovieva et al., 2020), Italian (Marras et al., 2014), and Polish Holstein (Szmatoła et al., 2019), these values were 0.108±0.006 (SE), 0.116±0.001 (SE), and 0.118±0.027 (SD), respectively. It is important to note that in the above studies, ROH data were based on the 50k array; thereby, ROH segments not shorter than 1 Mb were identified. Once again, we have to admit that, according to our data, an increase in the number of mostly short ROHs (1–2 Mb) by 395 points identified during consecu-

tive runs compared to sliding runs (Suppl. Materials 2 and 3) leads to only a slight increase in the inbreeding coefficient (Tables 4 and 5).

It can be assumed that there should be an event horizon for a herd or population, beyond which it is impossible to obtain valid information about inbreeding events in history of their breeding. I hypothesize that in our local population, a reduced effective population size, ongoing admixture and inbreeding throughout its history, accompanied by recombination, should lead to the largest number of short ROH less than 1 Mb in the herds currently studied. These short ROH can be considered as ancient ROH segments formed by some population events, such as drift, bottleneck, and inbreeding that occurred many decades ago. The bottleneck in our local herds has not previously been proven by Principal Component Analysis (Smaragdov, Kudinov, 2020). An accurate interpretation of these short ROH can be troublesome without knowledge of the herd management history. In addition, it is very important to know the true number of short ROHs in the analyzed animals resulting from inbreeding (see above-mentioned WGS data). Thus, the event horizon can depend on both pedigree information, ROH length profile (SNPs array or WGS used) as well as on the algorithm-defined approach to ROH identification. However, ROH segments shorter than 500 Kb can be considered to be beyond the event horizon due to strong LD and inconsistency with autozygosity. The short ROH characterized by strong LD among markers are not always considered autozygous, but nevertheless they may have formed due to mating with distantly related animals (McKay et al., 2007). Summarizing, it should be assumed that inbreeding data can be only relatively correct based on ROH larger than 1 Mb (no more than 50 generation back).

A number of studies have noted an uneven distribution of ROH in the bovine genome, e. g. (Ferencakovic et al., 2011; Sölkner et al., 2014; Howard et al., 2015). Giving the number of ROH in the chromosomes, we calculated their rank taking into account the proportion of chromosome length in the genome of the cattle (see Table 3). Out of 29 chromosomes, the most covered with ROH segments were BTA 14, BTA 16 and BTA 7 for both approaches used. D. Purfield et al. (2012) noticed that among the breeds studied, BTA 14 and BTA 16 had the highest degree of ROH segments overlap. The regions of the genome with the highest frequency of occurrence of ROH in the genome of the studied animals were called "ROH islands" (Nothnagel et al., 2010; Pemberton et al., 2012). The ROH islands on BTA 14 and BTA 16 were identified among Polish Holstein-Friesian animals (Szmatoła et al., 2019). In Holstein cows in our study, ROH islands were localized in BTA 7 and BTA 14 (unpublished results). In American Holstein, ROH distribution was more variable among the genomes of the selected animals, compared to a relatively even ROH distribution in unselected animals (Kim et al., 2013). Regions with a high proportion of ROH for American and New Zealand Jersey cows and bulls were revealed on BTA 3 and BTA 7 (Howard et al., 2015). On BTA 14 and BTA 16, one strongest ROH region was found common for Kholmogor and Holstein breeds and one region common for Yaroslavl and Holstein breeds (Zinovieva et al., 2020). Extremely non-uniform ROH patterns among bovine populations of Angus, Brown Swiss, and Fleckvieh breeds were mainly located on BTA 6, BTA 7, BTA 16, and BTA 21 (Sölkner et al., 2014). The highest number of ROH islands among all Neilore breed lineages was found on BTA 7 (Peripolli et al., 2018a). In addition, an enrichment of genes affecting traits of interest for dairy breeds was shown on BTA 14 in dairy Gyr breed (*Bos indicus*) (Peripolli et al., 2018b). D. Goszczynski et al. (2018) analyzed ROH >16 Mb (three generations from a common ancestor) in highly inbred Retinta bulls. Among other chromosomes, the highest occurrence of ROH was found on BTA 7. Summarizing the above studies, it can be suggested that BTA 7 is outstanding regarding ROH islands occurrence in the cattle genome but in general there is no overall direct relationship between the proportion of ROH segments in the chromosomes and ROH islands identified there.

As discussed above, the number of identified short ROH is highly dependent on the software used and also on the genotyping method. Moreover, it can be suggested that consecutive runs more accurately identified the ROH pattern in the cow genome. However, both methods coincide in assessing the distribution of ROH segments on chromosomes (see Table 3). Taking the findings together, it should be assumed that uneven distribution of ROH segments in the cow genome is a result of different inbreeding events that have occurred in their history.

Conclusion

Analysis of ROH data showed that consecutive runs most accurately identified ROH in the cattle genome. It has been shown that missing SNPs did not have a noticeable effect on the number of ROH, while an allowance of even one heterozygous SNP in the ROH segments had a significant effect. Therefore, care should be taken to allow any heterozygous SNPs in the ROH. The average number of ROH across herds was 95.4 ± 2.7 and their length varied from 1 Mb to more than 16 Mb. The class with the length of 1–2 Mb was the most numerous in the number of ROH. This confirms the long history of inbreeding in herds for many decades in the past. Moreover, the number of ROH in the chromosomes does not depend on their length. ROH segments mainly cover BTA 14, BTA 16, and BTA 7. The average inbreeding coefficient for our local Holstein herds was 0.111 ± 0.003, which is not much different from the Holstein cattle inbreeding coefficient worldwide. This value indicates competent management of the studied herds. In addition, the inbreeding coefficient obtained on cows is consistent with the inbreeding coefficient of 0.078 ± 0.005 calculated in our study for Holstein bulls from other countries. These bulls have been used in breeding our local Holstein cattle two generations ago.

References

Addo S., Klingel S., Hinrichs D., Thaller G. Runs of homozygosity and NetView analyses provide new insigt into the genome-wide diversity and admixture of three German cattle breeds. *PLoS One.* 2019;14(12):e0225847. DOI 10.1371/journal.pone.0225847.

Bhati M., Kadri N., Crysnanto D., Pausch H. Assesing genomic diversity and signatures of selection in Original Braunvich cattle using whole-genome sequencing data. *BMC Genomics.* 2020;21(1):27. DOI 10.1186/s12864-020-6446-y.

Biscarini F., Cozzi P., Gaspa G., Marras G. detectRUNS: detect runs of homozygosity and runs of heterozygosity in diploid genomes. R package version 0.9.5. 2018.

Bjelland D., Weigel K., Vukasinovic N., Nkrumah J. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J. Dairy Sci.* 2013;96(7):4697-4706. DOI 10.3168/jds.2012-6435.

Browning S., Browning B. High-resolution detection of identity by descent in unrelated individuals. *Am. J. Hum. Genet.* 2010;86(4): 526-539. DOI 10.1016/j.ajhg.2010.02.021.

Ceballos F., Hazelhurst S., Ramsay M. Assessing runs of Homozygosity: a comparison of SNP array and whole genome sequence low coverage data. *BMC Genomics.* 2018a;19(1):106. DOI 10.1186/s12864-018-4489-0.

Ceballos F., Joshi P., Clark D., Ramsay M., Wilson J. Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* 2018b;19(4):220-234. DOI 10.1038/nrg.2017.109.

Ferencakovic M., Hamzic E., Gredler B., Curik I., Johann S. Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh cattle. *Agric. Conspec. Sci.* 2011;76(4):325-329.

Ferenčaković M., Sölkner J., Curik I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet. Sel. Evol.* 2013;45:42.

Forutan M., Mahyari S., Baes C., Melzer N., Schenkel F., Sargolzaei M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC Genomics.* 2018;19(1):98. DOI 10.1186/s12864-018-4453-z.

Goszczynski D., Molina A., Tera E., Morales-Durand H., Ross P., Cheng H., Giovambattista G., Demyda-Peyrás S. Runs of homozygosity in a selected cattle population with extremely inbred bulls: descriptive and functional analyses revealed highly variable patterns. *PLoS One.* 2018;13(7):e0200069. DOI 10.1371/journal.pone.0200069.

Gusev A., Lowe J., Stoffel M., Daly M., Altshuler D., Breslow J., Friedman J., Peer I. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* 2009;19(2):18-26. DOI 10.1101/gr.081398.108.

Howard J., Maltecca C., Haile-Mariam M., Hayes B., Pryce J. Characterizing homozygosity across United States, New Zealand and Australian Jersey cow and bull populations. *BMC Genomics.* 2015; 16(1):187. DOI 10.1186/s12864-015-1352-4.

Howrigan D., Simonson M., Keller M. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics.* 2011;12:460. DOI 10.1186/1471-2164-12-460.

Karimi Z. Runs of homozygosity patterns in taurine and indicine cattle breeds. Doctoral thesis. Vienna: Univ. of Natural Resources and Life Sciences, 2013.

Kim E., Cole J., Huson H., Wiggans G., Van Tassel C., Crooker B., Liu G., Da Y., Sonstegard T. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. *PLoS One.* 2013;8(11):e80813. DOI 10.1371/journal.pone.0080813.

Kim E., Sonstegard T., Van Tassell C., Wiggans G., Rothschild M. The relationship between runs of homozygosity and inbreeding in Jersey cattle under selection. *PLoS One.* 2015;10(7):e0129967. DOI 10.1371/journal.pone.0129967.

Kudinov A., Mäntysaari E., Pitkänen T., Saksa E., Aamand G., Uimari P., Strandén I. Single-step genomic evaluation of Russian dairy cattle using internal and external information. *J. Anim. Breed. Genet.* 2022;139(3):259-270. DOI 10.1111/jbg.12660.

Luan T., Yu X., Dolezal M., Bagnato A., Meuwissen T.H. Genomic prediction based on runs of homozygosity. *Genet. Sel. Evol.* 2014; 46(1):64. DOI 10.1186/s12711-014-0064-6.

Marras G., Gaspa G., Sorbolini S., Dimauro C., Ajmone-Marsan P., Valentini A., Williams J., Macciotta N. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim. Genet.* 2014;46(2):110-121. DOI 10.1111/age.12259.

Mastrangelo S., Tolone M., Di Gerlando R., Fontanesi L., Sardina M., Portolano B. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal.* 2016;10(5):746-754. DOI 10.1017/S1751731115002943.

Matukumalli L., Cynthia T., Lawley C., Robert D., Schnabel R., Taylor J., Allan M., Heaton M., O'Connell J., Moore S., Smith T., Sonstegard T., Van Tassell C. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One.* 2009;4(4): e5350. DOI 10.1371/journal.pone.0005350.

McKay S.D., Schnabel R.D., Murdoch B.M., Matukumalli L.K., Aerts J., Coppieters W., Crews D., Dias Neto E., Gill C.A., Gao C., Mannen H., Stothard P., Wang Z., Van Tassell C.P., Williams J.L., Taylor J.F., Moore S.S. Whole genome linkage disequilibrium maps in cattle. *BMC Genetics.* 2007;8:74. DOI 10.1186/1471-2156-8-74.

Meuwissen T., Luo Z. Computing inbreeding coefficient in large populations. *Genet. Sel. Evol.* 1992;24(4):305-313. DOI 10.1186/1297-9686-24-4-305.

Mulim H., Brito L., Pinto L., Ferraz J., Grigoletto L., Silva M., Pedrosa V. Characterization of runs of homozygosity – enriched regions, and population structure in cattle populations selected for different breeding goals. *BMC Genomics.* 2022;23(1):209. DOI 10.1186/s12864-022-08384-0.

Narasimhan V., Danecek P., Scally A., Xue Y., Tyler-Smith C., Durbin R. BCFtools/ROH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics.* 2016;32(11):1749-1751. DOI 10.1093/bioinformatics/btw044.

Nothnagel M., Lu T., Kayser M., Krawczak M. Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum. Mol. Genet.* 2010;19(15):2927-2935. DOI 10.1093/hmg/ddq198.

Pemberton T., Absher D., Feldman M., Myers R., Rosenberg N., Li J. Genomic: patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* 2012;91(2)*:*275-292. DOI 10.1016/j.ajhg.2012.06.014.

Peripolli E., Munari D., Silva M., Lima A., Irgang R., Baldi F. Runs of homozygosity: current knowledge and applications in livestock. *Anim. Genet.* 2016;48(3):255-271. DOI 10.1111/age.12526.

Peripolli E., Metzger J., de Lemos M.V.A., Stafuzza N.B., Kluska S., Olivieri B.F., Feitosa F.L.B., Berton M.P., Lopes F.B., Munari D.P., Lôbo R.B., Magnabosco C.U., Di Croce F., Osterstock J., Denise S., Pereira A.S.C., Baldi F. Autozygosity islands and ROH patterns in Nellore lineages: evidence of selection for functionally important traits. *BMC Genomics*. 2018a;19(1):680. DOI 10.1186/s12864-018-5060-8.

Peripolli E., Stafuzza N.B., Munari D.P., Lima A.L.F., Irgang R., Machado M.A., Panetto J.C.D.C., Ventura R.V., Baldi F., da Silva M.V.G.B. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. *BMC Genomics*. 2018b;19(1):34. DOI 10.1186/s12864-017-4365-3.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., de Bakker P.I., Daly M.J., Sham P.C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007;81(3):559-575. DOI 10.1086/519795.

Purfield D., Berry D., McParland S., Bradley D. Runs of homozygosity and population history in cattle. *BMC Genetics.* 2012;13:70. DOI 10.1186/1471-2156-13-70.

Smaragdov M., Kudinov A. Assessing the power of principal components and wright's fixation index analyzes applied to reveal the genomewide genetic differences between herds of Holstein cows. *BMC Genetics.* 2020;21:47. DOI 10.1186/s12863-020-00848-0.

Smaragdov M., Kudinov A., Uimari P. Assessing the genetic differentiation of Holstein cattle herds in the Leningrad region using F_{st} statistics. *Agric. Food Sci.* 2018;27(2):96-101. DOI 10.23986/afsci.69777.

Sölkner J., Ferenčaković M., Karimi Z., Pérez O'Brien A.M., Mészáros G., Eaglen S., Boison S., Curik I. Extremely non-uniform: patterns of runs of homozygosity in bovine populations. In: Proceedings of the 10th World Congress on Genetics Applied to Livestock Production (August 17–22, 2014, Vancouver, BC Canada). 2014.

Szmatoła T., Gurgul A., Jasielczuk I., Zabek T., Ropka-Molik K., Litwinczuk Z., Bugno-Poniewierska M. Comprehensive analysis of runs of homozygosity of eleven cattle breeds representing different production types. *Animals*. 2019;9(12):1024. DOI 10.3390/ani9121024.

Zavarez L.B., Utsunomiya Y.T., Carmo A.S., Neves H.H., Carvalheiro R., Ferenčaković M., Pérez O'Brien A.M., Curik I., Cole J.B., Van Tassell C.P., da Silva M.V., Sonstegard T.S., Sölkner J., Garcia J.F. Assessment of autozygosity in Nellore cows (*Bos indicus*) through high density SNP genotypes. *Front. Genet.* 2015;6:5. DOI 10.3389/fgene.2015.00005.

Zhang L., Orloff M., Reber S., Li S., Zhao Y., Eng C. cgaTOH: extended approach for identifying tracts of homozygosity. *PLoS One.* 2013;8:e57772. DOI 10.1371/journal.pone.0057772.

Zhang Q., Calus M., Guldbrandtsen B., Lund M., Sahana G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genetics.* 2015a;16:88. DOI 10.1186/s12863-015-0227-7.

Zhang Q., Guldbrandtsen B., Bosse M., Lund M., Sahana G. Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics.* 2015b;16:542. DOI 10.1186/s12864-015-1715-x.

Zinovieva N., Dotsev A., Sermyagin A., Deniskova T., Abdelmanova A., Kharzinova V., Solkner J., Reyer H., Wimmers K., Brem G. Selection signatures in two oldest Russian native cattle breeds revealed using high density single nucleotide polymorphism analysis. *PLoS One.* 2020;15:e0242200. DOI 10.1371/journal.pone.0242200.

ORCID ID

Smaragdov M.G. orcid.org/0000-0002-5087-6444

Acknowledgements. This work was supported by the Ministry of Science and Higher Education of the Russian Federation (State Assignment Program No. 121052600352-3).

Conflict of interest. The author declares no conflict of interest.

Received December 4, 2022. Revised February 3, 2023. Accepted February 27, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):480-487
DOI 10.18699/VJGB-23-58

CD-1 mice females recognize male reproductive success via volatile organic compounds in urine

A.S. Khotskina[1] ✉, E.L. Zavjalov[1], E.P. Shnayder[1], L.A. Gerlinskaya[1], S.O. Maslennikova[1], D.V. Petrovskii[1], M.N. Baldin[2], A.L. Makas[2], V.M. Gruznov[2], M.L. Troshkov[2], M.P. Moshkin[1]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Trofimuk Institute of Petroleum Geology and Geophysics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ dotcenko@bionet.nsc.ru

Abstract. Sexual selection is considered as one of the leading factors of evolutionary development. In the conditions of incessant competition, specialized methods of attracting individuals of the opposite sex as well as criteria for assessing the quality of a sexual partner have been formed. In order for animals to rely on signaling from sexual partners, the signal must reflect the morpho-physiological status of animals. A high reproductive efficiency of male mice is a good advantage for mate selection and thus must be somehow demonstrated to potential mates. The aim of our study was to find out if male mice could demonstrate their reproductive efficiency through urine volatile organic compounds. The experiment implies cohabiting one male with two mature females for 6 days. The reproductive success of the male was assessed by the presence or absence of pregnant females. At the same time, naive females, who did not participate in reproduction, assessed the urine of the successful males as more attractive, which was expressed in shorter Latency time of sniffs in the Olfactory test. Using a rapid headspace GC/MS analysis, we have found volatile organic compounds (VOCs) in male urine that correlated with female behavior. It turned out that these substances are derivatives of mouse pheromone 6-hydroxy-6-methyl-3-heptanone. The amplitude of peaks corresponding to this pheromone correlated with the testosterone level in blood and the weight of preputial glands. The amplitude of peaks increased in males after mating with whom the females turned out to be pregnant. It is important to note that body weight, weight of testes, weight of seminal vesicles, weight of preputial glands, and plasma testosterone level alone are not reliable indicators of male reproductive success. Thus, the content of the pheromone 6-hydroxy-6-methyl-3-heptanone in the urine of males can serve as a good predictor of the quality of the male as a sexual partner for female CD-1 mice.
Key words: chemical signals; dihydrofuran; GC/MS; 6-hydroxy-6-methyl-3-heptanone; mating preference; olfactory preference; reproductive success.

For citation: Khotskina A.S., Zavjalov E.L., Shnayder E.P., Gerlinskaya L.A., Maslennikova S.O., Petrovskii D.V., Baldin M.N., Makas A.L., Gruznov V.M., Troshkov M.L., Moshkin M.P. CD-1 mice females recognize male reproductive success via volatile organic compounds in urine. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):480-487. DOI 10.18699/VJGB-23-58

Самки мышей CD-1 распознают репродуктивно успешных самцов по летучим органическим соединениям их мочи

А.С. Хоцкина[1] ✉, Е.Л. Завьялов[1], Е.П. Шнайдер[1], Л.А. Герлинская[1], С.О. Масленникова[1], Д.В. Петровский[1], М.Н. Балдин[2], А.Л. Макась[2], В.М. Грузнов[2], М.Л. Трошков[2], М.П. Мошкин[1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Институт нефтегазовой геологии и геофизики им. А.А. Трофимука Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ dotcenko@bionet.nsc.ru

Аннотация. Половой отбор рассматривается в качестве одного из ведущих факторов эволюционного развития. В условиях постоянной конкуренции сформировались специализированные способы привлечения особей противоположного пола, а также критерии оценки качества полового партнера. Самцы, способные оставить наибольшее количество потомков, должны быть более привлекательными половыми партнерами, что требует от них каким-либо образом демонстрировать самкам свои преимущества. При этом ключевое условие для реализации такого отбора – это способность сигнала достоверно отражать информацию о физиологическом состоянии организма особи. У мышей одним из ведущих каналов передачи информации являются хемосигналы. Целью нашего исследования было выяснить, могут ли самцы мышей демонстрировать самкам свою репродуктивную эффективность через летучие органические соединения мочи. В ходе эксперимента самцу подсаживали двух половозрелых самок на шесть дней для размножения. Репродуктивный успех самца оценивали по наличию или отсутствию потомства. В то же время половозрелые самки, ранее не участвовавшие в размножении, оценивали мочу этих самцов. Оказалось, что более привлекательной была моча самцов, способных оставить потомство. Это выражалось в более коротком латентном времени обнюхивания самкой мочи самца в ольфак-

торном тесте. При помощи газовой хроматографии в моче самцов были обнаружены летучие органические соединения, которые коррелировали с поведением самок. Масс-спектрометрический анализ образцов мочи показал, что эти соединения являются производными феромона мышей 6-гидрокси-6-метил-3-гептанона. Амплитуда пиков, соответствующих данному феромону, коррелировала с уровнем тестостерона в крови и массой препуциальных желез и была выше у самцов с фертильными покрытиями. Важно отметить, что ключевые морфофизиологические показатели, такие как масса тела, масса семенников, препуциальных желез, семенных пузырьков, а также уровень тестостерона в плазме крови самцов, не отражали их способность оставить потомство. Таким образом, содержание феромона 6-гидрокси-6-метил-3-гептанона в моче самцов CD-1 может служить хорошим критерием их репродуктивной эффективности.
Ключевые слова: хемосигналы; дигидрофуран; ГХ/МС; 6-гидрокси-6-метил-3-гептанон; половое предпочтение; ольфакторное предпочтение; репродуктивный успех.

Introduction

Animal olfactory cues contain a vast amount of information that plays an important role in their life and population processes. The odors play a special role in the relations between the sexes and ensure the process of mating (Brennan, Zufall, 2006; Arakawa et al., 2008). Most often, the rodents use urinary tags for information transfer (Hurst, Beynon, 2004). Urine could be considered as the body fluid with the highest capability to "yield" different volatile organic compounds (VOC) that could be used in chemocommunication (Novotny et al., 1999b). The complexity of mouse urine volatile profile has been described in a number of publications (Novotny et al., 2007; Schaefer et al., 2010). Through GC/MS analysis, more than two hundred VOCs were found on the chromatographic profiles of mouse urine, and for nearly half of them chemical structure was identified (Schwende et al., 1986; Jemiolo et al., 1987; Röck et al., 2007; Zhang et al., 2007; Schaefer et al., 2010; Liu et al., 2017). Several substances were described as unique mouse urine constituents, which are not present in urine of any other species – the mouse pheromones (Novotny et al., 2007). The biological activity of their majority has already been studied (Novotny et al., 1985, 1990, 1999b; Jemiolo et al., 1986). The role of other urinary volatile metabolites has been studied less than that of pheromones but they are also involved in the process of chemocommunication. Urinary metabolites form an odor background, which reflects individual features such as diet, stress level, genotype and others (Zhang et al., 2007; Schaefer et al., 2010).

Most known male pheromones, such as: 2-sec-butyl-4,5-dihydrothiazole and 3,4-dihydro-exo-brevicomine (Jemiolo et al., 1985), 1-hexadecanol and 1-hexadecanol acetate (Zhang et al., 2007), α- and β-farnesene (Jemiolo et al., 1991), MTMT (Lin et al., 2005), and darcin (Liu et al., 2017) are highly attractive for female mice. A lot of male pheromones, such as: 2-sec-butyl-4,5-dihydrothiazole, 3,4-dihydro-exo-brevicomine, α- and β-farnesene, 6-hydroxi-6-methyl-3-heptanone (Novotny et al., 1999a), and 2-isopropyl-4,5-dihydrothiazole have a stimulating effect on puberty in females (Osada et al., 2008). Wherein some of them: 2-sec-butyl-4,5-dihydrothiazole, 3,4-dihydro-exo-brevicomine (Jemiolo et al., 1986), α- and β-farnesene (Ma et al., 1999) can stimulate the estrus synchronization in female population (Whitten et al., 1968). Now it is known as Whitten effect. Beside this, all above-mentioned pheromones induce the estrus cycle (Jemiolo et al., 1986; Ma et al., 1999).

Moreover, it was shown that the females of mice, rats, and voles could discriminate males. For examples, only by the scent of urine females can discriminate the genotype of males (Penn, Potts, 1998; Roberts, Gosling, 2003; Ilmonen et al., 2009; Manser et al., 2015), their maturity (Osada et al., 2003, 2008), hierarchy status (Drickamer, 1992; Veyrac et al., 2011), parasite load (Kavaliers, Colwell, 1995; Willis, Poulin, 2000), immunocompetence (Zala et al., 2004; Gerlinskaya et al., 2012), and infection status (Moshkin et al., 2001, 2002; Zala et al., 2015).

Now it is known that the free choice of a partner ensures the birth of the most viable offspring. In experiments on various species of animals it was shown that the survival rate from the moment of birth till reaching sexual maturity is significantly higher in individuals born when mating occurs in accordance with the free behavioral choice of a partner, compared to that when crossing contrary to choice (Drickamer et al., 2000; Gowaty et al., 2007; Nelson et al., 2013; Raveh et al., 2014). However, this result was obtained in experiments where females had direct contact with a partner by hearing, seeing and sniffing them. In these experiments wild-caught animals or their outbred offspring were used. To explain the positive effect of sexual choice, the hypotheses of "good genes" (Kokko, 2001), phenogenetic complementarity of the mother and father (Andersson, 2006), heterozygosity (Ilmonen et al., 2009), and Fisher's "attractive sons" (Kokko, 2001) are used. The basis of all these theories is the choice of a partner based on his genotype. Theoretically, paternal effects may be associated with traits acquired during ontogeny and not dependent on genes, but this theory has not enough evidence at the moment. Therefore, for this study, we chose CD-1 mice, which have genetic diversity and are a frequently used model object of research.

In this study we attempt to determine whether the females will be able to recognize the successfully mated males only by urine tag, and, moreover, what kind of components detected in their urine by gas chromatography correlate with attractiveness of males for females in Olfactory test.

Materials and methods

Mice and sample collection. We used 19 males and 86 (38+48) females of an outbred CD-1 mouse strain (2–3 months old) from the Centre for Genetic Resources of Laboratory Animals at the Institute of Cytology and Genetics, Siberian Branch of RAS (Novosibirsk, Russia) in this study. All animals had SPF-status.

The experiment was carried out in the spring-summer period. Mice were kept for 2 weeks in single-sex groups of four-five animals per standard cage (35 cm × 25 cm × 12 cm) with sawdust bedding at room temperature (20–22 °C) under a 14/10 h light/dark cycle (lights off at 18:00). Water and

food pellets (Zoomir, St. Petersburg, Russia) were available ad libitum. Following the recommendation of Lombardi and Vandenbergh (Lombardi, Vandenbergh, 1977), we added soiled male bedding to female cages daily to support regular estrus cycles. Five days before the experiment males were placed into individual cages. On the day 1 of the experiment, 2 females were placed into each male cage at the time of lights off. Females were housed with males for 6 days, except for mated females (see below). Urine samples were collected on the 6th day. Urine was collected through gentle abdominal massage while the male was held over an Eppendorf microcentrifuge tube. The urine samples were divided into two aliquots, then immediately frozen and stored at –80 °C. Due to the fact that we did not spend more than 1 minute obtaining a urine sample from an individual mouse, only 16 out of 19 secondary aliquots were collected.

In procedures with animals, we used the principles specified in the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes. All animal protocols were approved by the Institutional Bioethics Committee of the Institute of Cytology and Genetics (No. 81).

Male reproductive success. Two females housed with males during 6 days were examined daily 2–3 hours after the lights were turned on for the presence of vaginal plugs. Females with vaginal plugs were removed and housed individually. After 6 days males were removed and sacrificed to assess the testosterone level in blood, body weight, weight of testes, seminal vesicles and preputial glands. Seminal vesicles were removed and weighed together with coagulation glands. A male was considered reproductively successful if at least one female kept with him turned out to be pregnant. We got 7 successful and 9 unsuccessful males. The time of cohabiting a male with females was chosen in accordance with (Gerlinskaya et al., 2012). Since in 96 % of females that have access to the smell of a male, the length of the estrous cycle does not exceed 5.5 days (Jemiolo et al., 1986), within 6 days, a female introduced at any stage of the cycle will be in oestrus at least once, which is necessary for fertile mating.

Olfactory test. Separate females, who did not participate in reproduction, were tested. In behavioral testing, a preliminary acquaintance with the smell of contaminated bedding can have a decisive effect on the ability of females to recognize and increase interest in the volatile components of male urine (Moncho-Bogani et al., 2002). Therefore, bedding contaminated by males was added to the cages of females on daily basis. The day before the experiment, the females were placed individually and the test was performed in a home cage. Urine was thawed for 20 minutes at room temperature. 20 µl of urine were applied to filter paper and placed in a vial (a single-use 5 mm truncated tip of an automatic pipette). That is why the females had access only to the volatile compounds of urine. The tip was fixed to the mesh lid of the cage in the corner. Females observed one accidentally selected urine sample (Dougherty, Shuker, 2015; Dougherty, 2020). 16 samples were examined. During the 10-minute test, the number of approaches to sniff (number of sniffs), the time spent sniffing the sample in seconds (total time of sniffs) and the time of the first approach to the stimulus in seconds (latency time of sniffs) were taken into account. After the test, a swab was taken from the females to determine the stage of the estrous cycle. A sample from each male was tested on 3 females, the data on their testing were averaged for further calculations (or Repeated measure ANOVA was used if specified). The stage of the cycle had a significant effect on the behavior of females: at the proestrus stage, the females approached the urine samples significantly earlier than the females at the diestrus stage ($p = 0.039$, LSD test). There were no significant differences in other stages and no effect on other types of behavior. To exclude the influence of the cycle stage on further statistical analysis, residual variances were used.

Preparation of urine samples and concentration procedure. Mouse urine was thawed and 20 µl of urine from each male was then transferred to 7 ml glass vials with caps containing gastight PTFE/Silicone septum (Supelco). The vials with urine were then heated for 15 min at 40 °C for equilibration, and also for denaturation of urinary proteins that bind some volatiles. Immediately after the heating sample headspace was concentrated on 6 mg of Tenax (Chrompack, Netherlands) using special sorbent traps designed for EKHO-A-PID gas chromatograph (IPGG SB RAS, Novosibirsk, Russia). For concentration we used filtered air flow at flow-rate of 40 ml/min. In total we pumped 80 ml urine headspace and air mix through the sorbent layer.

The samples were thawed and prepared consistently one after another with 20 min intervals, so preparation of samples was done consecutively so each sample was run on the Gas Chromatograph (GC) within 20 min (including 15 min of heating) after defrosting. Filtered air samples were routinely run as a control.

GC analysis of urine VOCs and data preparation. An EKHO-A-PID gas chromatograph with original software (Sorbat, IPGG SB RAS, Novosibirsk, Russia) was used in our study. 19 samples of mouse urine were run in the GC using the non-polar GC column (polidimethylsiloxane polycapillary, $n = 920$ capillaries) SE-30 22 cm × 0.6 mm with 40 µm coating (IPGG SB RAS). The temperature of the column was constant at 50 °C during the whole separation. The temperature of the injection port was set at 180 °C. Filtered ambient air with a constant flow-rate of 20 ml/min was used as a carrier gas throughout the analysis. The duration of each analysis was set at 300 s. For further statistical calculations, we used the amplitudes of the peaks.

MS analysis. Urine samples were treated on a custom-made GC/MS system (IPGG SB RAS, Novosibirsk, Russia), specialized for fast VOC analysis in air with a non-polar column HP-5 similar to SE-30 (Makas, Troshkov, 2004). Heated urine headspace was concentrated on Tenax as described above. A non-polar column HP-5, 15 m × 0.32 mm with 1 µm film (Agilent technologies, USA), was used for separation. The temperature of the column was constant at 45 °C for 5 minutes, then it was programmed with the rate of 10 °C/min to 150 °C. The temperature of the injection port was set at 280 °C. Helium was used as the gas-carrier with a flow-rate of 2 ml/min.

The operating parameters for the mass spectrometer were set as follows: scan rate 0.5 s from 45 to 250 m/z; ion source temperature set to 180 °C, with electron impact ionization energy at 70 eV. Identification of compounds was performed using NIST/EPA/NIH libraries (ver. 2.0.2008) and information from literature (Schwende et al., 1986; Novotny et al., 2007).

The GC/MS data analysis was performed by the AMDIS program (NIST, USA).

Additionally retention times of several standard compounds were analyzed on both columns to use them as reference points in subsequent procedures. The standard compounds were purchased from Acros Organics (Belgium) and Sigma-Aldrich (USA). For each peak in question we found reference compounds among pure chemicals to make the retention time of the reference compound as close as possible to the retention time of the target peak. Thus, we were able to surround the target peaks on the chromatogram with one or two reference points. For this purpose benzene, toluene, m-xylene, 2,5-dimethylpyrazine and nonane were chosen. Thus, we were able to confine the intervals where the target peaks could be found. Next, we compared the area and the amplitude of the peaks detected in the localized intervals with the same parameters of the target peaks obtained on EKHO-A-PID and found three peaks (RT 4.2, RT 7.9 and RT 87.7) caught on GC/MS that satisfied all requirements. We identified the peaks as three dihydrofuran (DHF) derivatives – the dehydratation products of lactol – using data reported in the literature. The characteristic losses of m/z 126, m/z 111, m/z 97, m/z 83, m/z 69, and m/z 57 were identical to those obtained in the earlier studies with synthetic analogues of cyclic enol ethers (Novotny et al., 2007). Moreover, the ratio of target peaks in our study (1 : 0.32 : 0.12) was nearly the same as the ratio of cyclic enol ethers (1 : 0.30 : 0.10) in mouse urine calculated in one of the studies mention above (Harvey et al., 1989). This provides additional confidence that our target compounds were dihydrofuran derivatives: 5,5-dimethyl-2-ethyl-4,5-DHF, E-5,5-dimethyl-2-ethylenetetrahydrofuran, and Z-5,5-dimethyl-2-ethylenetetrahydrofuran.

Statistical analysis. To analyze behavior, we used Repeated measure ANOVA, since each male sample was tested by 3 females. To compare the groups by the content of components, we used One way ANOVA. We used Spearman correlation coefficient to count correlation ratios in this study. No data were removed from calculations. All experimental data were obtained blindly, the belonging of the animals to the groups was indicated only at the stage of data analysis. The level of significance used was $p < 0.05$.

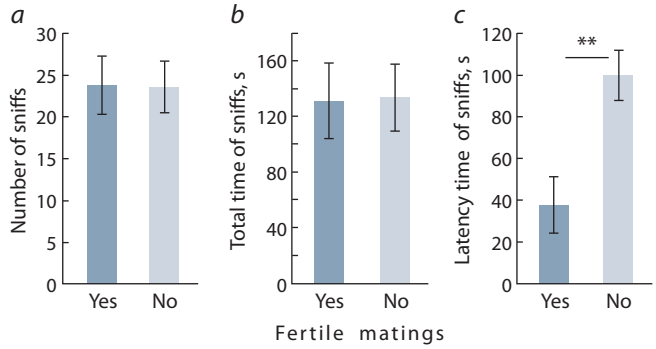


Fig. 1. Behavioral reaction of naive females on smell of urine of male mice with various number of fertile matings: *a*, number of sniffs; *b*, total time of sniffs; *c*, latency time of sniffs.

N(No) = 9♂; *N*(Yes) = 7♂; ** $p < 0.01$, Repeated measure ANOVA.

Results

In Olfactory test, females had a shorter latency time of sniffs when studying urine samples from males who had already successfully procreated offsprings after 6 days of being kept with other females $F_{1,14} = 0.00$, $p = 0.963$ (Fig. 1, *c*). The presence of fertile matings in a male did not have a significant effect on the number of sniffs ($F_{1,14} = 0.00$, $p = 0.963$) and total time of sniffs ($F_{1,14} = 0.00$, $p = 0.945$).

Chromatographic study of volatile components of male urine samples revealed 12 peaks (Table 1). To understand whether the behavioral response of females is really related to differences in the content of the detected components, we

Table 1. The relationship between the response of females in Olfactory test and the amplitude of chromatographic peaks in the urine of males

Peaks' retention time (RT)	Behavioral traits		
	Number of sniffs *N* = 16	Total time of sniffs *N* = 16	Latency time of sniffs *N* = 16
2.9	$r_S = -0.25$, $p = 0.351$	$r_S = -0.14$, $p = 0.608$	$r_S = -0.31$, $p = 0.247$
3.3	$r_S = -0.23$, $p = 0.398$	$r_S = -0.17$, $p = 0.537$	$r_S = 0.16$, $p = 0.545$
4.2	$r_S = 0.04$, $p = 0.880$	$r_S = 0.15$, $p = 0.587$	$r_S = -0.48$, $p = 0.063$
5.4	$r_S = -0.39$, $p = 0.138$	$r_S = -0.04$, $p = 0.879$	$r_S = -0.03$, $p = 0.927$
7.9	$r_S = 0.00$, $p = 0.991$	$r_S = 0.04$, $p = 0.897$	**$r_S = -0.57$, $p = 0.021$**
9.9	$r_S = -0.38$, $p = 0.147$	$r_S = -0.15$, $p = 0.579$	$r_S = -0.27$, $p = 0.305$
56.8	$r_S = -0.48$, $p = 0.060$	$r_S = 0.09$, $p = 0.753$	$r_S = -0.18$, $p = 0.513$
64.8	$r_S = -0.03$, $p = 0.897$	$r_S = 0.08$, $p = 0.762$	$r_S = -0.06$, $p = 0.820$
74.8	$r_S = -0.06$, $p = 0.824$	$r_S = 0.12$, $p = 0.652$	$r_S = 0.13$, $p = 0.637$
87.7	$r_S = 0.04$, $p = 0.880$	$r_S = -0.03$, $p = 0.922$	**$r_S = -0.68$, $p = 0.004$**
110.6	$r_S = 0.03$, $p = 0.914$	$r_S = -0.24$, $p = 0.368$	$r_S = -0.42$, $p = 0.107$
170.45	$r_S = 0.29$, $p = 0.272$	$r_S = -0.01$, $p = 0.983$	$r_S = -0.47$, $p = 0.063$

Note. Significant differences are given in bold.

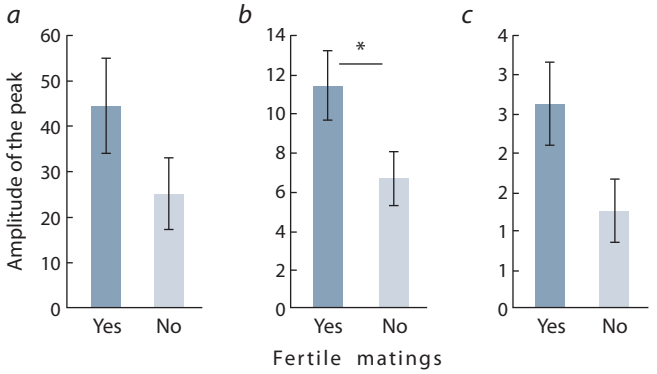


Fig. 2. Amplitude differences in peaks RT 4.2 (*a*), RT 7.9 (*b*) and RT 87.7 (*c*) found in the urine of male mice that had or didn't have fertile mating.
N(0) = 12♂; *N*(1) = 7♂; * $p < 0.05$, One way ANOVA.

performed a correlation analysis of the behavior characteristics and the obtained peaks of urine. A significant negative correlation was found between the latency time of sniffs and the amplitude of peaks RT 7.9 and RT 87.7 (see Table 1). No significant relationship was found between the amplitude of the peaks and the number of sniffs, as well as the total time of sniffs.

Mass spectrometric analysis of male urine samples showed that the RT 7.9 peak corresponds to the known compound E-5,5-dimethyl-2-ethylenetetrahydrofuran and the RT 87.7 peak corresponds to Z-5,5-dimethyl-2-ethylenetetrahydrofuran. At the same time, one more dihydrofuran derivative was identified: 5,5-dimethyl-2-ethyl-4,5-DHF, RT 4.2. This compound did not show a significant correlation value with the behavioral characteristics of females (see Table 1). All three compounds are derivatives of the known male mouse pheromone 6-hydroxy-6-methyl-3-heptanone (HMH) (Harvey et al., 1989).

We evaluated the relationship between the amplitude of the chromatographic peaks of dihydrofurans and the reproductive success of males. The amplitude of the RT 7.9 peak was found to be significantly higher in the urine of males who had fertilized at least one female, compared to the urine of males who had fertilized no females (Fig. 2, Fig. 3). The RT 4.2 and RT 87.7 peaks showed the same trend, but the *p*-value was below the threshold of statistical significance ($F_{1,17} = 2.19$, $p = 0.157$ and $F_{1,17} = 4.20$, $p = 0.056$, respectively).

Analysis of the relationship between the amplitude of the three peaks under study and the level of testosterone in the blood plasma showed significant positive correlation of the amplitude of the RT 4.2, RT 7.9 and RT 87.7 peaks. A similar correlation was observed for the relationship between preputial gland weight and the RT 4.2 and RT 7.9 peaks. Correlation analysis of the studied peaks with body weight, weight of seminal vesicles and testes showed no significant values (Table 2).

It is important to note that body weight, weight of testes, seminal vesicles, preputial glands, and plasma testosterone level alone are not reliable indicators of male reproductive success (Table 3).

Discussion

Two decades ago, it was shown for the first time that females can identify males, mating with which leads to greater reproductive success (Drickamer et al., 2000). This result has been repeated many times in different animal species (Drickamer et al., 2003; Gowaty et al., 2007; Nelson et al., 2013; Raveh et al., 2014). Nevertheless, direct contact with a partner and free access to all signals from a potential sexual partner in these experiments did not allow getting closer to the understanding of the selection mechanisms. In the present study, we used only volatile compounds of the male urine to be tested by females, and found that based on the smell of urine only, females could identify reproductively effective males. It turned out that the experimental females in Olfactory test approached the urine of successful males earlier (see Fig. 1, *c*) despite the fact that they performed an equal number of sniffs and had the same total time of sniffs (see Fig. 1, *a*, *b*).

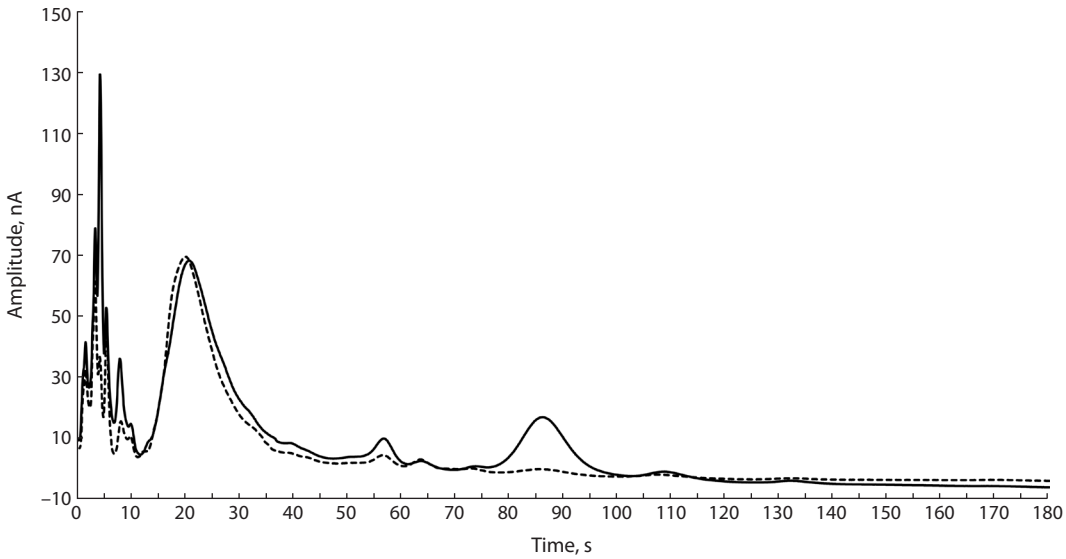


Fig. 3. Sample chromatograms from a male with fertile mating (solid line) and without fertile mating (dotted line).

Table 2. Correlation between physiological characteristics, testosterone level and amplitudes of target peaks (Spearman correlation coefficients)

Physiological characteristics	Target peaks		
	RT 4.2 *N* = 19	RT 7.9 *N* = 19	RT 87.7 *N* = 19
Body weight, g	$r_s = 0.21$, $p = 0.383$	$r_s = 0.17$, $p = 0.491$	$r_s = 0.18$, $p = 0.464$
Weight of testes, mg	$r_s = 0.11$, $p = 0.651$	$r_s = 0.14$, $p = 0.579$	$r_s = 0.10$, $p = 0.681$
Weight of seminal vesicles, mg	$r_s = 0.23$, $p = 0.339$	$r_s = 0.26$, $p = 0.279$	$r_s = 0.09$, $p = 0.713$
Weight of preputial glands, mg	**$r_s = 0.57$, $p = 0.011$**	**$r_s = 0.58$, $p = 0.010$**	**$r_s = 0.48$, $p = 0.038$**
Plasma testosterone, ng/ml	**$r_s = 0.46$, $p = 0.046$**	**$r_s = 0.51$, $p = 0.027$**	$r_s = 0.27$, $p = 0.260$

Note. Significant differences are given in bold.

Table 3. Mean values of physiological characteristics depending on the reproductive success of males

Physiological characteristics	Fertile matings, Mean ± SE	No fertile matings, Mean ± SE	One way ANOVA
Body weight, g	33.70 ± 1.09	32.54 ± 0.52	$F_{1,17} = 1.17$, $p = 0.295$
Weight of testes, mg	201.00 ± 8.29	210.75 ± 6.54	$F_{1,17} = 0.84$, $p = 0.373$
Weight of seminal vesicles, mg	177.57 ± 10.57	181.67 ± 9.15	$F_{1,17} = 0.08$, $p = 0.781$
Weight of preputial glands, mg	70.00 ± 6.79	61.50 ± 2.19	$F_{1,17} = 2.11$, $p = 0.164$
Plasma testosterone, ng/ml	6.28 ± 1.57	7.29 ± 1.35	$F_{1,17} = 0.22$, $p = 0.642$

The search for markers of male reproductive efficiency using gas chromatographic analysis of male urine samples showed that the behavior of females correlated with the amplitude of the dihydrofuran peaks (see Table 1).

Previously, DHFs have already been detected in significant amounts in the chromatograms of male mouse urine (Schwende et al., 1986; Jemiolo et al., 1987; Harvey et al., 1989; Novotny et al., 1999b). While studying the origin of these cyclic enol ethers in mouse urine S. Harvey et al. (Harvey et al., 1989) showed that DHFs originate from the tautomeric mixture of 6-hydroxy-6-methyl-3-heptanone and lactol via dehydratation in the inlet port of gas chromatograph under high temperature. It turned out that DHFs are not presented in the mouse urine by themselves, but their peaks in the chromatogram reflect the content of their precursor 6-hydroxy-6-methyl-3-heptanone (Harvey et al., 1989). Mutual precursors of target compounds explain very high coefficients of intercorrelation between these components, exceeding 0.90 ($p < 0.001$). Thus, the behavior of females correlated with compounds that were previously shown to reflect the content of 6-hydroxy-6-methyl-3-heptanone in urine of males.

When studying the effects of HMH, it turned out that 6-hydroxy-6-methyl-3-heptanone interacts with vomeronasal receptors (Del Punta et al., 2002), and therefore can trigger behavioral and physiological responses in females. HMH is known as a male mouse pheromone that accelerates puberty in female mice (Novotny et al., 1999a). Here we demonstrated that quantity of HMH in urine of a male reflects its ability to make fertile matings (see Fig. 2). Males with a lower level of this pheromone did not mate any of the two females for 6 days of joint maintenance. Perhaps exactly this indirectly explains the effect of a decrease in fertile matings in aged male mice, as shown earlier (Parkening et al., 1988), since HMH decreases in aged males (Osada et al., 2008; Varshavi et al., 2018), and in the work of Schaefer with colleagues HMH has been associated with the maturation state (Schaefer et al., 2010). In our work, when analyzing correlations with androgen-dependent characteristics of males, it turned out that amplitudes of the peaks correlated positively with plasma testosterone and weight of the preputial glands (see Table 2). However, reproductive success was not directly related to physiological characteristics of males (see Table 3). Taken together, these data show that 6-hydroxy-6-methyl-3-heptanone reflects reproductive quality of male mice.

On the other hand, it has previously been shown that HMH does not affect the attractiveness of male urine samples to females (Osada et al., 2008). These conclusions about male attractiveness were based only on the total time of sniffs, which is consistent with our results. We also demonstrated no relationship between the level of this pheromone in urine and the total time of sniffs (see Table 1). In addition to this, when assessing behavioral response of females to the stimulus, we found that the HMH content correlated with the latency time of sniffs (see Table 1), which indicates the importance of assessing other parameters of the females' behavior, and not just the total time of sniffs.

Conclusion

Identification of markers of reproductively successful males is an important task. Its solution does not have only fundamental importance, but will also allow targeted selection of more successful males when breeding animals. Urine, which is an accessible and unlimited resource, is of greatest interest in this regard. The observed relationships between reproduc-

tive success and the HMH content in urine require further research to understand the dynamics of this compound excretion at different stages of male ontogenesis, with different reproductive experience, in different conditions of social and microbiological environment.

References

Andersson M. Condition-dependent indicators in sexual selection: development of theory and tests. In: Lucas J.R., Simmons L.W. (Eds.) Essays in Animal Behaviour: Celebrating 50 Years of Animal Behaviour. Elsevier, 2006;253-267.

Arakawa H., Blanchard D.C., Arakawa K., Dunlap C., Blanchard R.J. Scent marking behavior as an odorant communication in mice. *Neurosci. Biobehav. Rev*. 2008;32(7):1236-1248. DOI 10.1016/j.neubiorev.2008.05.012.

Brennan P.A., Zufall F. Pheromonal communication in vertebrates. *Nature.* 2006;444(7117):308-315. DOI 10.1038/nature05404.

Del Punta K., Leinders-Zufall T., Rodriguez I., Jukam D., Wysocki C.J., Ogawa S., Zufall F., Mombaerts P. Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. *Nature*. 2002;419(6902):70-74. DOI 10.1038/nature00955.

Dougherty L.R. Designing mate choice experiments. *Biol. Rev*. 2020;95(3):759-781. DOI 10.1111/brv.12586.

Dougherty L.R., Shuker D.M. The effect of experimental design on the measurement of mate choice: a meta-analysis. *Behav. Ecol*. 2015;26(2):311-319. DOI 10.1093/beheco/aru125.

Drickamer L.C. Behavioral selection of odor cues by young female mice affects age of puberty. *Dev. Psychobiol*. 1992;25(6):461-470. DOI 10.1002/dev.420250606.

Drickamer L.C., Gowaty P.A., Holmes C.M. Free female mate choice in house mice affects reproductive success and offspring viability and performance. *Anim. Behav*. 2000;59(2):371-378. DOI 10.1006/anbe.1999.1316.

Drickamer L.C., Gowaty P.A., Wagner D.M. Free mutual mate preferences in house mice affect reproductive success and offspring performance. *Anim. Behav*. 2003;65(1):105-114. DOI 10.1006/anbe.2002.2027.

Gerlinskaya L.A., Shnayder E.P., Dotsenko A.S., Maslennikova S.O., Zavjalov E.L., Moshkin M.P. Antigen-induced changes in odor attractiveness and reproductive output in male mice. *Brain Behav. Immun*. 2012;26(5):451-458. DOI 10.1016/j.bbi.2011.11.010.

Gowaty P.A., Anderson W.W., Bluhm C.K., Drickamer L.C., Kim Y.K., Moore A.J. The hypothesis of reproductive compensation and its assumptions about mate preferences and offspring viability. *Proc. Natl. Acad. Sci. USA*. 2007;104(38):15023-15027. DOI 10.1073/pnas.0706622104.

Harvey S., Wiesler D., Novotny M. Formation of cyclic enol ethers from a labile biological precursor: an example of analytical artifacts. *J. Chromatogr*. 1989;491(1):27-36. DOI 10.1016/S0378-4347(00)82816-9.

Hurst J.L., Beynon R.J. Scent wars: the chemobiology of competitive signalling in mice. *Bioessays*. 2004;26(12):1288-1298. DOI 10.1002/bies.20147.

Ilmonen P., Stundner G., Thoß M., Penn D.J. Females prefer the scent of outbred males: good-genes-as-heterozygosity? *BMC Evol. Biol*. 2009;9:104. DOI 10.1186/1471-2148-9-104.

Jemiolo B., Alberts J., Sochinski-Wiggins S., Harvey S., Novotny M. Behavioral and endocrine responses of female mice to synthetic analogues of volatile compounds in male urine. *Anim. Behav*. 1985;33(4):1114-1118. DOI 10.1016/S0003-3472(85)80170-6.

Jemiolo B., Harvey S., Novotny M. Promotion of the Whitten effect in female mice by synthetic analogs of male urinary constituents. *Proc. Natl. Acad. Sci. USA*. 1986;83(12):4576-4579. DOI 10.1073/pnas.83.12.4576.

Jemiolo B., Andreolini F., Wiesler D., Novotny M. Variations in mouse (*Mus musculus*) urinary volatiles during different periods of pregnancy and lactation. *J. Chem. Ecol*. 1987;13(9):1941-1956. DOI 10.1007/BF01014677.

Jemiolo B., Xie T.M., Novotny M. Socio-sexual olfactory preference in female mice: attractiveness of synthetic chemosignals. *Physiol. Behav.* 1991;50(6):1119-1122. DOI 10.1016/0031-9384(91)90570-E.

Kavaliers M., Colwell D.D. Discrimination by female mice between the odours of parasitized and non-parasitized males. *Proc. R. Soc. Lond. B.* 1995;261(1360):31-35. DOI 10.1098/rspb.1995.0113.

Kokko H. Fisherian and "good genes" benefits of mate choice: how (not) to distinguish between them. *Ecol. Lett*. 2001;4(4):322-326. DOI 10.1046/j.1461-0248.2001.00224.x.

Lin D.Y., Zhang S.Z., Block E., Katz L.C. Encoding social signals in the mouse main olfactory bulb. *Nature*. 2005;434(7032):470-477. DOI 10.1038/nature03414.

Liu Y.J., Guo H.F., Zhang J.X., Zhang Y.H. Quantitative inheritance of volatile pheromones and darcin and their interaction in olfactory preferences of female mice. *Sci. Rep*. 2017;7(1):2094. DOI 10.1038/s41598-017-02259-1.

Lombardi J.R., Vandenbergh J.G. Pheromonally induced sexual maturation in females: regulation by the social environment of the male. *Science*. 1977;196(4289):545-546. DOI 10.1126/science.557838.

Ma W., Miao Z., Novotny M.V. Induction of estrus in grouped female mice (*Mus domesticus*) by synthetic analogues of preputial gland constituents. *Chem. Senses*. 1999;24(3):289-293. DOI 10.1093/chemse/24.3.289.

Makas A.L., Troshkov M.L. Field gas chromatography-mass spectrometry for fast analysis. *J. Chromatogr. B*. 2004;800(1-2):55-61. DOI 10.1016/j.jchromb.2003.08.054.

Manser A., König B., Lindholm A.K. Female house mice avoid fertilization by *t* haplotype incompatible males in a mate choice experiment. *J. Evol. Biol*. 2015;28(1):54-64. DOI 10.1111/jeb.12525.

Moncho-Bogani J., Lanuza E., Hernández A., Novejarque A., Martínez-García F. Attractive properties of sexual pheromones in mice: innate or learned? *Physiol. Behav*. 2002;77(1):167-176. DOI 10.1016/S0031-9384(02)00842-9.

Moshkin M.P., Kolosova I.E., Novikov E.A., Litvinova E.A., Mershieva L.V., Mak V.V., Petrovskii D.V. Co-modulation of the immune function and the reproductive chemosignals. *Asian-Aust. J. Anim. Sci*. 2001;14:43-51.

Moshkin M.P., Gerlinskaya L.A., Morozova O., Bakhvalova V., Evsikov V. Behaviour, chemosignals and endocrine functions in male mice infected with tick-borne encephalitis virus. *Psychoneuroendocrinology*. 2002;27(5):603-608. DOI 10.1016/S0306-4530(01)00096-8.

Nelson A.C., Colson K.E., Harmon S., Potts W.K. Rapid adaptation to mammalian sociality via sexually selected traits. *BMC Evol. Biol*. 2013;13:81. DOI 10.1186/1471-2148-13-81.

Novotny M., Harvey S., Jemiolo B., Alberts J. Synthetic pheromones that promote inter-male aggression in mice. *Proc. Natl. Acad. Sci. USA*. 1985;82(7):2059-2061. DOI 10.1073/pnas.82.7.2059.

Novotny M., Harvey S., Jemiolo B. Chemistry of male dominance in the house mouse, *Mus domesticus*. *Experientia*. 1990;46(1):109-113. DOI 10.1007/BF01955433.

Novotny M.V., Jemiolo B., Wiesler D., Ma W., Harvey S., Xu F., Xie T., Carmack M. A unique urinary constituent, 6-hydroxy-6-methyl-3-heptanone, is a pheromone that accelerates puberty in female mice. *Chem. Biol*. 1999a;6(6):377-383. DOI 10.1016/S1074-5521(99)80049-0.

Novotny M.V., Ma W., Wiesler D., Zidek L. Positive identification of the puberty-accelerating pheromone of the house mouse: the volatile ligands associating with the major urinary protein. *Proc. R. Soc. Lond. B.* 1999b;266(1432):2017-2022. DOI 10.1098/rspb.1999.0880.

Novotny M.V., Soini H.A., Koyama S., Wiesler D., Bruce K.E., Penn D.J. Chemical identification of MHC-influenced volatile compounds in mouse urine. I: Quantitative proportions of major chemosignals. *J. Chem. Ecol*. 2007;33(2):417-434. DOI 10.1007/s10886-006-9230-9.

Osada K., Yamazaki K., Curran M., Bard J., Smith B.P., Beauchamp G.K. The scent of age. *Proc. R. Soc. Lond. B.* 2003;270(1518):929-933. DOI 10.1098/rspb.2002.2308.

Osada K., Tashiro T., Mori K., Izumi H. The identification of attractive volatiles in aged male mouse urine. *Chem. Senses*. 2008;33(9):815-823. DOI 10.1093/chemse/bjn045.

Parkening T.A., Collins T.J., Au W.W. Paternal age and its effects on reproduction in C57BL/6NNia mice. *J. Gerontol*. 1988;43(3): B79-B84. DOI 10.1093/geronj/43.3.B79.

Penn D.J., Potts W.K. Untrained mice discriminate MHC-determined odors. *Physiol. Behav*. 1998;64(3):235-243. DOI 10.1016/S0031-9384(98)00052-3.

Raveh S., Sutalo S., Thonhauser K.E., Thoß M., Hettyey A., Winkelser F., Penn D.J. Female partner preferences enhance offspring ability to survive an infection. *BMC Evol. Biol.* 2014;14:14. DOI 10.1186/1471-2148-14-14.

Roberts S.C., Gosling L.M. Genetic similarity and quality interact in mate choice decisions by female mice. *Nat. Genet*. 2003;35(1):103-106. DOI 10.1038/ng1231.

Röck F., Hadeler K.P., Rammensee H.G., Overath P. Quantitative analysis of mouse urine volatiles: in search of MHC-dependent differences. *PLoS One*. 2007;2(5):e429. DOI 10.1371/journal.pone.0000429.

Schaefer M.L., Wongravee K., Holmboe M.E., Heinrich N.M., Dixon S.J., Zeskind J.E., Kulaga H.M., Brereton R.G., Reed R.R., Trevejo J.M. Mouse urinary biomarkers provide signatures of maturation, diet, stress level, and diurnal rhythm. *Chem. Senses*. 2010; 35(6):459-471. DOI 10.1093/chemse/bjq032.

Schwende F.J., Wiesler D., Jorgenson J.W., Carmack M., Novotny M. Urinary volatile constituents of the house mouse, *Mus musculus*, and their endocrine dependency. *J. Chem. Ecol*. 1986;12(1):277-296. DOI 10.1007/BF01045611.

Varshavi D., Scott F.H., Varshavi D., Veeravalli S., Phillips I.R., Veselkov K., Strittmatter N., Takats Z., Shephard E.A., Everett J.R. Metabolic biomarkers of ageing in C57BL/6J wild-type and flavin-containing monooxygenase 5 (FMO5)-knockout mice. *Front. Mol. Biosci*. 2018;5:28. DOI 10.3389/fmolb.2018.00028.

Veyrac A., Wang G., Baum M.J., Bakker J. The main and accessory olfactory systems of female mice are activated differentially by dominant versus subordinate male urinary odors. *Brain Res*. 2011; 1402:20-29. DOI 10.1016/j.brainres.2011.05.035.

Whitten W.K., Bronson F.H., Greenstein J.A. Estrus-inducing pheromone of male mice: transport by movement of air. *Science*. 1968; 161(3841):584-585. DOI 10.1126/science.161.3841.5.

Willis C., Poulin R. Preference of female rats for the odours of non-parasitized males: the smell of good genes. *Folia Parasitol*. 2000; 47(1):6-10. DOI 10.14411/fp.2000.002.

Zala S.M., Potts W.K., Penn D.J. Scent-marking displays provide honest signals of health and infection. *Behav. Ecol*. 2004;15(2):338-344. DOI 10.1093/beheco/arh022.

Zala S.M., Bilak A., Perkins M., Potts W.K., Penn D.J. Female house mice initially shun infected males, but do not avoid mating with them. *Behav. Ecol. Sociobiol*. 2015;69:715-722. DOI 10.1007/s00265-015-1884-2.

Zhang J.X., Rao X.P., Sun L., Zhao C.H., Qin X.W. Putative chemical signals about sex, individuality, and genetic background in the preputial gland and urine of the house mouse (*Mus musculus*). *Chem. Senses*. 2007;32(3):293-303. DOI 10.1093/chemse/bjl058.

ORCID ID

A.S. Khotskina orcid.org/0000-0001-5379-9977
E.L. Zavjalov orcid.org/0000-0002-9412-3874
L.A. Gerlinskaya orcid.org/0000-0002-8118-1362
D.V. Petrovskii orcid.org/0000-0002-0623-0363
M.P. Moshkin orcid.org/0000-0002-5388-2946

Funding. The study of females behavior was supported by the budget project (FWNR-2022-0004). The GC/MS study was supported by the Russian Science Foundation grant No. 20-14-00055. The studies are implemented using the equipment of the Center for Genetic Resources of Laboratory Animals at ICG SB RAS, supported by the Ministry of Science and Higher Education of Russian Federation (Unique identifier of the project RFMEFI62119X0023).

Acknowledgements. We thank Professor Milos Novotny for the data of mass-spectra of synthetic cyclic enol ethers obtained in his laboratory in Indiana University.

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

Received December 27, 2022. Revised March 10, 2023. Accepted March 13, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):488-494
DOI 10.18699/VJGB-23-59

Original Russian text https://vavilovj-icg.ru/

Differentiation of *Bos grunniens* and *Bos taurus* based on STR locus polymorphism

K.B. Chekirov[1], Zh.T. Isakova[2] ✉, V.N. Kipen[3], M.I. Irsaliev[2], S.B. Mukeeva[2], K.A. Aitbaev[2], G.A. Sharshenalieva[4], S.B. Beyshenalieva[4], B.U. Kydyralieva[1]

[1] Kyrgyz-Turkish Manas University, Bishkek, Kyrgyz Republic
[2] Research Institute of Molecular Biology and Medicine, Bishkek, Kyrgyz Republic
[3] Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus
[4] Kyrgyz State University named after I. Arabaev, Bishkek, Kyrgyz Republic
✉ jainagul@mail.ru

Abstract. Differentiation of closely related biological species using molecular genetic analysis is important for breeding farm animals, creating hybrid lines, maintaining the genetic purity of breeds, lines and layering. *Bos grunniens* and *Bos taurus* differentiation based on STR locus polymorphism will help maintain the genetic isolation of these species and identify hybrid individuals. The aim of this study is to assess the differentiating potential of 15 microsatellite loci to distinguish between domestic yak (*B. grunniens*) bred in the Kalmak-Ashuu highland region (Kochkor district, Naryn region, Kyrgyz Republic) and cattle (*B. taurus*) of three breeds (Aberdeen-Angus, Holstein and Alatau) using molecular genetic analysis. The samples were genotyped at 15 microsatellite loci (ETH3, INRA023, TGLA227, TGLA126, TGLA122, SPS115, ETH225, TGLA53, BM2113, BM1824, ETH10, BM1818, CSSM66, ILSTS006 and CSRM60). Twelve of the loci were from the standard markers panel recommended by ISAG. Statistical analysis was performed using GenAlEx v.6.503, Structure v.2.3.4, PAST v.4.03, and POPHELPER v1.0.10. The analysis of the samples' subpopulation structure using the Structure v.2.3.4 and 15 STR locus genotyping showed that the accuracy of assigning a sample to *B. taurus* was 99.6 ± 0.4 %, whereas the accuracy of assigning a sample to *B. grunniens* was 99.2 ± 2.6 %. Of the 15 STRs, the greatest potential to differentiate *B. grunniens* and *B. taurus* was found in those with the maximal calculated FST values, including BM1818 (0.056), BM1824 (0.041), BM2113 (0.030), CSSM66 (0.034) and ILSTS006 (0.063). The classification accuracy of *B. grunniens* using only these five microsatellite loci was 98.8 ± 3.4 %, similar for *B. taurus*, 99.1 ± 1.2 %. The proposed approach, based on the molecular genetic analysis of 5 STR loci, can be used as an express test in Kyrgyzstan breeding and reproduction programs for *B. grunniens*.
Key words: domestic yak; *Bos grunniens*; cattle; *Bos taurus*; DNA; microsatellite markers; STR; genotyping; differentiation.

For citation: Chekirov K.B., Isakova Zh.T., Kipen V.N., Irsaliev M.I., Mukeeva S.B., Aitbaev K.A., Sharshenalieva G.A., Beyshenalieva S.B., Kydyralieva B.U. Differentiation of *Bos grunniens* and *Bos taurus* based on STR locus polymorphism. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):488-494. DOI 10.18699/VJGB-23-59

Дифференциация *Bos grunniens* и *Bos taurus* на основании полиморфизма STR-локусов

К.Б. Чекиров[1], Ж.Т. Исакова[2] ✉, В.Н. Кипень[3], М.И. Ирсалиев[2], С.Б. Мукеева[2], К.А. Айтбаев[2], Г.А. Шаршеналиева[4], С.Б. Бейшеналиева[4], Б.У. Кыдыралиева[1]

[1] Кыргызско-Турецкий университет «Манас», Бишкек, Кыргызская Республика
[2] Научно-исследовательский институт молекулярной биологии и медицины, Бишкек, Кыргызская Республика
[3] Институт генетики и цитологии Национальной академии наук Беларуси, Минск, Республика Беларусь
[4] Кыргызский государственный университет им. И. Арабаева, Бишкек, Кыргызская Республика
✉ jainagul@mail.ru

Аннотация. Дифференциация близкородственных биологических видов с использованием методов молекулярно-генетического анализа имеет важное значение для селекционных процессов при разведении сельскохозяйственных животных, создании гибридных линий, поддержании генетической чистоты пород, линий, отводок. Подход к дифференциации *Bos grunniens* и *Bos taurus* на основании полиморфизма STR-локусов будет способствовать поддержанию генетической обособленности данных видов и, как следствие, выявлению гибридных особей. Целью исследования была оценка дифференцирующего потенциала 15 микросателлитных локусов для различения особей домашнего яка (*B. grunniens*), разводимых в высокогорном регионе Калмак-Ашуу (Кочкорский район, Нарынская область, Кыргызская Республика), и крупного рогатого скота (*B. taurus*) трех пород (абердин-ангусская, голштинская и алатауская) с использованием молекулярно-генетического анализа. Образцы были генотипированы по 15 микросателлитным локусам (ETH3, INRA023, TGLA227, TGLA126, TGLA122, SPS115, ETH225, TGLA53, BM2113, BM1824, ETH10, BM1818, CSSM66, ILSTS006 и CSRM60). Двенадцать из рассматриваемых

STR-локусов составляли стандартную панель маркеров, рекомендованную ISAG. Статистический анализ данных проводили с использованием программ GenAlEx v.6.503, Structure v.2.3.4, PAST v.4.03 и POPHELPER v1.0.10. Анализ субпопуляционной структуры исследуемых выборок в программе Structure v.2.3.4 по данным генотипирования 15 STR-локусов показал, что точность отнесения образца к *B. taurus* составила 99.6 ± 0.4 %, к *B. grunniens* – 99.2 ± 2.6 %. Наибольшим потенциалом для дифференциации *B. grunniens* и *B. taurus* обладали те локусы, для которых рассчитанные значения показателя FST оказались максимальными – BM1818 (0.056), BM1824 (0.041), BM2113 (0.030), CSSM66 (0.034) и ILSTS006 (0.063). Точность классификации *B. grunniens* с использованием только этих пяти микросателлитных локусов составила 98.8 ± 3.4 %, *B. taurus* – 99.1 ± 1.2 %. Предложенный нами подход, основанный на молекулярно-генетическом анализе пяти STR-локусов, может быть использован в качестве экспресс-теста в селекционных и воспроизводительных программах Кыргызстана для *B. grunniens*.
Ключевые слова: домашний як; *Bos grunniens*; крупный рогатый скот; *Bos taurus*; ДНК; микросателлитные маркеры; STR; генотипирование; дифференциация.

Introduction

Kyrgyzstan is characterized by a variety of natural and climatic conditions, therefore, animal husbandry may vary a lot between locations. Thus, breeding yaks at high altitude makes much sense given that the natural conditions are favorable for the species. In contrast, breeding cattle develops more at low and middle altitudes. Compared to cattle, yaks use low-growth pasture feed better, and in winter they extract it from under a snow cover 10–15 cm thick. Yak meat is not inferior to beef and is rich in proteins, as well as trace elements vital for humans. Although yak milk output is low, their milk is known for the high content of fat (5.5–8.6 %), phosphorus (0.28 %) and calcium (0.30 %) (Abdykerimov, 2001). The yak not only produces milk, meat, skin and wool, but is also used for transport by people in the highlands of Asia (Chertkiev, Chortonbaev, 2007). Yak is a very strong alternative for domestic cattle, easy to breed at high altitude with a very harsh and cold climate. Yaks have thick subcutaneous fat, covered with thick long hair, as well as sharp "steel" hooves that allow them to move along very steep, rocky trails, unattainable for any other livestock.

Unlike the common cattle, which is currently bred on all continents, domestic yak has a very small geographical distribution area, which is limited to the mountainous regions of Central Asia (Jacques et al., 2021). The reason for this, according to (Luz, 1936), is the animal itself. Domestic yak, as well as its wild relative, the Tibetan yak, is perfectly adapted to the conditions of high altitude and mountain plateaus (Lyz, 1936). They both live in the harsh climate of the highlands, where the annual temperature is close to zero for more than eight months a year, and the minimal temperature can drop to –50 °C. In such harsh conditions, yaks live all year round in the open air on pasture.

One of the ways to further intensify yak breeding as an independent branch of animal husbandry is to improve breeding technology, yak breeding and productive qualities, expand knowledge on their biology, as well as increase meat productivity. The study of yak genetic characteristics allowing them to live in the harsh climate of the highlands is of great practical interest.

Currently, the most convenient genetic markers describing genetic structure of different animal species, including yaks and cattle, are polymorphic microsatellite DNA loci (STR, short tandem repeat), which have a codominant nature of inheritance and serve as an indispensable tool to study genetic differences not only between animals, but also populations of the same breeds, as well as between breeds.

The aim of this study is to evaluate the differentiating potential of 15 STR loci (ETH3, INRA023, TGLA227, TGLA126, TGLA122, SPS115, ETH225, TGLA53, BM2113, BM1824, ETH10, BM1818, CSSM66, ILSTS006 and CSRM60) to distinguish individuals of *Bos grunniens* and *Bos taurus*.

Materials and methods

The biological material for molecular genetic research was blood samples taken from adult livestock, including 55 domestic yaks (*B. grunniens*) bred in the Kalmak-Ashuu highland region (Kochkor district, Naryn region, Kyrgyz Republic), which comprised a sample called YAK, as well as blood DNA samples taken from an adult herd of 145 cows (*B. taurus*) of three breeds, including Aberdeen-Angus (n = 45, sample ABR), Holstein (n = 50, sample HOL) and Alatau (n = 50, sample ALA). All applicable international, national and/or institutional principles for the care and use of animals have been observed.

DNA isolation was carried out by phenol-chloroform extraction (Sambrook, Russell, 2001). The samples were genotyped by 15 microsatellite loci. Of the analyzed STR loci, 12 were the standard markers panel recommended by the International Society of Animal Genetics (ISAG), including ETH3, INRA023, TGLA227, TGLA126, TGLA122, SPS115, ETH225, TGLA53, BM2113, BM1824, ETH10 and BM1818. Microsatellite loci CSSM66, ILSTS006 and CSRM60 were analyzed additionally. Oligonucleotides sequence is shown in Table 1.

PCR was analyzed using capillary electrophoresis via an automatic genetic analyzer with a laser-induced fluorescence detection Applied Biosystems 3500 (ThermoFisher, USA). Samples validated using the COrDIS Cattle kit (LLC "GORDIZ", Russian Federation) were used as a reference for allelic calculation.

Statistical analysis was carried out using GenAlEx v.6.503 (Peakall, Smouse, 2012), Structure v.2.3.4 (Pritchard et al., 2000), PAST v.4.03 (Hammer et al., 2001) and POPHELPER v1.0.10 (Francis, 2016). GenAlEx v.6.503 was used to estimate genetic distances using the AMOVA (analysis of molecular variation) method; Structure v.2.3.4 was used to calculate the Q criterion, which characterizes the attribution of each individual to the corresponding cluster (subgroup within the group); POPHELPER v1.0.10 web application was utilized

Table 1. Oligonucleotides sequence for 15 STR loci

STR locus	Primer-F (5′>3′)	Primer-R (5′>3′)	Reference
CSSM66	AATTTAATGCACTGAGGAGCTTGG	ACACAAATCCTTTCTGCCAGCTGA	Barendse et al., 1994
BM1824	GAGCAAGGTGTTTTTCCAATC	CATTCTCCAACTGCTTCCTTG	
SPS115	AAAGTGACACAACAGCTTCACCAG	AACCGAGTGTCCTAGTTTGGCTGTG	Bovine Genome Project
CSRM60	AAGATGTGATCCAAGAGAGAGGCA	AGGACCAGATCGTGAAAGGCATAG	
BM1818	AGCTGGGAATATAACCAAAGG	AGTGCTTTCAAGGTCCATGC	Bishop et al., 1994
ILSTS006	TGTCTGTATTTCTGCTGTGG	ACACGGAAGCGATCTAAACG	Brezinsky et al., 1993
TGLA227	GGAATTCCAAATCTGTTAATTTGCT	ACAGACAGAAACTCAATGAAAGCA	Georges, Massey, 1992
TGLA126	CTAATTTAGAATGAGAGAGGCTTCT	TTGGTCCTCTATTCTCTGAATATTCC	
TGLA122	AATCACATGGCAAATAAGTACATAC	CCCTCCTCCAGGTAAATCAGC	
TGLA53	GCTTTCAGAAATAGTTTGCATTCA	ATCTTCACATGATATTACAGCAGA	
ETH3	GAACCTGCCTCTCCTGCATTGG	ACTCTGCCTGTGGCCAAGTAGG	Toldo et al., 1993
ETH10	GTTCAGGACTGGCCCTGCTAACA	CCTCCAGCCCACTTTCTCTTCTC	
ETH225	GATCACCTTGCCACTATTTCCT	ACATGACAGCCAGCTGCTACT	Steffen et al., 1993
BM2113	GCTGCCTTCTACCAAATACCC	CTTCCTGAGAGAAGCAACACC	
INRA023	GAGTAGAGCTACAAGATAAACTTC	TAACTACAGGGTGTTAGATGAACTC	Vaiman et al., 1994

for graphical interpretation of results obtained in Structure v.2.3.4, whereas PAST v.4.03 was used to plot the main components based on the calculation of genetic distances using the AMOVA method.

Results and discussion

The analysis of the subpopulation structure of *B. grunniens* and *B. taurus* using the Structure v.2.3.4 program on the genotyping data of 15 STR loci, as well as the graphical representation showing the assignment of individuals to a specific group, produced by the POPHELPER v1.0.10 web application, is shown in Figure 1.

As a result of the simulation (the duration of the burn-in 5000, the number of MCMC (Markov chain Monte Carlo) repetitions after the burn-in 50,000, 10 iterations), we found four distinct clusters (K = 4, ΔK = 83.2). Structure v.2.3.4, according to the method of J.K. Pritchard (Pritchard et al., 2000), allowed to compute the Q criterion, which characterized the assignment of each individual to a group (species) for four samples, including ABR, HOL, ALA and YAK. Q ≥ 75 % in the ABR sample was found in 88.9 % (40/45) individuals, accuracy 94.4 ± 5.7 %; HOL, in 82.0 % (41/50), accuracy 95.8 ± 3.3 %; ALA, in 90.0 % (45/50), accuracy 96.3 ± 4.3 %; and YAK, in 98.2 % (54/55), accuracy 98.3 ± 3.2 %. When combining three *B. taurus* samples into one (COW) and analyzing only two groups – COW and YAK, Q ≥ 75 % was identified in 100 % (145/145) individuals, accuracy 99.6 ± 0.4 % in the first group and in 98.2 % (54/55), accuracy 99.2 ± 2.6 % in the second group.

Based on the genetic distances analysis calculated using the AMOVA algorithm, we constructed a graph of principal component analysis (PCA) (Fig. 2). COW and YAK groups on the graph are spaced relative to each other and form two non-overlapping arrays.

Of all 15 STR loci analyzed in the study, including ETH3, INRA023, TGLA227, TGLA126, TGLA122, SPS115, ETH225, TGLA53, BM2113, BM1824, ETH10, BM1818, CSSM66, ILSTS006 and CSRM60, those with the highest calculated F_{ST} had the greatest potential to differentiate *B. grunniens* and *B. taurus* (Table 2).

A similar approach aimed to develop an algorithm to differentiate evolutionarily close animals using STR loci was described in (Rębała et al., 2016; Nosova et al., 2020).

The highest calculated F_{ST} values are shown for BM1818, BM1824, BM2113, CSSM66 and ILSTS006 STR loci. Table 3 summarizes the allelic diversity and allele frequency for the STR loci listed above.

As a result, the representation of major alleles was very different between COW and YAK groups. In particular, ‘256’, ‘258’ and ‘262’ (the total frequency of prevalence was 84.1 %) were the major alleles for BM1818 in the COW group, whereas ‘262’ (occurrence 56.4 %) was major for the YAK group. For BM1824, the difference in the frequency of ‘195’ allele in two groups was 24.0 % (COW – 18.3 %, YAK – 41.8 %), and 22.0 % (COW – 35.2 %, YAK – 12.7 %) for ‘187’ allele. The most common alleles for the BM2113 STR locus in the YAK group were ‘128’ (23.6 %) and ‘130’ (27.3 %), while total frequency of these alleles in the COW group was only 17.2 %.

A similar trend was observed for the CSSM66 locus, and there was a significant difference in the frequency of ‘172’, ‘178’, ‘180’, ‘184’ and ‘190’ alleles. ‘294’ (42.7 %) was the most common allele in the ILSTS006 locus for the YAK group,

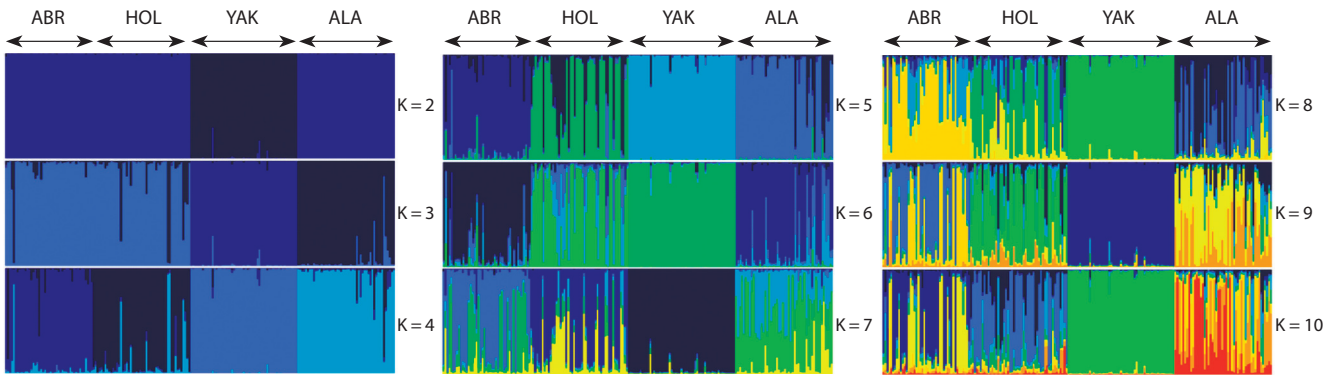


Fig. 1. Genetic structure of the studied samples for the most probable cluster count (K) from 2 to 10.
ABR – Aberdeen-Angus, HOL – Holstein, YAK – domestic yaks, ALA – Alatau.

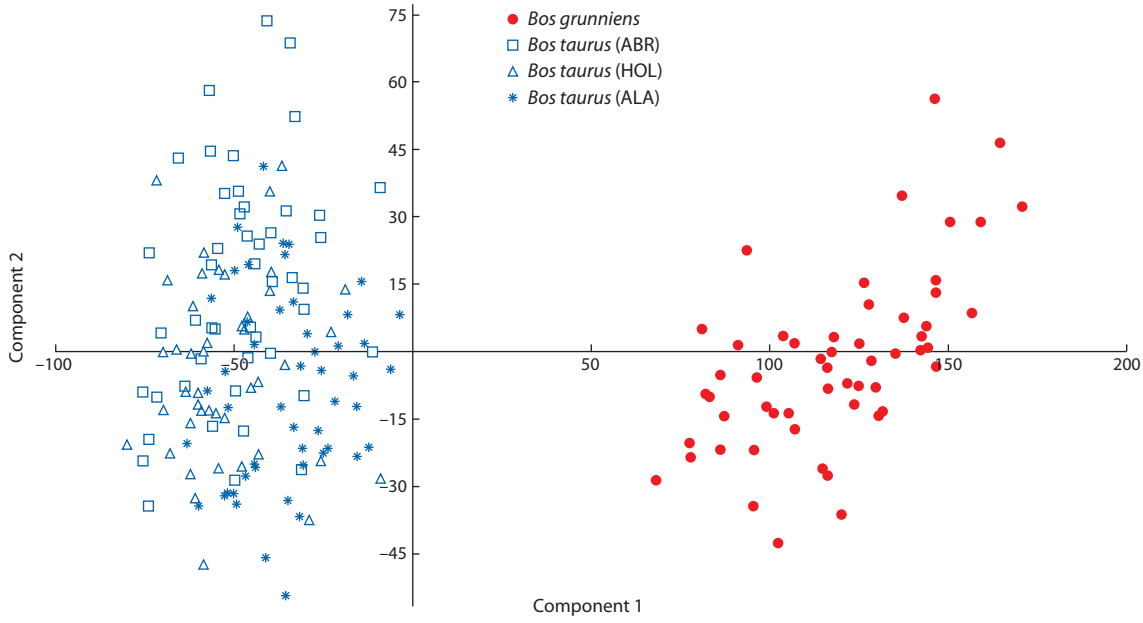


Fig. 2. Principal component analysis (all 15 STR loci).

Table 2. Differentiating potential of 15 STR loci (locus-by-locus AMOVA analysis for COW and YAK groups)

STR-locus	F_{ST}	*p*-value	STR-locus	F_{ST}	*p*-value	STR-locus	F_{ST}	*p*-value
BM1818	**0.056**	<0.001	ETH10	0.017	<0.001	SPS115	0.024	<0.001
BM1824	**0.041**	<0.001	ETH225	0.013	<0.001	TGLA122	0.018	<0.001
BM2113	**0.030**	<0.001	ETH3	0.028	<0.001	TGLA126	0.017	<0.001
CSRM60	0.018	<0.001	ILSTS006	**0.063**	<0.001	TGLA227	0.022	<0.001
CSSM66	**0.034**	<0.001	INRA023	0.023	<0.001	TGLA53	0.029	<0.001

Note. Values in bold are greater than 0.03.

whereas '286' (11.4 %), '290' (12.8 %), '272' (19.3 %) and '292' (23.8 %) were the most prevalent in COW. Table 4 summarizes private alleles for these STR loci.

Based on the data obtained, a repeated analysis of the subpopulation structure was completed using Structure v.2.3.4 only for 5 out of 15 STR loci, as a result of genotyping analysis (Table 5). Table 3 presents allelic diversity and allele prevalence for these STR loci.

As a result of the simulation with Structure v.2.3.4 (the duration of the burn-in 5000, the number of MCMC repetitions

Table 3. The allele frequency of five STR loci with the highest differentiating potential according to F_{ST} values for *B. grunniens* and *B. taurus*

Allele	*B. taurus*	*B. grunniens*	Allele	*B. taurus*	*B. grunniens*
BM1818			CSSM66		
250	0.010	–	172	0.134	0.045
254	0.007	0.036	176	0.028	0.118
256	0.303	0.191	178	0.010	0.145
258	0.297	0.091	180	0.169	0.036
260	0.066	0.091	182	0.148	0.227
262	0.241	0.564	184	0.148	0.027
264	0.021	–	186	0.214	0.055
266	0.048	–	190	0.062	0.200
272	0.003	–	192	0.031	–
274	0.003	–	194	0.045	0.091
276	–	0.027	196	0.010	0.045
BM1824			198	–	0.009
185	0.203	0.309	ILSTS006		
187	0.352	0.127	270	0.069	0.018
189	0.155	0.136	272	0.193	0.064
193	0.010	–	274	0.028	0.009
195	0.183	0.418	278	0.055	–
197	0.097	0.009	282	0.010	0.036
BM2113			284	0.003	0.018
124	0.059	–	286	0.114	0.218
128	0.066	0.236	288	0.052	–
130	0.107	0.273	290	0.128	0.182
132	0.003	–	292	0.238	0.009
134	0.166	0.082	294	0.083	0.427
136	0.128	0.064	296	0.028	–
138	0.045	0.127	298	–	0.009
140	0.124	0.027	300	–	0.009
142	0.100	0.009			
144	0.100	0.173			
146	0.034	0.009			
152	0.069	–			

after the burn-in 50,000, 10 iterations), we found four distinct clusters (K = 4, ΔK = 119.7). Structure v.2.3.4, according to the method of J.K. Pritchard (Pritchard et al., 2000), allowed to compute the Q criterion, which characterized the assignment of each individual to a group (species) for four samples, including ABR, HOL, ALA and YAK. Q ≥ 75 % in the ABR sample was found in 88.9 % (40/45) of the individuals, accuracy 92.6 ± 5.8 %; HOL, in 68.0 % (34/50), accuracy 92.4 ± 6.2 %; ALA, in 82.0 % (41/50), accuracy 93.2 ± 5.6 %; and YAK, in 96.4 % (53/55), accuracy 97.7 ± 3.4 %. To improve the accuracy of individuals differentiation in the Holstein breed, the list of analyzed STR loci should be further expanded, starting with ETH3, TGLA126 and TGLA122.

In total, differentiation accuracy based on BM1818, BM1824, BM2113, CSSM66 and ILSTS006 STR loci in YAK (*B. grunniens*) was 98.8 ± 3.4 %, and 99.1 ± 1.2 % in COW (*B. taurus*). Thus, differentiation accuracy was not lost even when 5 STR loci out of 15 were analyzed.

Earlier, Inter Simple Sequence Repeats of yak-cattle hybrids were studied at the Institute of General Genetics RAS, and a species-specific pattern of eight ISSR fragments for yak was found in yak and F_1 hybrids populations (Stolpovsky et al., 2014). Also, the allele depository of yaks and their hybrids with *B. taurus* was assessed earlier using microsatellite analysis, yielding high genetic diversity for F_1 hybrids in comparison with the original species (Al-Kaisy, 2011). Our study did not confirm hybrid individuals of *B. grunniens* and *B. taurus.*

Table 4. Private allele frequency for COW and YAK

Group	STR locus	Allele	Frequency
COW	BM1818	266	0.048
YAK		276	0.027
COW		264	0.021
COW		250	0.010
COW		272	0.003
COW		274	0.003
COW	BM1824	193	0.010
COW	BM2113	152	0.069
COW		124	0.059
COW		132	0.003
COW	CSSM66	192	0.031
YAK		198	0.009
COW	ILSTS006	278	0.055
COW		288	0.052
COW		296	0.028
YAK		298	0.009
YAK		300	0.009

Table 5. Differentiating potential of five STR loci (locus-by-locus AMOVA analysis)

STR locus	YAK/COW		YAK/ABR/HOL/ALA	
	F_{ST}	*p*-value	F_{ST}	*p*-value
BM1818	0.056	<0.001	0.207	<0.001
BM1824	0.041	<0.001	0.158	<0.001
BM2113	0.030	<0.001	0.130	<0.001
CSSM66	0.034	<0.001	0.113	<0.001
ILSTS006	0.063	<0.001	0.188	<0.001

Conclusion

This study assessed the differentiating potential of 15 STR loci, including ETH3, INRA023, TGLA227, TGLA126, TGLA122, SPS115, ETH225, TGLA53, BM2113, BM1824, ETH10, BM1818, CSSM66, ILSTS006 and CSRM60 for *B. grunniens* and *B. taurus* individuals, as well as attempted to identify hybrids of these species.

According to the subpopulation structure analysis, following genotyping of 15 STR loci, the classification accuracy of *B. grunniens* individuals was 99.1 ± 1.2 %, and 99.6 ± 0.4 % for *B. taurus*. When the number of STI loci used for decision was limited to five, including BM1818, BM1824, BM2113, CSSM66 and ILSTS006, the differentiating potential of which, according to F_{ST}, was the greatest and varied from 0.030 to 0.063, the classification accuracy for *B. grunniens* was 98.8 ± 3.4 %, and 99.1 ± 1.2 % for *B. taurus*.

Thus, we conclude that the analysis of even a small number of STR loci allows to ascertain differentiation of domestic yak and three breeds of cattle (Aberdeen-Angus, Holstein and Alatau) bred in Kyrgyzstan. At the same time, further research is needed in the longer run to more accurately classify differentiation potential for selected loci.

References

Abdykerimov A. Theory and Practice of Raising Yaks in Kyrgyzstan. Bishkek, 2001. (in Russian)

Al-Kaisy T.V. Comparative study of the allele pool of yaks and their hybrids with cattle using microsatellites. Cand. Biol. Sci. Diss. Dubrovitsy, 2011. (in Russian)

Barendse W., Armitage S.M., Kossarek L.M., Shalom A., Kirkpatrick B.W., Ryan A.M., Clayton D., Li L., Neibergs H.L., Zhang N., Grosse W.M., Weiss J., Creighton P., McCarthy F., Ron M., Teale A.J., Fries R., McGraw R.A., Moore S.S., Georges M., Soller M., Womack J.E., Hetzel D.J.S. A genetic linkage map of the bovine genome. *Nat. Genet*. 1994;6(3):227-235. DOI 10.1038/ng0394-227.

Bishop M.D., Kappes S.M., Keele J.W., Stone R.T., Sunden S.L.F., Hawkins G.A., Toldo S.S., Fries R., Grosz M.D., Yoo J., Beattie C.W. A genetic linkage map for cattle. *Genetics*. 1994;136(2):619-639. DOI 10.1093/genetics/136.2.619.

Bovine Genome Project. Baylor College of Medicine Human Genome Sequencing Center. Genome Bovine Whole Genome Assembly release Btau_3.1. URL: http://www.hgsc.bcm.tmc.edu/projects/bovine/ (Application 25.07.2022).

Brezinsky L., Kemp S.J., Teale A.J. ILSTS006: a polymorphic bovine microsatellite. *Anim. Genet*. 1993;24(1):73. DOI 10.1111/j.1365-2052.1993.tb00933.x.

Chertkiev Sh.Ch., Chortonbaev T.J. Scientific Basis for the Formation of Meat Productivity of Yaks in Ontogeny. Bishkek, 2007. (in Russian)

Francis R.M. POPHELPER: an R package and web app to analyse and visualise population structure. *Mol. Ecol. Resour*. 2016;17(1):27-32. DOI 10.1111/1755-0998.12509.

Georges M., Massey J. Polymorphic DNA markers in Bovidae. Patent application WO PUBL NO 92/13102. World Intellectual Property Organization. Geneva, 1992.

Hammer Ø., Harper D.A.T., Ryan P.D. Past: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron*. 2001;4(1):4.

Jacques G., d'Alpoim Guedes J., Zhang S. Yak domestication: a review of linguistic, archaeological, and genetic evidence. *Ethnobiol. Lett*. 2021;12(1):103-114. DOI 10.14237/ebl.12.1.2021.1755.

Luz Ya.Ya. (Ed.) Domestic Animals of Mongolia. Proceedings of the livestock detachment of the Mongolian expedition of the Academy of Sciences of the USSR in 1931. Moscow–Leningrad: Publishing House of the USSR Academy of Sciences, 1936. (in Russian)

Nosova A.Yu., Kipen V.N., Tsar A.I., Lemesh V.A. Differentiation of hybrid progeny of silver carp (*Hypophthalmichthys molitrix* Val.) and bighead carp (*H. nobilis* Rich.) based on microsatellite polymorphism. *Russ. J. Genet*. 2020;56(3):317-323. DOI 10.1134/S1022795420030126.

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. DOI 10.1093/genetics/155.2.945.

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.

Sambrook J., Russell D.W. Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press, 2001.

Steffen P., Eggen A., Dietz A.B., Womack J.E., Stranzinger G., Fries R. Isolation and mapping of polymorphic microsatellites in cattle. *Anim. Genet*. 1993;24(2):121-124. DOI 10.1111/j.1365-2052.1993.tb00252.x.

Stolpovsky Yu.A., Kol N.V., Evsyukov A.N., Nesteruk L.V., Dorzhu Ch.M., Tsendsuren Ts., Sulimova G.E. Comparative analysis of ISSR marker polymorphism in population of yak (*Bos mutus*) and in F_1 hybrids between yak and cattle in the Sayan-Altai region. *Russ. J. Genet.* 2014;50(10):1163-1176. DOI 10.1134/S1022795414100135.

Toldo S., Fries R., Steffen P., Neibergs H.L., Barendse W., Womack J.E., Hetzel D.J., Stranzinger G. Physically mapped, cosmid-derived microsatellite markers as anchor loci on bovine chromosomes. *Mamm. Genome*. 1993;4(12):720-727. DOI 10.1007/BF00357796.

Vaiman D., Mercier D., Moazami-Goudarzi K., Eggen A., Ciampolini R., Lepingle A., Velmala R., Kaukinen J., Varvio S.L., Martin P., Leveziel H., Guerin G. A set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism. *Mamm. Genome*. 1994;5(5):288-297. DOI 10.1007/BF00389543.

ORCID ID

K.B. Chekirov orcid.org/0000-0001-6146-6750
Zh.T. Isakova orcid.org/0000-0002-3681-6939
V.N. Kipen orcid.org/0000-0002-7822-0746
M.I. Irsaliev orcid.org/0000-0002-8364-8982
S.B. Mukeeva orcid.org/0000-0001-9584-4860
K.A. Aitbaev orcid.org/0000-0003-4973-039X
G.A. Sharshenalieva orcid.org/0000-0002-5016-2492
S.B. Beyshenalieva orcid.org/0000-0003-4461-6993
B.U. Kydyralieva orcid.org/0000-0001-9392-2773

Acknowledgements. This research was completed within the frame of the governmental assignment of the Ministry of Education and Science of Kyrgyz Republic (Contract No. 30-21 dated 15/02/2021 No. 129/1).

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

Received August 19, 2022. Revised February 6, 2023. Accepted March 24, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):495-501
DOI 10.18699/VJGB-23-60

Original Russian text https://vavilovj-icg.ru/

Quantifying human genome parameters in aging

V.P. Volobaev[1] ✉, S.S. Kunizheva[1, 2, 3], L.I. Uralsky[1, 2], D.A. Kupriyanova[1], E.I. Rogaev[1, 2, 3, 4]

[1] Sirius University of Science and Technology, Scientific Center for Genetics and Life Sciences, Sochi, Russia
[2] Vavilov Institute of General Genetics, Russian Academy of Sciences, Department of Genomics and Human Genetics, Moscow, Russia
[3] Lomonosov Moscow State University, Center for Genetics and Genetic Technologies, Faculty of Biology, Moscow, Russia
[4] University of Massachusetts Chan Medical School, Department of Psychiatry, Shrewsbury, USA
✉ volobaev.vp@gmail.com

Abstract. Healthy human longevity is a global goal of the world health system. Determining the causes and processes influencing human longevity is the primary fundamental goal facing the scientific community. Currently, the main efforts of the scientific community are aimed at identifying the qualitative characteristics of the genome that determine the trait. At the same time, when evaluating qualitative characteristics, there are many challenges that make it difficult to establish associations. Quantitative traits are burdened with such problems to a lesser extent, but they are largely overlooked in current genomic studies of aging and longevity. Although there is a wide repertoire of quantitative trait analyses based on genomic data, most opportunities are ignored by authors, which, along with the inaccessibility of published data, leads to the loss of this important information. This review focuses on describing quantitative traits important for understanding aging and necessary for analysis in further genomic studies, and recommends the inclusion of the described traits in the analysis. The review considers the relationship between quantitative characteristics of the mitochondrial genome and aging, longevity, and age-related neurodegenerative diseases, such as the frequency of extensive mitochondrial DNA (mtDNA) deletions, mtDNA half-life, the frequency of A>G replacements in the mtDNA heavy chain, the number of mtDNA copies; special attention is paid to the mtDNA methylation sign. A separate section of this review is devoted to the correlation of telomere length parameters with age, as well as the association of telomere length with the amount of mitochondrial DNA. In addition, we consider such a quantitative feature as the rate of accumulation of somatic mutations with aging in relation to the lifespan of living organisms. In general, it may be noted that there are quite serious reasons to suppose that various quantitative characteristics of the genome may be directly or indirectly associated with certain aspects of aging and longevity. At the same time, the available data are clearly insufficient for definitive conclusions and the determination of causal relationships.
Key words: genome quantification; aging; longevity; neurodegenerative disorders; mtDNA; telomere length; somatic mutations.

For citation: Volobaev V.P., Kunizheva S.S., Uralsky L.I., Kupriyanova D.A., Rogaev E.I. Quantifying human genome parameters in aging. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):495-501. DOI 10.18699/VJGB-23-60

Количественные параметры генома человека при старении

В.П. Волобаев[1] ✉, С.С. Кунижева[1, 2, 3], Л.И. Уральский[1, 2], Д.А. Куприянова[1], Е.И. Рогаев[1, 2, 3, 4]

[1] Научно-технологический университет «Сириус», Научный центр генетики и наук о жизни, Сочи, Россия
[2] Институт общей генетики им. Н.И. Вавилова Российской академии наук, отдел геномики и генетики человека, Москва, Россия
[3] Московский государственный университет им. М.В. Ломоносова, Центр генетики и генетических технологий, биологический факультет, Москва, Россия
[4] Медицинская школа Чан Массачусетского университета, отделение психиатрии, Шрусбери, США
✉ volobaev.vp@gmail.com

Аннотация. Здоровое долголетие человека – глобальная цель мировой системы здравоохранения. В то же время неуклонное старение населения стало серьезным вызовом для систем здравоохранения многих стран мира, в том числе из-за возросшего риска развития многих тяжелых нейродегенеративных заболеваний, включая болезнь Альцгеймера (БА) и болезнь Паркинсона (БП). Определение причин и процессов, влияющих на старение и продолжительность жизни человека, а также выявление механизмов развития возрастных патологий – первостепенная фундаментальная задача, стоящая перед научным сообществом. В настоящее время основные усилия направлены на идентификацию качественных характеристик генома, детерминирующих признак. Вместе с тем при их оценке существует множество проблем, затрудняющих установление ассоциаций. Количественные признаки обременены таковыми проблемами в меньшем объеме, но в большинстве случаев упускаются при проведении современных геномных исследований, посвященных вопросам старения и долголетия. Несмотря на наличие широкого круга возможностей проведения анализа геномных данных по количественным признакам, большинство возможностей не используется, что наряду с недоступностью опубликованных данных ведет к по-

тере этой важной информации. Настоящий обзор посвящен описанию количественных признаков, важных для понимания процесса старения и необходимых для анализа в дальнейших геномных исследованиях, и является рекомендацией для включения описанных признаков в анализ. Рассматривается взаимосвязь количественных характеристик ядерного и митохондриального генома со старением, долголетием и возрастными нейродегенеративными заболеваниями, таких как частота обширных делеций митохондриальной ДНК (mtDNA), время полураспада mtDNA, частота замен A>G в тяжелой цепи mtDNA, количество копий mtDNA, длина теломер, частота соматических мутаций. В целом можно отметить, что есть достаточно серьезные причины полагать, что различные количественные характеристики генома могут быть прямо или косвенно ассоциированы с теми или иными аспектами старения и продолжительности жизни. Но имеющихся данных недостаточно для окончательных выводов и выявления причинно-следственных связей.
Ключевые слова: количественные параметры генома; старение; долголетие; нейродегенеративные заболевания; mtDNA; длина теломер; соматические мутации.

Introduction

Human longevity is a complex trait that is influenced by environmental factors, lifestyle, random events, and individual genetic traits. Studies have shown that genetics plays a significant role in longevity, with individuals from families of long-livers having a higher chance of living longer (van den Berg, 2020). However, identifying specific genetic determinants associated with longevity has been challenging. Currently, only two genes, *APOE* and *FOXO3A*, have been shown to be important for human longevity across different samples and research groups (Deelen et al., 2019). Other results have been inconsistent, possibly due to population differences and the effect of multiple comparisons.

Despite the difficulty in identifying specific genetic determinants, maintaining body health is crucial for longevity. Aging of the brain and the development of cerebrovascular and neurodegenerative diseases are major causes of disability and death in older adults (Debette et al., 2019). However, the genetic basis of age-related brain diseases is complex and inconsistent. In contrast to qualitative characteristics of the genome, quantitative traits such as telomere length, the number of mitochondrial DNA copies, the frequency of heterozygous variants of mitochondrial DNA, and the frequency of somatic mutations are less affected by population and statistical factors. Despite their importance, little attention has been paid to these quantitative traits in the study of the genetic basis of longevity.

This review aims to analyze existing information on quantitative genetic traits affecting aging and human longevity.

Quantification of changes in mtDNA structure due to aging

In recent years, there has been much interest in the role of mtDNA as a determinant of aging, lifespan processes, and age-related diseases. Mitochondrial dysfunction is considered one of the key aging biomarkers (Miva et al., 2022), and changes in quantitative and qualitative characteristics of mtDNA are directly associated with longevity. Given the spatial proximity of mtDNA to the electron transport chain, it is exposed to the damaging effects of free radicals, which, along with a limited ability to repair, due to the fact that mtDNA is not protected by histones and is in a single-stranded form for a considerable part of its replication time, determines the vulnerability of mtDNA structure to damage and degradation. All of these factors lead to a higher rate of chemical modifications and mutation accumulation in mtDNA compared to the cell nucleus DNA.

Damage and deletion of mtDNA sites can lead to mitochondrial dysfunction due to an increased proportion of molecules containing an extensive deletion (mtDNAdel), since mtDNA with an extensive deletion has a replicative advantage over wild-type mtDNA (Kowald, Kirkwood, 2018). The replicative advantage is probably determined by the smaller size of the replicating molecule, which leads to a higher replication rate (Diaz, 2002), and at the same time, a lower chance of damage to the molecule by active oxygen species. This results in less active mitophagy of mtDNAdel-rich organelles compared to normal organelles (de Grey, 1997). Moreover, whereas in actively proliferating tissues, cells containing dysfunctional mitochondria are subject to elimination and replacement, tissues characterized by a high number of postmitotic cells accumulate a burden of such mutations, which probably leads to a decrease in the functional parameters of the tissue (Herbst et al., 2017).

The proportion of mtDNAdel in muscle tissue has been shown to increase approximately 19-fold, from 0.008 to 0.15 %, from 50 to 86 years (Herbst et al., 2021b). A similar phenomenon has been observed in nerve tissue (Nido et al., 2018). It has also been noted that significant accumulation of mtDNAdel is observed in patients with Parkinson's disease in substantia nigra neurons (Bender et al., 2006; Grünewald et al., 2016) and the striatum (Ikebe et al., 1990). Moreover, there is an opinion that accumulation of mtDNAdel can trigger neuroprotective mechanisms (Perier et al., 2013).

The state of mtDNA heterogeneity in which several clones of mtDNA with different nucleotide sequences exist in mitochondria is called heteroplasmy. It is known that heteroplasmy can occur either *de novo* during ontogenesis or by maternal inheritance (Sallevelt et al., 2017). Heteroplasmic mutations also appear to be associated with macroinflammation (Just et al., 2015). For example, R. Zhang and colleagues noted that, on average, individuals over the age of 70 had 58.5 % more mtDNA heteroplasmic mutations than individuals under the age of 40 (Zhang et al., 2017). This fact becomes of great significance when we consider that there is substantial evidence linking mtDNA heteroplasmy with neurodegenerative diseases directly associated with longevity: Alzheimer's disease (AD) (Tranah et al., 2012) and Parkinson's disease (PD) (Hudson et al., 2013). At the same time, there are reports showing a positive role of heteroplasmy for longevity (Rose et al., 2010; Sondheimer et al., 2011), which is probably because mtDNA heteroplasmy is a reservoir of genetic variability that can introduce new functions and increase the ability of cells to

cope with environmental and physiological stressors during life. It can be assumed that both of these phenomena take place and their importance for longevity is determined by the localization of somatic mtDNA mutation accumulation and by the fact that congenital heteroplasmy can have a positive effect to a greater extent, while acquired one has a greater chance to carry negative properties.

Another mitochondrial marker likely associated with longevity may be the frequency of accumulation of mitochondrial somatic mtDNA (mtSNV) A>G mutations in the mtDNA heavy chain. In a recent study (Mikhailova et al., 2022), the authors determined a positive correlation between the frequency of $A_H>G_H$ (H – heavy chain) substitutions and the lifespan of different mammalian species: the more long-lived a species is, the higher the frequency of $A_H>G_H$ substitutions is observed in it. At the same time, the authors suggest that the observed accumulation of G_H nucleotides is a consequence of oxidative mutagenesis and aging processes rather than a cause.

The half-life of mtDNA also seems to be an important factor determining the rate of tissue dysfunction onset. It has been suggested that cell lifespan depends on mtDNA half-life (Poovathingal et al., 2012; Chan et al., 2013). In modeling the effect of half-life on cell survival time, it has been determined that a moderate increase in mtDNA half-life has a profound effect on increasing cell survival time, thereby reducing the replicative advantage of mtDNA with extensive deletions (Holt, Davies, 2021). Equally importantly, a decrease in mtDNA half-life significantly affects the process of mtSNV accumulation in tissues characterized by a high number of postmitotic cells. It has been shown that if the half-life is three months, pathogenic mtSNV acquired in a neuronal progenitor cell early in development and present in the postmitotic neuronal population at an average frequency of 1 %, by 70 years of human life, will be contained in most neurons with a frequency of ~14 % (Li et al., 2019). Accordingly, changing the half-life rate downward acts to inhibit mitochondrial heteroplasmy levels and vice versa.

In addition to mutational events, the mtDNA copy number (mtDNAcn) is an important quantitative trait. Changes in mtDNAcn are usually a reflection of the mitochondrial response to oxidative stress and are also associated with general dysfunction. Various studies have reported results showing a decrease in mtDNAcn as humans age (Herbst et al., 2017, 2021a). A decrease in mtDNA copies in whole blood has been found to occur with age, and a lower number of mtDNA copies is associated with poorer health (Lee et al., 2010; Mengel-From et al., 2014). High mtDNAcn levels are probably generally associated with better health outcomes in older individuals, including higher levels of cognitive function and lower mortality (Kim et al., 2013; Mengel-From et al., 2014). It has been noted that a decreased mtDNAcn score is strongly associated with the risk of age-related neurodegenerative diseases such as dementia, PD, AD, etc. (Yang et al., 2021).

It should be noted that both systemic trends toward a decrease in mtDNAcn in individuals with AD and a local decrease in mtDNAcn by 30–50 % in the frontal lobe of the large hemisphere cortex and hippocampus compared to healthy controls have been observed (Coskun et al., 2004; Rice et al., 2014). At the same time, there is a publication that describes an increase in mtDNAcn in patients of African descent with Parkinson's disease (Müller-Nedebock et al., 2022).

In studies examining changes in mtDNAcn in the blood leukocytes of long-livers as a model of healthy aging, contradictory results have been obtained. Y.H. He et al. (2014) showed a significant increase in the amount of mtDNAcn in centenarians compared to elderly people (He et al., 2014), but van Leeuwen et al. did not observe such a pattern (van Leeuwen et al., 2014), which may be due to different methodological approaches. It should be noted that different tissues may show different age dynamics of mtDNAcn. For example, while an inverse correlation was observed in skeletal muscle samples, a positive correlation was observed in liver or substantia nigra samples (Dölle et al., 2016; Wachsmuth et al., 2016).

The mtDNAcn index seems to be related to the telomere length (TL) parameter (Qiu et al., 2015; Tyrka et al., 2015; Dolcini et al., 2020). It is assumed that this relationship is based on the negative correlation between mtDNAcn levels and levels of reactive oxygen species (ROS) and further negative effects of ROS on telomere length (Melicher et al., 2018).

Telomere length as a cause or consequence of longevity

Telomere length is a well-known biomarker of aging (Sanders, Newman, 2013). Although the relationship between TL and cellular aging is undeniable in model cell cultures (Victorelli, Passos, 2017), the conclusions for multicellular organisms are not so unambiguous (Blackburn et al., 2015). It has been suggested that telomere shortening dynamics, rather than total telomere length, can serve as a quantitative biomarker of macroorganism lifespan (Vera et al., 2012). For example, in cross-sectional studies on five bird species, it was shown that short-lived bird species lose more telomere repeats with age than species with longer lifespans (Haussmann et al., 2003). A similar correlation has been observed in mammals, suggesting that long-lived animals have more effective mechanisms of protection against replicative aging, such as higher telomerase activity throughout life (Haussmann et al., 2007).

In humans, shorter telomere length is associated with higher mortality rates from various age-related pathologies, including some neurodegenerative diseases such as dementia (Levstek et al., 2021). However, reports on the role of telomere length in the risk of AD are ambiguous. Some studies noted that TL length is lower in people with AD than in control samples (Thomas et al., 2008; Forero et al., 2016), while P. Thomas et al. noted an inverse relationship in some tissues such as the hippocampus. Interestingly, longer telomeres have a negative effect on disease dynamics and severity (Movérare-Skrtic et al., 2012; Mahoney et al., 2019). Short TL is a good prognostic marker for determining the long-term risk of AD in APOE4-negative individuals (Hackenhaar et al., 2021). Moreover, TL is associated with cognitive function in both elderly and middle-aged individuals (Hägg et al., 2017; Gampawar, 2022).

It has been estimated that leukocyte telomeres in adults shorten at an average rate of 24.7 bp per year (Müezzinler et al., 2013). A number of different factors can influence TL

and the rate of telomere depletion. For example, TL has been shown to be higher in older women compared to men (Benetos et al., 2001) and in African Americans compared to Caucasians (Hunt et al., 2008). First of all, it should be noted that in addition to the large number of studies that have noted a negative correlation of TL with age and the association of this parameter with mortality in the older age group, there are also studies in which these patterns were not confirmed (Sanders, Newman, 2013).

Initially, it was assumed that such discrepancies are associated with the peculiarities of specific studies, such as the methodology of sample formation, the presence of population stratification, the type of tissue studied, and the methods of studying the index. For example, in an extensive study of TL in various tissues, it was determined that in 21 types of tissue, there is a negative correlation of TL with age (the strongest correlations for whole blood and gastric tissue), while no correlation was observed for testes, ovaries, cerebellum, vagina, skeletal muscle, thyroid gland, and gastroesophageal junction tissue (Demanelis et al., 2020).

When studying long-livers as a model of healthy aging, it was hypothesized that TL primarily depends on the physiological state of the organism rather than age. It was shown that in "high-performing" long-livers (with a low number of diseases and high physical activity), TL was significantly higher than TL in "low-performing" long-livers (with a high number of diseases and low physical activity). Therefore, it has been suggested that it is probably not the telomere length factor that affects the ability to live to one hundred years, but the health condition associated with telomere length (Terry et al., 2008; Tedone et al., 2019). This theory is also supported by a study of TL in same-sex twins over the age of 70, which noted a clear association between blood white cell TL and physical health, including between twins (Bendix et al., 2011). Thus, the study of telomere dynamics in long-lived individuals is of particular importance because they may have developed mechanisms that actively postpone aging and provide effective protection against the negative effects of aging processes.

Somatic mutations and their role in longevity

The current theory of aging suggests that the accumulation of DNA mutations in somatic cells (copy number variations, CNVs) with age leads to a decrease in cell function due to the inactivation or disruption of important genes (Kennedy et al., 2012). Indeed, it has been shown that the accumulation of somatic mutations occurs with age and at a differential rate for different tissues. For example, in human proximal bronchial basal cells, the rate of mutation accumulation is approximately 29 CNVs per cell per year (CNVs/pcpy) (Huang et al., 2022); in prefrontal cortex and hippocampal neurons, it is 16–21 CNVs/pcpy (Lodato et al., 2018; Miller et al., 2022); in subcutaneous preadipocytes, it is 18 CNVs/pcpy; in visceral adipose tissue preadipocytes, it is 27 CNVs/pcpy (Franco et al., 2019); in memory T cells, it is approximately 25 CNVs/pcpy; in naive B-lymphocytes, it is approximately 15 CNVs/pcpy; in hematopoietic stem cells and progenitor cells, it is approximately 16 CNVs/pcpy (Machado et al., 2022); and in spermatogonia, it is approximately 2 CNVs/pcpy (Milholland et al., 2017).

A vivid illustration of the significance of CNVs for lifespan is provided by studies of the rate of mutation accumulation in the crypts of the large intestine in mammals with different lifespans (Cagan et al., 2022). While the rate for humans is approximately 47 CNVs/pcpy, for giraffes at 25–35 years of life, it is approximately 99 CNVs/pcpy; for ferrets at 14 years of life, it is approximately 496 CNVs/pcpy; and for mice at 2 years of life, it is approximately 796 CNVs/pcpy. Thus, the dependence of lifespan and the rate of mutation accumulation is well established.

At the same time, it has been shown that the frequency of somatic mutations in humans in old age is much lower than that required for the loss of gene function in a significant number of cells, indicating an indirect relationship between the indices (Vijg, Dong, 2020). In a study of a large sample of Chinese centenarians compared to controls, it was observed that CNV levels were significantly higher in the sample of centenarians than in the control sample, indicating that the frequency of CNVs does not directly affect the probability of living beyond the population norm (Zhao et al., 2018). On the other hand, a study of centenarians from Italy obtained different data, observing that centenarians had significantly lower levels of CNVs than controls (Garagnani et al., 2021). Given the contradictory results obtained in these two studies, more research on this issue is needed.

Conclusion

Thus, there are reasons to suggest that there is a significant association between aging dynamics, life expectancy, healthy aging, the risk of neurodegenerative diseases, and various quantitative genomic characteristics. At the same time, what is the cause and what is the effect in most cases is not determined, which, along with the sporadic nature of the available publications, highlights the need for additional research.

References

Bender A., Krishnan K.J., Morris C.M., Taylor G.A., Reeve A.K., Perry R.H., Jaros E., Hersheson J.S., Betts J., Klopstock T., Taylor R.W., Turnbull D.M. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat. Genet*. 2006;38(5):515-517. DOI 10.1038/ng1769.

Bendix L., Gade M.M., Staun P.W., Kimura M., Jeune B., Hjelmborg J.V.B., Aviv A., Christensen K. Leukocyte telomere length and physical ability among Danish Twins age 70+. *Mech. Ageing Dev*. 2011;132(11-12):568-572. DOI 10.1016/j.mad.2011.10.003.

Benetos A., Okuda K., Lajemi M., Kimura M., Thomas F., Skurnick J., Labat C., Bean K., Aviv A. Telomere length as an indicator of biological aging: the gender effect and relation with pulse pressure and pulse wave velocity. *Hypertension*. 2001;37(2):381-385. DOI 10.1161/01.hyp.37.2.381.

Blackburn E.H., Epel E.S., Lin J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science*. 2015;350(6265):1193-1198. DOI 10.1126/science.aab3389.

Cagan A., Baez-Ortega A., Brzozowska N., Abascal F., Coorens T.H.H., Sanders M.A., Lawson A.R.J., … Bochynska D., Smith E.St.J., Gerstung M., Campbell P.J., Murchison E.P., Stratton M.R., Martincorena I. Somatic mutation rates scale with lifespan across mammals. *Nature*. 2022;604(7906):517-524. DOI 10.1038/s41586-022-04618-z.

Chan S.W., Chevalier S., Aprikian A., Chen J.Z. Simultaneous quantification of mitochondrial DNA damage and copy number in circulat-

ing blood: a sensitive approach to systemic oxidative stress. *BioMed Res. Int*. 2013;2013:157547. DOI 10.1155/2013/157547.

Coskun P.E., Beal M.F., Wallace D.C. Alzheimer's brains harbor somatic mtDNA control-region mutations that suppress mitochondrial transcription and replication. *Proc. Natl. Acad. Sci. USA*. 2004; 101(29):10726-10731. DOI 10.1073/pnas.0403649101.

Debette S., Schilling S., Duperron M., Larsson S.C., Markus H.S. Clinical significance of magnetic resonance imaging markers of vascular brain injury: A systematic review and meta-analysis. *JAMA Neurol*. 2019;76(1):81-94. DOI 10.1001/jamaneurol.2018.3122.

Deelen J., Evans D.S., Arking D.E., Tesi N., Nygaard M., Liu X., Wojczynski M.K., … Zeng Y., Zheng W., Holstege H., Kiel D.P., Lunetta K.L., Slagboom P.E., Murabito J.M. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat. Commun*. 2019;10(1):3669. DOI 10.1038/s41467-019-11558-2.

de Grey A.D.N.J. A proposed refinement of the mitochondrial free radical theory of aging. *BioEssays*. 1997;19(2):161-166. DOI 10.1002/bies.950190211.

Demanelis K., Jasmine F., Chen L.S., Chernoff M., Tong L., Delgado D., Zhang C., Shinkle J., Sabarinathan M., Lin H., Ramirez E., Oliva M., Kim-Hellmuth S., Stranger B.E., Lai T.-P., Aviv A., Ardlie K.G., Aguet F., Ahsan H., Doherty J.A., Kibriya M.G., Pierce B.L. Determinants of telomere length across human tissues. *Science*. 2020;369(6509):eaaz6876. DOI 10.1126/science.aaz6876.

Diaz F. Human mitochondrial DNA with large deletions repopulates organelles faster than full-length genomes under relaxed copy number control. *Nucleic Acids Res*. 2002;30(21):4626-4633. DOI 10.1093/nar/gkf602.

Dolcini J., Wu H., Nwanaji-Enwerem J.C., Kiomourtozlogu M.A., Cayir A., Sanchez-Guerra M., Vokonas P., Schwarz J., Baccarelli A.A. Mitochondria and aging in older individuals: an analysis of DNA methylation age metrics, leukocyte telomere length, and mitochondrial DNA copy number in the VA normative aging study. *Aging*. 2020;12(3):2070-2083. DOI 10.18632/aging.102722.

Dölle C., Flønes I., Nido G.S., Miletic H., Osuagwu N., Kristoffersen S., Lilleng P.K., Larsen J.P., Tysnes O.B., Haugarvoll K., Bindoff L.A., Tzoulis C. Defective mitochondrial DNA homeostasis in the substantia nigra in Parkinson disease. *Nat. Commun*. 2016;22(7):13548. DOI 10.1038/ncomms13548.

Forero D.A., González-Giraldo Y., López-Quintero C., Castro-Vega L.J., Barreto G.E., Perry G. Meta-analysis of telomere length in Alzheimer's disease. *J. Gerontol. A Biol. Sci. Med. Sci*. 2016;71(8): 1069-1073. DOI 10.1093/gerona/glw053.

Franco I., Helgadottir H.T., Moggio A., Larsson M., Vrtačnik P., Johansson A., Norgren N., Lundin P., Mas-Ponte D., Nordström J., Lundgren T., Stenvinkel P., Wennberg L., Supek F., Eriksson M. Whole genome DNA sequencing provides an atlas of somatic mutagenesis in healthy human cells and identifies a tumor-prone cell type. *Genome Biol*. 2019;20(1):285. DOI 10.1186/s13059-019-1892-z.

Gampawar P., Schmidt R., Schmidt H. Telomere length and brain aging: A systematic review and meta-analysis. *Ageing Res. Rev.* 2022;80:101679. DOI 10.1016/j.arr.2022.101679.

Garagnani P., Marquis J., Delledonne M., Pirazzini C., Marasco E., Kwiatkowska K.M., Iannuzzi V., … Bertamini L., Martinelli N., Girelli D., Olivieri O., Giuliani C., Descombes P., Franceschi C. Whole-genome sequencing analysis of semi-supercentenarians. *eLife*. 2021;10:e57849. DOI 10.7554/eLife.57849.

Grünewald A., Rygiel K.A., Hepplewhite P.D., Morris C.M., Picard M., Turnbull D.M. Mitochondrial DNA depletion in respiratory chain-deficient Parkinson disease neurons. *Ann. Neurol*. 2016;79(3):366-378. DOI 10.1002/ana.24571.

Hackenhaar F.S., Josefsson M., Adolfsson A.N., Landfors M., Kauppi K., Hultdin M., Adolfsson R., Degerman S., Pudas S. Short leukocyte telomeres predict 25-year Alzheimer's disease incidence in non-APOE ε4-carriers. *Alzheimers Res. Ther*. 2021;13(1):130. DOI 10.1186/s13195-021-00871-y.

Haussmann M.F., Winkler D.W., O'Reilly K.M., Huntington C.E., Nisbet I.C., Vleck C.M. Telomeres shorten more slowly in long-lived birds and mammals than in short-lived ones. *Proc. Biol. Sci*. 2003; 270(1522):1387-1392. DOI 10.1098/rspb.2003.2385.

Haussmann M.F., Winkler D.W., Huntington C.E., Nisbet I.C., Vleck C.M. Telomerase activity is maintained throughout the lifespan of long-lived birds. *Exp. Gerontol*. 2007;42(7):610-618. DOI 10.1016/j.exger.2007.03.004.

Hägg S., Zhan Y., Karlsson R., Gerritsen L., Ploner A., van der Lee S.J., Broer L., … Kuh D., Starr J.M., Deary I.J., Slagboom P.E., van Duijn C.M., Codd V., Pedersen N.L. Short telomere length is associated with impaired cognitive performance in European ancestry cohorts. *Transl. Psychiatry*. 2017;7(4):e1100. DOI 10.1038/tp.2017.73.

He Y.H., Lu X., Wu H., Cai W.W., Yang L.Q., Xu L.Y., Sun H.P., Kong Q.P. Mitochondrial DNA content contributes to healthy aging in Chinese: a study from nonagenarians and centenarians. *Neurobiol. Aging*. 2014;35(7):1779.e1-1779.e4. DOI 10.1016/j.neurobiolaging.2014.01.015.

Herbst A., Widjaja K., Nguy B., Lushaj E.B., Moore T.M., Hevener A.L., McKenzie D., Aiken J.M., Wanagat J. Digital PCR quantitation of muscle mitochondrial DNA: age, fiber type, and mutation-induced changes. *J. Gerontol. A Biol. Sci. Med. Sci*. 2017;72(10):1327-1333. DOI 10.1093/gerona/glx058.

Herbst A., Lee C.C., Vandiver A.R., Aiken J.M., McKenzie D., Hoang A., Allison D., Liu N., Wanagat J. Mitochondrial DNA deletion mutations increase exponentially with age in human skeletal muscle. *Aging Clin. Exp. Res*. 2021a;33(7):1811-1820. DOI 10.1007/s40520-020-01698-7.

Herbst A., Prior S.J., Lee C.C., Aiken J.M., McKenzie D., Hoang A., Liu N., Chen X., Xun P., Allison D.B., Wanagat J. Skeletal muscle mitochondrial DNA copy number and mitochondrial DNA deletion mutation frequency as predictors of physical performance in older men and women. *Geroscience*. 2021b;43(3):1253-1264. DOI 10.1007/s11357-021-00351-z.

Holt A.G., Davies A.M. The effect of mitochondrial DNA half-life on deletion mutation proliferation in long lived cells. *Acta Biotheor*. 2021;69(4):671-695. DOI 10.1007/s10441-021-09417-z.

Huang Z., Sun S., Lee M., Maslov A.Y., Shi M., Waldman S., Marsh A., Siddiqui T., Dong X., Peter Y., Sadoughi A., Shah C., Ye K., Spivack S.D., Vijg J. Single-cell analysis of somatic mutations in human bronchial epithelial cells in relation to aging and smoking. *Nat. Genet*. 2022;54(4):492-498. DOI 10.1038/s41588-022-01035-w.

Hudson G., Nalls M., Evans J.R., Breen D.P., Winder-Rhodes S., Morrison K.E., Morris H.R., Williams-Gray C.H., Barker R.A., Singleton A.B., Hardy J., Wood N.E., Burn D.J., Chinnery P.F. Two-stage association study and meta-analysis of mitochondrial DNA variants in Parkinson disease. *Neurology*. 2013;80(22):2042-2048. DOI 10.1212/WNL.0b013e318294b434.

Hunt S.C., Chen W., Gardner J.P., Kimura M., Srinivasan S.R., Eckfeldt J.H., Berenson G.S., Aviv A. Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study. *Aging Cell*. 2008;7(4):451-458. DOI 10.1111/j.1474-9726.2008.00397.x.

Ikebe S., Tanaka M., Ohno K., Sato W., Hattori K., Kondo T., Mizuno Y., Ozawa T. Increase of deleted mitochondrial DNA in the striatum in Parkinson's disease and senescence. *Biochem. Biophys. Res. Commun.* 1990;170(3):1044-1048. DOI 10.1016/0006-291x(90)90497-b.

Just R.S., Irwin J.A., Parson W. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Sci. Int. Genet*. 2015;18:131-139. DOI 10.1016/j.fsigen.2015.05.003.

Kennedy S.R., Loeb L.A., Herr A.J. Somatic mutations in aging, cancer and neurodegeneration. *Mech. Ageing Dev*. 2012;133(4):118-126. DOI 10.1016/j.mad.2011.10.009.

Kim J.-H., Kim H.K., Ko J.-H., Bang H., Lee D.-C. The relationship between leukocyte mitochondrial DNA copy number and telomere

length in community-dwelling elderly women. *PLoS One*. 2013; 8(6):e67227. DOI 10.1371/journal.pone.0067227.

Kowald A., Kirkwood T. Resolving the enigma of the clonal expansion of mtDNA deletions. *Genes (Basel)*. 2018;9(3):126. DOI 10.3390/genes9030126.

Lee J.W., Park K.D., Im J.A., Kim M.Y., Lee D.C. Mitochondrial DNA copy number in peripheral blood is associated with cognitive function in apparently healthy elderly women *Clin. Chim. Acta.* 2010; 411(7-8):592-596. DOI 10.1016/j.cca.2010.01.024.

Levstek T., Redenšek S., Trošt M., Dolžan V., Podkrajšek K.T. Assessment of the telomere length and its effect on the symptomatology of Parkinson's disease. *Antioxidants (Basel)*. 2021;10(1):137. DOI 10.3390/antiox10010137.

Li H., Slone J., Fei L., Huang T. Mitochondrial DNA variants and common diseases: A mathematical model for the diversity of age-related mtDNA mutations. *Cells*. 2019;8(6):608. DOI 10.3390/cells 8060608.

Lodato M.A., Rodin R.E., Bohrson C.L., Coulter M.E., Barton A.R., Kwon M., Sherman M.A., Vitzthum C.M., Luquette L.J., Yandava C.N., Yang P., Chittenden T.W., Hatem N.E., Ryu S.C., Woodworth M.B., Park P.J., Walsh C.A. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science.* 2018;359(6375):555-559. DOI 10.1126/science.aao4426.

Machado H.E., Mitchell E., Øbro N.F., Kübler K., Davies M., Leongamornlert D., Cull A., Maura F., Sanders M.A., Cagan A.T.J., McDonald C., Belmonte M., Shepherd M.S., Vieira Braga F.A., Osborne R.J., Mahbubani K., Martincorena I., Laurenti E., Green A.R., Getz G., Polak P., Saeb-Parsy K., Hodson D.J., Kent D.G., Campbell P.J. Diverse mutational landscapes in human lymphocytes. *Nature*. 2022;608(7924):724-732. DOI 10.1038/s41586-022-05072-7.

Mahoney E.R., Dumitrescu L., Seto M., Nudelman K.N.H., Buckley R.F., Gifford K.A., Saykin A.J., Jefferson A.J., Hohman T.J. Telomere length associations with cognition depend on Alzheimer's disease biomarkers. *Alzheimers Dement. (NY)*. 2019;5:883-890. DOI 10.1016/j.trci.2019.11.003.

Melicher D., Illés A., Pállinger É., Kovács Á.F., Littvay L., Tárnoki Á.D., Tárnoki D.L., Bikov A., Molnár M.J., Buzás E.I., Falus A. Tight co-twin similarity of monozygotic twins for hTERT protein level of T cell subsets, for telomere length and mitochondrial DNA copy number, but not for telomerase activity. *Cell. Mol. Life Sci*. 2018;75(13):2447-2456. DOI 10.1007/s00018-017-2738-z.

Mengel-From J., Thinggaard M., Dalgård C., Kyvik K.O., Christensen K., Christiansen L. Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. *Hum. Genet*. 2014;133(9):1149-1159. DOI 10.1007/s00439-014-1458-9.

Mikhailova A.G., Mikhailova A.A., Ushakova K., Tretiakov E.O., Iliushchenko D., Shamansky V., Lobanova V., Kozenkov I., Efimenko B., Yurchenko A.A., Kozenkova E., Zdobnov E.M., Makeev V., Yurov V., Tanaka M., Gostimskaya I., Fleischmann Z., Annis S., Franco M., Wasko K., Denisov S., Kunz W.S., Knorre D., Mazunin I., Nikolaev S., Fellay J., Reymond A., Khrapko K., Gunbin K., Popadin K. A mitochondria-specific mutational signature of aging: increased rate of A > G substitutions on the heavy strand. *Nucleic Acids Res*. 2022;50(18):10264-10277. DOI 10.1093/nar/gkac779.

Milholland B., Dong X., Zhang L., Hao X., Suh Y., Vijg J. Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun*. 2017;8(1):15183. DOI 10.1038/ncomms15183.

Miller M.B., Huang A.Y., Kim J., Zhou Z., Kirkham S.L., Maury E.A., Ziegenfuss J.S., Reed H.C., Neil J.E., Rento L., Ryu S.C., Ma C.C., Luquette L.J., Ames H.M., Oakley D.H., Frosch M.P., Hyman B.T., Lodato M.A., Lee E.A., Walsh C.A. Somatic genomic changes in single Alzheimer's disease neurons. *Nature*. 2022;604(7907):714-722. DOI 10.1038/s41586-022-04640-1.

Miwa S., Kashyap S., Chini E., von Zglinicki T. Mitochondrial dysfunction in cell senescence and aging. *J. Clin. Invest*. 2022;132(13): e158447. DOI 10.1172/JCI158447.

Movérare-Skrtic S., Johansson P., Mattsson N., Hansson O., Wallin A., Johansson J.O., Zetterberg H., Blennow K., Svensson J. Leukocyte telomere length (LTL) is reduced in stable mild cognitive impairment but low LTL is not associated with conversion to Alzheimer's disease: a pilot study. *Exp. Gerontol*. 2012;47(2):179-182. DOI 10.1016/j.exger.2011.12.005.

Müezzinler A., Zaineddin A.K., Brenner H. A systematic review of leukocyte telomere length and age in adults. *Ageing Res*. *Rev.* 2013; 12:509-519. DOI 10.1016/j.arr.2013.01.003.

Müller-Nedebock A.C., Meldau S., Lombard C., Abrahams S., van der Westhuizen F.H., Bardien S. Increased blood-derived mitochondrial DNA copy number in African ancestry individuals with Parkinson's disease. *Parkinsonism Relat. Disord*. 2022;101:1-5. DOI 10.1016/j.parkreldis.2022.06.003.

Nido G.S., Dölle C., Flønes I., Tuppen H.A., Alves G., Tysnes O.-B., Haugarvoll K., Tzoulis C. Ultradeep mapping of neuronal mitochondrial deletions in Parkinson's disease. *Neurobiol. Aging*. 2018;63: 120-127. DOI 10.1016/j.neurobiolaging.2017.10.024.

Perier C., Bender A., García-Arumí E., Melià M.J., Bové J., Laub C., Klopstock T., Elstner M., Mounsey R.B., Teismann P., Prolla T., Andreu A.L., Vila M. Accumulation of mitochondrial DNA deletions within dopaminergic neurons triggers neuroprotective mechanisms. *Brain*. 2013;136(Pt. 8):2369-2378. DOI 10.1093/brain/awt196.

Poovathingal S.K., Gruber J., Lakshmanan L., Halliwell B., Gunawan R. Is mitochondrial DNA turnover slower than commonly assumed? *Biogerontology*. 2012;13(5):557-564. DOI 10.1007/s10522-012-9390-7.

Qiu C., Enquobahrie D., Gelaye B., Hevner K., Williams M. The association between leukocyte telomere length and mitochondrial DNA copy number in pregnant women: A pilot study. *Clin. Lab*. 2015; 61(3-4):363-369. DOI 10.7754/Clin.Lab.2014.140313.

Rice A.C., Keeney P.M., Algarzae N.K., Ladd A.C., Thomas R.R., Bennett J.P. Jr. Mitochondrial DNA copy numbers in pyramidal neurons are decreased and mitochondrial biogenesis transcriptome signaling is disrupted in Alzheimer's disease hippocampi. *J. Alzheimer's Dis*. 2014;40(2):319-330. DOI 10.3233/JAD-131715.

Rose G., Romeo G., Dato S., Crocco P., Bruni A.C., Hervonen A., Majamaa K., Sevini F., Franceschi C., Passarino G. Somatic point mutations in mtDNA control region are influenced by genetic background and associated with healthy aging: A GEHA study. *PLoS One*. 2010;5(10):e13395. DOI 10.1371/journal.pone.0013395.

Sallevelt S.C., de Die-Smulders C.E., Hendrickx A.T., Hellebrekers D.M., de Coo I.F., Alston C.L., Knowles C., Taylor R.W., McFarland R., Smeets H.J. De novo mtDNA point mutations are common and have a low recurrence risk. *J. Med. Genet.* 2017;54(2):73-83. DOI 10.1136/jmedgenet-2016-103876.

Sanders J.L., Newman A.B. Telomere length in epidemiology: A biomarker of aging, age-related disease, both, or neither? *Epidemiol. Rev*. 2013;35(1):112-131. DOI 10.1093/epirev/mxs008.

Sondheimer N., Glatz C.E., Tirone J.E., Deardorff M.A., Krieger A.M., Hakonarson H. Neutral mitochondrial heteroplasmy and the influence of aging. *Hum. Mol. Genet*. 2011;20(8):1653-1659. DOI 10.1093/hmg/ddr043.

Tedone E., Huang E., O'Hara R., Batten K., Ludlow A.T., Lai T.-P., Arosio B., Mari D., Wright W.E., Shay J.W. Telomere length and telomerase activity in T cells are biomarkers of high-performing centenarians. *Aging Cell*. 2019;18(1):e12859. DOI 10.1111/acel.12859.

Terry D.F., Nolan V.G., Andersen S.L., Perls T.T., Cawthon R. Association of longer telomeres with better health in centenarians. *J. Gerontol. A Biol. Sci. Med. Sci*. 2008;63(8):809-812. DOI 10.1093/gerona/63.8.809.

Thomas P., O' Callaghan N.J., Fenech M. Telomere length in white blood cells, buccal cells and brain tissue and its variation with ageing and Alzheimer's disease. *Mech. Ageing Dev.* 2008;129(4):183-190. DOI 10.1016/j.mad.2007.12.004.

Tranah G.J., Nalls M.A., Katzman S.M., Yokoyama J.S., Lam E.T., Zhao Y., Mooney S., Thomas F., Newman A.B., Liu Y., Cummings S.R., Harris T.B., Yaffe K. Mitochondrial DNA sequence

variation associated with dementia and cognitive function in the elderly. *J. Alzheimers. Dis*. 2012;32(2):357-372. DOI 10.3233/JAD-2012-120466.

Tyrka A.R., Carpenter L.L., Kao H.-T., Porton B., Philip N.S., Ridout S.J., Ridout K.K., Price L.H. Association of telomere length and mitochondrial DNA copy number in a community sample of healthy adults. *Exp. Gerontol*. 2015;66:17-20. DOI 10.1016/j.exger.2015.04.002.

van den Berg N. Family matters in unraveling human longevity. *Aging*. 2020;12(22):22354-22355. DOI 10.18632/aging.104218.

van Leeuwen N., Beekman M., Deelen J., van den Akker E.B., de Craen A.J.M., Slagboom P.E., 't Hart L.M. Low mitochondrial DNA content associates with familial longevity: the Leiden Longevity Study. *Age (Dordr.)*. 2014;36(3):9629. DOI 10.1007/s11357-014-9629-0.

Vera E., Bernardes de Jesus B., Foronda M., Flores J.M., Blasco M.A. The rate of increase of short telomeres predicts longevity in mammals. *Cell Rep*. 2012;2(4):732-737. DOI 10.1016/j.celrep.2012.08.023.

Victorelli S., Passos J.F. Telomeres and cell senescence – size matters not. *EBioMedicine*. 2017;21:14-20. DOI 10.1016/j.ebiom.2017.03.027.

Vijg J., Dong X. Pathogenic mechanisms of somatic mutation and genome mosaicism in aging. *Cell*. 2020;182(1):12-23. DOI 10.1016/j.cell.2020.06.024.

Wachsmuth M., Hübner A., Li M., Madea B., Stoneking M. Age-related and heteroplasmy-related variation in human mtDNA copy number. *PLoS Genet*. 2016;12(3):e1005939. DOI 10.1371/journal.pgen.1005939.

Yang S.Y., Castellani C.A., Longchamps R.J., Pillalamarri V.K., O'Rourke B., Guallar E., Arking D.E. Blood-derived mitochondrial DNA copy number is associated with gene expression across multiple tissues and is predictive for incident neurodegenerative disease. *Genome Res*. 2021;31(3):349-358. DOI 10.1101/gr.269381.120.

Zhang R., Wang Y., Ye K., Picard M., Gu Z. Independent impacts of aging on mitochondrial DNA quantity and quality in humans. *BMC Genomics*. 2017;18(1):890. DOI 10.1186/s12864-017-4287-0.

Zhao X., Liu X., Zhang A., Chen H., Huo Q., Li W., Ye R., Chen Z., Liang L., Liu Q.A., Shen J., Jin X., Li W., Nygaard M., Liu X., Hou Y., Ni T., Bolund L., Gottschalk W., Tao W., Gu J., Tian X.L., Yang H., Wang J., Xu X., Lutz M.W., Min J., Zeng Y., Nie C. The correlation of copy number variations with longevity in a genome-wide association study of Han Chinese. *Aging (Albany NY)*. 2018;10(6):1206-1222. DOI 10.18632/aging.101461.

ORCID ID

V.P. Volobaev orcid.org/0000-0001-7355-9882
S.S. Kunizheva orcid.org/0000-0003-1882-0667
L.I. Uralsky orcid.org/0000-0002-5565-7961
D.A. Kupriyanova orcid.org/0009-0002-2228-6276
E.I. Rogaev orcid.org/0000-0002-0351-8783

Acknowledgements. This work was supported financially by the Russian Science Foundation under Scientific Project No. 19-75-30039 (R.E.I., K.S.S., U.L.I.) and the Sirius University of Science and Technology under Scientific Project GEN-RND-2019 (V.V.P., K.D.A.).

Received November 28, 2022. Revised February 7, 2023. Accepted February 7, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):502-511
DOI 10.18699/VJGB-23-61

Original Russian text https://vavilovj-icg.ru/

Mosaic loss of the Y chromosome in human neurodegenerative and oncological diseases

I.L. Kuznetsova[1, 2] ✉, L.I. Uralsky[1, 2], T.V. Tyazhelova[1], T.V. Andreeva[1, 2, 3], E.I. Rogaev[1, 2, 3]

[1] Vavilov Institute of General Genetics, Russian Academy of Sciences, Department of Genomics and Human Genetics, Moscow, Russia
[2] Sirius University of Science and Technology, Scientific Center for Genetics and Life Sciences, Sochi, Russia
[3] Lomonosov Moscow State University, Center for Genetics and Genetic Technologies, Faculty of Biology, Moscow, Russia
✉ irakuzn@gmail.com

Abstract. The development of new biomarkers for prediction and early detection of human diseases, as well as for monitoring the response to therapy is one of the most relevant areas of modern human genetics and genomics. Until recently, it was believed that the function of human Y chromosome genes was limited to determining sex and controlling spermatogenesis. Thanks to occurance of large databases of the genome-wide association study (GWAS), there has been a transition to the use of large samples for analyzing genetic changes in both normal and pathological conditions. This has made it possible to assess the association of mosaic aneuploidy of the Y chromosome in somatic cells with a shorter lifespan in men compared to women. Based on data from the UK Biobank, an association was found between mosaic loss of the Y chromosome (mLOY) in peripheral blood leukocytes and the age of men over 70, as well as a number of oncological, cardiac, metabolic, neurodegenerative, and psychiatric diseases. As a result, mLOY in peripheral blood cells has been considered a potential marker of biological age in men and as a marker of certain age-related diseases. Currently, numerous associations have been identified between mLOY and genes based on GWAS and transcriptomes in affected tissues. However, the exact cause of mLOY and the impact and consequences of this phenomenon at the whole organism level have not been established. In particular, it is unclear whether aneuploidy of the Y chromosome in blood cells may affect the development of pathologies that manifest in other organs, such as the brain in Alzheimer's disease, or whether it is a neutral biomarker of general genomic instability. This review examines the main pathologies and genetic factors associated with mLOY, as well as the hypotheses regarding their interplay. Special attention is given to recent studies on mLOY in brain cells in Alzheimer's disease.
Key words: mosaic loss of Y chromosome (mLOY); Alzheimer's disease; GWAS; age-related diseases; oncological diseases.

For citation: Kuznetsova I.L., Uralsky L.I., Tyazhelova T.V., Andreeva T.V., Rogaev E.I. Mosaic loss of the Y chromosome in human neurodegenerative and oncological diseases. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):502-511. DOI 10.18699/VJGB-23-61

Мозаичная потеря Y-хромосомы при нейродегенеративных и онкологических заболеваниях человека

И.Л. Кузнецова[1, 2] ✉, Л.И. Уральский[1, 2], Т.В. Тяжелова[1], Т.В. Андреева[1, 2, 3], Е.И. Рогаев[1, 2, 3]

[1] Институт общей генетики им. Н.И. Вавилова Российской академии наук, отдел геномики и генетики человека, Москва, Россия
[2] Научно-технологический университет «Сириус», Научный центр генетики и наук о жизни, Сочи, Россия
[3] Московский государственный университет им. М.В. Ломоносова, Центр генетики и генетических технологий, биологический факультет, Москва, Россия
✉ irakuzn@gmail.com

Аннотация. Одно из наиболее актуальных направлений современной генетики и геномики человека – разработка новых биомаркеров для прогнозирования и раннего выявления заболеваний человека, а также для мониторинга ответа на терапию. До недавнего времени считалось, что функция генов Y-хромосомы человека ограничена определением пола и контролем сперматогенеза. Благодаря созданию крупных баз данных, полученных на основе метода поиска ассоциаций (GWAS), произошел переход к использованию больших выборок при анализе генетических изменений в норме и при патологиях, в том числе стало возможно оценить связь мозаичной анеуплоидии по Y-хромосоме в соматических клетках с более короткой продолжительностью жизни у мужчин по сравнению с женщинами. На основе данных Британского биобанка (UK Biobank) была обнаружена ассоциация между мозаичной потерей Y-хромосомы (mLOY) в лейкоцитах периферической крови с возрастом мужчин старше 70 лет, а также с рядом онкологических, сердечных, метаболических, нейродегенеративных и психических заболеваний. В результате mLOY в клетках периферической крови стала рассматриваться в качестве потенциального маркера биологического возраста мужчин и как маркер определенных возрастных болезней. На сегодняшний день определено множество ассоциаций между mLOY и генами, выявленными на основании данных GWAS и транскриптомов в пораженных тканях, однако не установлены ни точная причина

возникновения mLOY, ни влияние и последствия этого феномена на уровне всего организма. В частности, неясно, влияет ли анеуплоидия по Y-хромосоме в клетках крови на развитие патологий, проявляющихся в других органах, например на мозг при болезни Альцгеймера, или представляет собой нейтральный биомаркер общей геномной нестабильности. В обзоре рассмотрены основные патологии и генетические факторы, ассоциированные с mLOY, и гипотезы их взаимосвязи. Особое внимание уделено последним исследованиям, посвященным mLOY в клетках мозга при болезни Альцгеймера.
Ключевые слова: мозаичная потеря Y-хромосомы (mLOY); болезнь Альцгеймера; GWAS; возрастные заболевания; онкологические заболевания.

Introduction

Y-chromosome aneuploidy in somatic cells that occurring with age was first identified 60 years ago by karyotyping blood cells (Jacobs et al., 1963) and was attributed to the common age-related phenomenon of accumulation of post-zygotic genetic aberrations of both sex chromosomes and autosomes, unrelated to diseases (Pierre, Hoagland, 1972). Somewhat later karyotyping of bone marrow cells from male patients with acute myeloid leukemia, myelodysplastic syndrome (MDS), myeloproliferative disorders (MPD), and healthy donors was performed to evaluate the contribution of LOY to the malignancy of the diseases (Pierre, Hoagland, 1972, Loss of the Y chromosome…, 1992).

Although aneuploidy is characteristic of many types of tumors, the incidence and level of mLOY both in patients with hematological cancers and healthy donors depend on male age and not on the malignancy of the disease and therefore mLOY cannot be considered a marker of a malignant clone. As a consequence, scientists did not classify mLOY to be a vital factor for several decades, because women survive without the Y chromosome.

More recently the hypothesis of the absence of a connection between mLOY and diseases has been challenged in several studies that used the fluorescent *in situ* hybridization (FISH) method. It was established that the frequency of mLOY is increased in various solid tumors (van Dekken et al., 1990; Sandberg, 1992; König et al., 1994; Takahashi et al., 1994), autoimmune thyroiditis, hematological cancers (Persani et al., 2012), and primary biliary cirrhosis (Lleo et al., 2013). In recent years, with the advent of next-generation sequencing methods and large GWAS databases, including the UK Biobank, with detailed clinical descriptions of donors, it became possible to accurately estimate the frequency of mLOY through statistical analysis. It has been established that 40 % of men over 70 years lack the Y chromosome in more than 5 % of peripheral immune blood cells, and that mLOY in peripheral blood is associated with an increased risk of all-cause mortality (Thompson et al., 2019).

Over the past decade, stable associations were found between mLOY in the blood and various age-related diseases, including hematological (Forsberg et al., 2014; Cáceres et al., 2020) and non-hematological types of cancer (Forsberg et al., 2012, 2014; Cáceres et al., 2020), where the frequency of mLOY varies from 15 to 80 % (Zhang et al., 2007; Bianchi, 2009; Silva Veiga et al., 2012; Duijf et al., 2013), macular degeneration (Forsberg et al., 2014; Cáceres et al., 2020; Duan et al., 2022), cardiovascular diseases (Sano et al., 2022), and neurodegenerative diseases (Dumanski et al., 2016; Vermeulen et al., 2022). In addition, the impact of external factors such as smoking and environmental pollution on mLOY levels in blood cells was revealed (Dumanski et al., 2015; Qin et al., 2019).

Based on these results, it was hypothesized that the biological significance of the Y chromosome goes beyond defining sex and ensuring normal spermatogenesis, and it may contribute to pathological processes, whose molecular mechanisms are yet to be studied. The majority of mLOY studies in pathologies are related to cancer and Alzheimer's disease, which we will examine in more detail.

Pathologies associated with mLOY

LOY in solid tumor cells

mLOY has been detected by FISH in various types of stomach cancer, where the frequency of mLOY varied 61–69 % (Hunter et al., 1993; Beuzen et al., 2000), 33–36 % of cases pancreatic cancer (Chia-Hsien Cheng et al., 2001; Kowalski et al., 2007), and 23–34 % of cases of bladder cancer (Sauter et al., 1995; Minner et al., 2010), and in 12.7 % of breast cancer in men (Agahozo et al., 2020). LOY in various types of renal cell carcinoma has been well studied using FISH, where the frequency of LOY varies from 77 % in papillary renal cell carcinoma to 39 % in the most common type – clear cell renal cell carcinoma (Büscheck et al., 2021).

It has been suggested that aberrations of the Y chromosome may be responsible for the higher incidence of clear cell renal cell carcinoma in men compared to women. Indeed, in addition to the previously known association of copy number variations in autosomes and this disease, analysis of whole-genome data and multiplex PCR have shown that LOY is detected in tumor cells of more than 30 % of male patients (Arseneault et al., 2017). Surprisingly, the frequency of LOY in prostate cancer tumor samples was found to be very low. For example, in a study involving of 2053 specimens from patients who underwent prostatectomy, only 12 cases of Y chromosome aneuploidy were detected (Stahl et al., 2012). This low frequency of LOY may be explained by the fact that mLOY reflects the overall chromosomal instability of tumor cells, which is lower in prostate cancer compared to many other types of cancer (de Matos et al., 2019).

Due to the prevalence of LOY in blood, it was expected to detect an increased frequency of LOY in bone marrow samples from patients with various hematological disorders. In the study of 237 patient samples, including those with MDS, MPN, acute myeloid leukemia, chronic myeloid leukemia, multiple myeloma, and lymphoma, LOY was detected in only 10 % of cases (Zhang et al., 2007).

Unfortunately, the prognostic potential of LOY in tumors is not entirely clear. In head and neck cancer, LOY in tumor cells may be an indicator of therapy resistance (Hollows et al., 2019). In invasive urothelial carcinomas, LOY in tumors was not associated with survival, increased risk of recurrence, or increased risk of progression (Minner et al., 2010). For example, in prostate cancer, no significant associations were found between loss of the Y chromosome in tumor cells and patient age, tumor stage, or risk of recurrence. However, it was significantly associated with a high Gleason score, which is associated with a poor prognosis (Stahl et al., 2012).

In summary, the absence of the Y chromosome is a common event in tumors, but it significantly differs in frequency for different types of cancer, and not enough data have yet been accumulated to understand the reasons for these differences. In other words, the presence and level of mLOY in tumors cannot yet be considered as a marker of malignancy or important information for predicting treatment and recurrence.

mLOY in peripheral blood in hematological and non-hematological cancers

A study of a large cohort from the Uppsala Longitudinal Study of Adult Men (ULSAM) database showed that men with mLOY in blood had an increased risk of cancer diagnosis and mortality from various non-hematological types of cancer (Forsberg et al., 2014). An analysis including data for all solid tumor cases submitted to the UK Biobank was also conducted, with results confirming an association between mLOY frequency and tumor presence (Loftfield et al., 2019). The association of mLOY in blood with the risk of cancer in other organs has been described in several independent studies on a smaller cohort, for example, an increased frequency of mLOY compared to age norms was shown for prostate cancer and colorectal cancer (Noveski et al., 2016).

Loss of the Y chromosome in leukocytes can also act as a marker of a tumor clone in hematological diseases. For example, associations of mLOY with a worse prognosis in leukemias such as AML (Holmes et al., 1985), chronic myeloid leukemia (Lippert et al., 2010), and chronic lymphocytic leukemia (Chapiro et al., 2014). However, in the case of MDS that has not progressed to AML, the presence of mLOY may be a favorable factor for recovery (García-Isidoro et al., 1997). Another identified condition for a stable association with MDS and the ability to predict its course is the presence of mLOY in a very high proportion (more than 75 %) of blood cells (Wiktor et al., 2000; Ouseph et al., 2021).

An attempt to narrow the search area to specific types of cells in the case of MDS was undertaken in the work (Ganster et al., 2015), where the ratio of the level of mLOY in CD34+ cells associated with MDS and mLOY in CD3+ cells not associated with MDS was analyzed. The results showed that the level of mLOY in both CD34+ and CD3+ cells were age-dependent in men without hematological diseases, and in the case of CD34+ cells, it was significantly higher in MDS patients compared to elderly control men without hematological pathologies. These data indicate that the level of LOY in the blood has an age-related basis, but is also associated with MDS. In addition, the authors determined the threshold level of LOY in CD34+ cells in peripheral blood to distinguish age-related changes from increased mLOY in MDS, which was 21.5 %.

Interesting data on a Chinese cohort were obtained in a study of mLOY in lung cancer, where it is a protective factor against the development of lung cancer, but only in non-smoking patients (Qin et al., 2019), which is the only report on a possible protective factor of mLOY, and is difficult to interpret.

Additionally, it is worth noting the results of data analysis from the UK Biobank, contradicting the above, which concluded that mLOY is primarily associated with age and smoking, but not with common types of cancer (Zhou et al., 2016). This conclusion was criticized, as the study used data not only from leukocytes but also from buccal epithelial cells (Forsberg et al., 2019). Later, it was shown that the addition of buccal epithelial cell data to mLOY analysis negated the result obtained for mLOY in leukocytes (Zhou et al., 2016). Thus, despite some disagreements in the results of the studies, it could be concluded that mLOY in human blood has significant potential as one of the biomarkers of malignant diseases, especially when the level of mosaicism is above 20 %.

Mechanism of mLOY geneses

The results of cytological studies suggest that the loss of the Y chromosome in somatic cells occurs during cell division. The most common theory is that mLOY is associated with a mechanism involving the formation of micronuclei with isolated Y chromosomes after improper chromosome segregation during mitosis, followed by the destruction of the micronucleus through autophagy, leading to the appearance of 45,X0 cells in older men (Guttenbach et al., 1994; Ly et al., 2019; Guo et al., 2020). Age-related factors, such as telomere shortening and centromere dysfunction, may increase chromosomal segregation errors. Compared to mosaic aneuploidy of autosomes and the X chromosome, mLOY is a more frequent event, presumably due to its overall high heterochromatinized status and small size (Clark, 2014; Wright et al., 2017). In addition, the human Y chromosome is enriched in repetitive sequences, which may play a role in chromosome segregation errors during mitosis (Jobling, Tyler-Smith, 2017).

It remains unclear at what stage of human leukocyte differentiation the loss of the Y chromosome occurs and the clone having a 45,X0 karyotype is formed. One hypothesis is related to primary loss of the Y chromosome in hematopoietic precursors. This process can be associated with clonal hematopoiesis of indeterminate potential (CHIP), which is defined as the detection of somatic mutations in genes typically associated with myeloid neoplasms in individuals without signs of hematologic malignancy. CHIP occurs due to aging of hematopoietic stem cells that have accumulated mutations, leading to proliferative advantage over their peers, and as a result, to clonal expansion (Genovese et al., 2014; Jaiswal et al., 2014).

CHIP is an age-related phenomenon, regularly observed in healthy elderly individuals with a frequency of up to 10 % at the age of 70, and CHIP is also associated with an increased risk of hematologic malignancies and cardiovascular disease (Genovese et al., 2014; Jaiswal et al., 2014, 2017). Whether mLOY is one of the manifestations associated with CHIP has been considered in the work of V. Ljungström et al. (2022), where the frequency of mutations and mLOY in monocytes of

men aged 65–90 was analyzed. The results of the study indicate a frequent coexistence of mLOY and CHIP in monocytes. Another study also revealed the co-occurrence of LOY and CHIP in bone marrow cells obtained from patients referred for clinical bone marrow examination (Ouseph et al., 2021). It should be noted that in the case of the monocyte study, cases of LOY without CHIP, and vice versa, were also detected. However, the sample size of both studies was very limited (24 and 73 participants), and the obtained result requires confirmation in further studies. It should also be taken into account that CHIP is observed in 10 % of 70-year-old men, and mLOY – in 40 %.

We should also note that mLOY stands out among aneuploidies in leukocytes by a higher frequency of occurrence in compared to other chromosomes, and may apparently be a biomarker of overall chromosomal instability, which is characteristic of general aging of the organism and many human diseases. However, the use of mLOY for diagnostic and prognostic purposes is premature due to the lack of precise mechanistic data on the nature of this phenomenon. It is also yet to be determined whether the observed level of mLOY depends on *de novo* events or high clonality, and how both possible mechanisms are related to stages of different diseases.

Genetic factors of mLOY

The human Y chromosome, according to the latest version (T2T-CHM13v2.0, hereafter T2T-Y) of the human genome assembly is over 62 million base pairs long (Rhie et al., 2022). The Y chromosome consists of three main regions: the ends of the chromosome contain pseudoautosomal regions (PAR1 and PAR2) that are homologous to the ends of the X chromosome, and the remainder (about 95 %) is the male-specific region of the Y chromosome (MSY), which does not undergo recombination (Colaco, Modi, 2018), resulting in the accumulation of repetitive sequences in the MSY.

According to the T2T-Y assembly, the Y chromosome contains 693 annotated genes and 888 transcripts, of which 107 (493 transcripts) are protein-coding. In addition to all the genes annotated in GRCh38-Y, the T2T-Y assembly contains 110 genes, of which 42 are predicted to be protein-coding. Y genes constantly degrade, which may be due to the lack of recombination on this chromosome. The presence of a large number of repeats, in turn, contributes to chromosomal rearrangements and intrachromosomal recombination (Jobling, Tyler-Smith, 2003; de Knijff, 2022). Unfortunately, with the advent of the GWAS era, Y chromosome variants have not been included in most GWAS due to the lack of recombination and high repeat content (Xue, Tyler-Smith, 2017; Parker et al., 2020), making the Y chromosome much less characterized in molecular genetic terms than other human chromosomes.

To identify the molecular genetic factors underlying the occurrence of mLOY in a certain proportion of men, GWAS data for autosomes are primarily used. The first genetic association discovered with mLOY was linked to a single nucleotide variation rs2887399 near the *TCL1A* gene (Zhou et al., 2016). The product of this gene, the TCL1A protein, is involved in carcinogenesis, mainly through chromosomal rearrangements (Laine et al., 2000). A strong association between rs2887399 and mLOY has been replicated in subsequent studies on other cohorts (Wright et al., 2017; Thompson et al., 2019). In total, over 150 genetic variants in autosomes associated with mLOY have been found (Thompson et al., 2019), including variants in genes involved in the regulation of cell cycle (*CCND2*, *CCND3*, *CDKN1B*, *CDKN1C*, *CDK5RAP1*, *ATM*), chromatin structure during mitosis (*NCAPG2*, *SMC2*), and kinetochore structure and function (*CENPN*, *CENPU*, *PMF1*, *ZWILCH*, *SPDL1*).

Associations with cancer susceptibility genes, as well as somatic tumor growth factors and anti-tumor therapy targets, have also been identified. Such genes include those encoding proteins involved in DNA damage response (*SETD2*, *DDB2*, *PARP1*, *ATM*, *TP53*, and *CHEK2*) and in apoptotic processes (*PMAIP1*, *SPOP*, *LTBR*, *SGMS1*, *TP53INP1*, *DAP*, BCL-2 family genes). These associations were confirmed in Japanese and European populations, collectively containing data for over 750,000 men (Thompson et al., 2019), and these variants have later been successfully used to predict mLOY using polygenic risk score estimates (Riaz et al., 2021).

The identified genetic variations in key cell cycle genes provide evidence that cells without a Y chromosome may avoid molecular processes that destroy aneuploid cells, resulting in their proliferation and accumulation in the tissues. Moreover, there are no genes on the Y chromosome associated with somatic cell survival or mitosis, so its absence should not be associated with limitation on cell division (Maan et al., 2017). On the other hand, many of the most commonly observed mutations associated with LOY are linked to general genomic instability. Indeed, based on data from a large-scale cohort GWAS, it has been shown that mosaicisms across autosomes are more common in men with LOY (Zhou et al., 2016).

Most of the mLOY-related genetic variants are often located near genes known as tumor growth encoding factors, targets for cancer treatment, or those that contribute to cancer susceptibility. This is one of the explanations for the association between mLOY and non-hematological cancers and the identified loci are associated not only with various types of male-specific cancers (prostate cancer and germ cell tumors), but also with gender-independent types of cancer (lung cancer, colorectal cancer, glioma, and renal cell carcinoma), as well as with an increased risk of developing specific non-hematological types of cancer in women (breast, ovarian, and endometrial cancer) (Thompson et al., 2019). Based on these results, it can be concluded that mLOY reflects a certain common autosomally determined state of the organism.

In addition to elucidating the root cause of mLOY, the challenge of genetic research on this phenomenon is to identify the consequences of mosaic absence of all genes on the sex chromosome in leukocytes in an elderly man. One hypothesis is that the loss of certain genes in connection with clonal absence of the Y chromosome can be considered as a trigger of molecular and biological processes that lead to age-related pathologies. In particular, it can be assumed that the loss of genes in the PAR1 and PAR2 regions may affect the level of expression of these genes.

One example is the *CD99* gene, located in the PAR1 region of the Y and X chromosomes, which is not subject to X-inactivation in women (Sharp et al., 2011), which may indicate the importance of its balanced expression. The *CD99* gene, most highly expressed in blood cells, encodes the transmembrane glycoprotein CD99, playing an important role in the functioning of the immune system and affects the

key properties of leukocytes (Sohn et al., 2001; Schenkel et al., 2002; Brémond et al., 2009; Pasello et al., 2018). Indeed, CITESeq analysis uncovered a reduced surface expression of CD99 in individual leukocytes lacking a Y chromosome, which may be indicative of a link between LOY and immune functions of leukocytes that is dependent on the homologous regions of sex chromosomes (Mattisson et al., 2021).

Another Y chromosome gene that can directly affect the development of malignant tumors is *KDM6C*, located in the MSY region and encoding histone demethylase. *KDM6C* has a functionally similar X-linked homolog, *KDM6A*, whose deficiency is particularly associated with the progression of clear cell renal cell carcinoma (Arseneault et al., 2017). Other interesting Y chromosome genes functionally corresponding to tumor suppressors are *KDM5B* and *KDM5D* (histone demethylases), *DDX3Y* (RNA helicase), *EIF1AY* (translation initiation factor), *RPS4Y1* (ribosome subunit), *ZFY* (transcription factor) (Dunford et al., 2017; Willis-Owen et al., 2021).

The development of omics technologies has also made it possible to consider in more detail the potential additional functions of Y chromosome genes. Thus, based on proteomic studies, it was found that the *DDX3Y* gene, located in the MSY region of the Y chromosome, can modulate neuronal differentiation (Vakilian et al., 2015), and the Y chromosome haplogroup may be a risk factor for prostate cancer (Cannon-Albright et al., 2014). It has been discovered that SRY (the MSY region that determines sex) may be an oncogenic factor (Murakami et al., 2014), and furthermore, SRY is involved in the molecular genetic pathway associated with pulmonary arterial hypertension (Yan et al., 2018).

There are initial indications of a possible effect of mLOY on autosomal gene expression. In particular, increased expression of the known oncogene *TCL1A* was detected based on single cell transcript sequencing (scRNAseq) data. The product of this gene, the TCL1 protein, is a stimulator of cell proliferation (Thompson et al., 2019). Tumors with mLOY show increased expression of genes involved in resistance to both radiation therapy and platinum-based chemotherapy drugs (Hollows et al., 2019), which partly explains the association of mLOY with treatment prognosis.

Thus, genetic studies suggest that mLOY is determined as a highly polygenic phenomenon. Unfortunately, it is difficult to trace from GWAS data whether single-nucleotide and structural variations in the Y chromosome itself influence the risk of mLOY. In particular, it will be relevant for population studies to determine the relationship between Y chromosome haplogroups and various mLOY indicators. Based on the genetic factors identified through association analysis, it can be concluded that mLOY is influenced by an increase in the frequency of errors during mitosis and a disruption of chromosomal balance recognition and apoptotic regulation.

mLOY in Alzheimer's disease

Alzheimer's disease is a progressive and irreversible neurodegenerative disorder of the central nervous system, responsible for approximately 70 % of all dementia cases. Mutations in the *PSEN1*, *PSEN2*, and *APP* genes were found as the main cause of familial AD (Sorbi et al., 1995; Masters et al., 2015).

Many variations in genes that are risk factors for sporadic AD are also known. Among these, the apolipoprotein E4 (ApoE4) allele is the greatest genetic risk factor, with homozygous ApoE4 carriers having a 14-fold increased susceptibility to AD (Yamazaki et al., 2019). Although the average lifespan of men is shorter than that of women, this age-related fatal disease develops less frequently in men. Therefore, the influence of the male sex chromosome on AD pathogenesis was unexpected. However, the results of the first case-control study of mLOY in AD patients using the European Alzheimer's Disease Initiative stage 1, ULSAM, and the Prospective Study of the Vasculature in Uppsala Seniors (PIVUS) database showed that mLOY is statistically significantly 2.8 times more common in leukocytes of men with AD compared to a control group without brain pathologies (Dumanski et al., 2016). These results were based on the analysis of high-quality sequencing data from 606 blood DNA samples from AD patients and 1005 control samples of all ages. To minimize the influence of age component on both AD and mLOY (risk), separate confirmation of the association between AD and mLOY was obtained for two sub-samples with narrow age ranges: men aged 70–75 and 75–80 years (Dumanski et al., 2016). The identified association persisted even after taking into account the influence of other age-related diseases (cardiovascular disease, diabetes), as well as unhealthy habits (alcohol and smoking). Longitudinal studies, in which participants were tested several times at an interval of about 10 years, allowed the evaluation of mLOY as a risk factor for AD. Thus, during the observation period of ULSAM and PIVUS, 140 individual cases of AD were registered. The results of the data analysis for these individuals, adjusted for the age of the patient at the moment of blood sampling and age-related diseases, showed that mLOY is a significant risk factor for AD, increasing the likelihood of its diagnosis by 6.8-fold.

Since the ApoE4 haplotype is one of the most important risk factors for the development of AD and currently the only confirmed genetic factor that affects lifespan on multiple independent samples (Nebel et al., 2011), its contribution is increasingly being evaluated in studies related to age-related diseases. It is known that the presence of the ApoE2/3/4 haplotype can affect the phenotypic expression of other genetic variants (Kuznetsova et al., 2022). In the case of mLOY, it has been shown that, firstly, the ApoE2/3/4 haplotype does not affect the level of mLOY in leukocytes (Dumanski et al., 2016). Also, the assessment of the joint effect of ApoE4 and mLOY gave a negative result, from which the authors concluded that the risk of developing AD from these factors manifests independently.

However, some studies suggest that an integrative effect of mLOY and genes related to the development of AD is possible. Thus, the influence of the ApoE4 haplotype on the presence of mLOY in dorsolateral prefrontal cortex cells was revealed (Graham et al., 2019), and in a study using neurons obtained from induced pluripotent stem cells from a patient with familial AD with a mutation in the *PSEN1* gene, it was demonstrated that LOY enhances the toxic effects of Aβ42, leading to impared neuron differentiation and premature cell death (Mendivil-Perez et al., 2019). In addition, a statistically significant correlation between mLOY and variation located in proximity to the *SPON1* gene, which is associated with

the severity of dementia, was demonstrated in the Japanese population (Sherva et al., 2014; Terao et al., 2019).

An attempt to answer the key question of whether LOY is a cause or consequence of age-related diseases has also been undertaken for Alzheimer's disease. Quantitative comparisons of LOY frequency showed that the percentage of mLOY is higher in leukocytes than in neurons (Graham et al., 2019), and is also associated with age. Studies using cell lines have shown that the absence of the Y chromosome is more commonly observed in fibroblasts than in neuron-like iPS cells. However, considering that LOY may depend on the proliferative capacity of cells, microglia and oligodendrocyte precursor cells may be more prone to LOY accumulation than terminally differentiated neuronal cells.

Based on scRNAseq data and single-nucleus RNA sequencing by (Vermeulen et al., 2022), LOY enrichment was shown in microglia, as opposed to neurons, astrocytes, and oligodendrocytes, in non-demented patients significantly higher mLOY levels in observed microglia in case of Alzheimer's disease. Comparative analysis of microglia in different brain regions revealed that in the somatosensory cortex of male AD patients 21 % of microglia classified as LOY compared to 1.81 % in controls, while in the entorhinal cortex, the frequency of LOY was 32.7 % in patients and 3.27 % in controls. According to the authors, the elevated level of LOY in the entorhinal cortex of AD patients is of particular interest because this part of the brain is the first to be affected in AD (Gómez-Isla et al., 1996; Kobro-Flatmoen et al., 2021).

Thus, based on several studies, it is becoming increasingly evident that the frequency of mLOY is elevated in AD and cancer, and is also a risk factor for these conditions (Forsberg et al., 2014; Dumanski et al., 2016; Noveski et al., 2016). A dual pathway of development for certain triggers leading to either cancer or AD has been noted in several works (Behrens et al., 2009; Lanni et al., 2021). Indeed, it has been shown that mLOY in blood is a competing risk factor between the onset of AD and solid cancers (Dumanski et al., 2016).

Summing up these results, it could be concluded that mLOY in blood, being a male-specific risk factor for both AD and cancer, may at least partially explain why men on average live shorter lives than women.

Recent studies have provided the first insights into changes in gene expression in brain cells associated with mLOY (Graham et al., 2019). Analysis using a population sample of differentially expressed genes in microglial cells with and without the Y chromosome revealed 193 genes with dysregulated expression upon Y chromosome loss, including genes involved in aging, basic glioma biology, and inflammation (Vermulen et al., 2022). Through the intersection of the list of autosomal genes with dysregulated expression influenced by mLOY in leukocytes and microglia, genes *TMEM176B*, *S100Z*, *TMEM71*, *CD226*, *B2M*, *SCMH1*, *LITAF*, and *IL15*, predominantly related to immune response and inflammation, were identified.

For genes located in the PARs of the sex chromosomes, a correlation was found between mLOY and the expression of *CD99*, previously identified in leukocytes, supporting the hypothesis of a possible disruption in immune system functioning directly associated with Y chromosome loss. In addition to the *CD99* gene, dysregulation of the genes *GTPBP6*, *IL3RA*, *SLC25A6*, *P2RY8*, *AKAP17A*, *DHRSX*, and *CSF2RA*, located in the PAR1 region of the Y chromosome, was observed in microglia. In particular, the *CSF2RA* gene, which is involved in a molecular pathway associated with neurodegeneration, is expressed only in microglia and macrophages associated with the brain. As this study was conducted by a single research group, these findings require confirmation in independent studies.

The association of mLOY in brain cells with AD prompted investigations into the link between this phenomenon and psychiatric disorders. Data have been obtained indicating either no association or weak association with schizophrenia (Hirata et al., 2018), as well as strong association of mLOY with suicidal tendencies, with blood mLOY levels being almost three times higher in the latter case. Interestingly, no changes in mLOY levels were detected in the dorsolateral prefrontal cortex of postmortem brains of individuals who died by suicide, but these data are limited in reliability due to the small sample size (Kimura et al., 2018).

Conclusion

Numerous studies in large cohorts have shown that the LOY in blood cells is a significant risk factor for mortality and various diseases in men. From the accumulated data on associations of mLOY with various pathologies, it becomes increasingly likely that analysis of LOY may become a sensitive biomarker for AD, solid tumors, hematological malignancies, and overall genomic instability. However, it should be kept in mind that no study has provided direct evidence on how mLOY arises, how it affects cells, and what consequences it has. In terms of a predictive potential perspective, it is necessary to understand whether mLOY is a barometer reflecting the presence of pathologies or, conversely, whether mLOY arises *de novo* and participates in the pathogenesis. In this regard, there are several questions that need to be answered, such as whether the pathogenic effect of LOY manifests in untransformed blood cells or in transformed clones in the case of hematological malignancies.

Another question that needs to be resolved is how LOY in normal blood cells can be linked to pathological processes in other organs, leading to tumors in other organs or neurodegeneration in the brain. Currently, the most attractive hypothesis is the immune surveillance hypothesis, explaining the mechanisms underlying associations between LOY in blood cells and enhanced neoplastic cell proliferation in tumors in other parts of the body or the development of Alzheimer's disease.

Comparisons of data for mLOY in different populations are limited by a small number of studies to date and differences in methods. Most of the obtained data are ethnically limited and may not be confirmed in poorly represented populations in databases, such as African or Middle Eastern populations. Based on two studies from the same research group, it can be preliminarily concluded that the prevalence of mLOY in men is higher in European populations than in African populations (Loftfield et al., 2018, 2019). However, it should be noted that such comparisons are currently limited.

Despite the many questions that are likely to be actively addressed in connection with the popularity of this topic, it can already be concluded that mLOY plays a role in determining the health of elderly men.

References

Agahozo M.C., Timmermans M.A.M., Sleddens H.F.B.M., Foekens R., Trapman-Jansen A.M.A.C., Schröder C.P., van Leeuwen-Stok E., Martens J.W.M., Dinjens W.N.M., van Deurzen C.H.M. Loss of Y-chromosome during male breast carcinogenesis. *Cancers*. 2020; 12(3):631. DOI 10.3390/cancers12030631.

Arseneault M., Monlong J., Vasudev N.S., Laskar R.S., Safisamghabadi M., Harnden P., Egevad L., Nourbehesht N., Panichnantakul P., Holcatova I., Brisuda A., Janout V., Kollarova H., Foretova L., Navratilova M., Mates D., Jinga V., Zaridze D., Mukeria A., Jandaghi P., Brennan P., Brazma A., Tost J., Scelo G., Banks R.E., Lathrop M., Bourque G., Riazalhosseini Y. Loss of chromosome Y leads to down regulation of KDM5D and KDM6C epigenetic modifiers in clear cell renal cell carcinoma. *Sci. Rep*. 2017;7(1):44876. DOI 10.1038/srep44876.

Behrens M., Lendon C., Roe C. A common biological mechanism in cancer and Alzheimer's disease? *Curr. Alzheimer Res*. 2009;6(3): 196-204. DOI 10.2174/156720509788486608.

Bellenguez C., Küçükali F., Jansen I.E., Kleineidam L., Moreno-Grau S., Amin N., Naj A.C., … Schellenberg G.D., van Duijn C.M., Sims R., van der Flier W.M., Ruiz A., Ramirez A., Lambert J.-C. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet*. 2022;54(4):412-436. DOI 10.1038/s41588-022-01024-z.

Beuzen F., Dubois S., Fléjou J.-F. Chromosomal numerical aberrations are frequent in oesophageal and gastric adenocarcinomas: a study using in-situ hybridization. *Histopathology*. 2000;37(3):241-249. DOI 10.1046/j.1365-2559.2000.00887.x.

Bianchi N.O. Y chromosome structural and functional changes in human malignant diseases. *Mutat. Res.* 2009;682(1):21-27. DOI 10.1016/j.mrrev.2009.02.001.

Brémond A., Meynet O., Mahiddine K., Coito S., Tichet M., Scotlandi K., Breittmayer J.-P., Gounon P., Gleeson P.A., Bernard A., Bernard G. Regulation of HLA class I surface expression requires CD99 and p230/golgin-245 interaction. *Blood*. 2009;113(2):347-357. DOI 10.1182/blood-2008-02-137745.

Büscheck F., Fraune C., Garmestani S., Simon R., Kluth M., Hube-Magg C., Ketterer K., Eichelberg C., Höflmayer D., Jacobsen F., Wittmer C., Wilczak W., Sauter G., Fisch M., Eichenauer T., Rink M. Y-chromosome loss is frequent in male renal tumors. *Ann. Transl. Med*. 2021;9(3):209. DOI 10.21037/atm-20-3061.

Cáceres A., Jene A., Esko T., Pérez-Jurado L.A., González J.R. Extreme downregulation of chromosome Y and cancer risk in men. *J. Natl. Cancer Inst*. 2020;112(9):913-920. DOI 10.1093/jnci/djz232.

Cannon-Albright L.A., Farnham J.M., Bailey M., Albright F.S., Teerlink C.C., Agarwal N., Stephenson R.A., Thomas A. Identification of specific Y chromosomes associated with increased prostate cancer risk. *Prostate*. 2014;74(9):991-998. DOI 10.1002/pros.22821.

Chapiro E., Antony-Debre I., Marchay N., Parizot C., Lesty C., Cung H.-A., Mathis S., Grelier A., Maloum K., Choquet S., Azgui Z., Uzunov M., Leblond V., Merle-Beral H., Sutton L., Davi F., Nguyen-Khac F. Sex chromosome loss may represent a disease-associated clonal population in chronic lymphocytic leukemia. *Genes Chromosom. Cancer*. 2014;53(3):240-247. DOI 10.1002/gcc.22134.

Chia-Hsien Cheng J., Chuang V.P., Cheng S.H., Lin Y.-M., Cheng T.-I., Yang P.-S., Jian J.J.-M., You D.-L., Horng C.-F., Huang A.T. Unresectable hepatocellular carcinoma treated with radiotherapy and/or chemoembolization. *Int. J. Cancer*. 2001;96(4):243-252. DOI 10.1002/ijc.1022.

Clark A.G. The vital Y chromosome. *Nature*. 2014;508(7497):463-464. DOI 10.1038/508463a.

Colaco S., Modi D. Genetics of the human Y chromosome and its association with male infertility. *Reprod. Biol. Endocrinol.* 2018;16: 1-24. DOI 10.1186/s12958-018-0330-5.

de Knijff P. On the forensic use of Y-chromosome polymorphisms. *Genes*. 2022;13(5):898. DOI 10.3390/genes13050898.

de Matos M., Posa I., Carvalho F., Morais V., Grosso A., de Almeida S. A systematic pan-cancer analysis of genetic heterogeneity reveals associations with epigenetic modifiers. *Cancers*. 2019;11(3):391. DOI 10.3390/cancers11030391.

Duan Q., Gao Y., Cao X., Wang S., Xu M., Jones O.D., Xu X. Mosaic loss of chromosome Y in peripheral blood cells is associated with age-related macular degeneration in men. *Cell Biosci*. 2022;12(1):73. DOI 10.1186/s13578-022-00811-9.

Duijf P.H.G., Schultz N., Benezra R. Cancer cells preferentially lose small chromosomes. *Int. J. Cancer*. 2013;132(10):2316-2326. DOI 10.1002/ijc.27924.

Dumanski J.P., Rasi C., Lönn M., Davies H., Ingelsson M., Giedraitis V., Lannfelt L., Magnusson P.K.E., Lindgren C.M., Morris A.P., Cesarini D., Johannesson M., Tiensuu Janson E., Lind L., Pedersen N.L., Ingelsson E., Forsberg L.A. Mutagenesis. Smoking is associated with mosaic loss of chromosome Y. *Science*. 2015;347(6217):81-83. DOI 10.1126/science.1262092.

Dumanski J.P., Lambert J.-C., Rasi C., Giedraitis V., Davies H., Grenier-Boley B., Lindgren C.M., Campion D., Dufouil C., Pasquier F., Amouyel P., Lannfelt L., Ingelsson M., Kilander L., Lind L., Forsberg L.A. Mosaic loss of chromosome Y in blood is associated with Alzheimer disease. *Am. J. Hum. Genet*. 2016;98(6):1208-1219. DOI 10.1016/j.ajhg.2016.05.014.

Dunford A., Weinstock D.M., Savova V., Schumacher S.E., Cleary J.P., Yoda A., Sullivan T.J., Hess J.M., Gimelbrant A.A., Beroukhim R., Lawrence M.S., Getz G., Lane A.A. Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. *Nat. Genet*. 2017;49(1):10-16. DOI 10.1038/ng.3726.

Forsberg L.A., Rasi C., Razzaghian H.R., Pakalapati G., Waite L., Thilbeault K.S., Ronowicz A., Wineinger N.E., Tiwari H.K., Boomsma D., Westerman M.P., Harris J.R., Lyle R., Essand M., Eriksson F., Assimes T.L., Iribarren C., Strachan E., O'Hanlon T.P., Rider L.G., Miller F.W., Giedraitis V., Lannfelt L., Ingelsson M., Piotrowski A., Pedersen N.L., Absher D., Dumanski J.P. Age-related somatic structural changes in the nuclear genome of human blood cells. *Am. J. Hum. Genet*. 2012;90(2):217-228. DOI 10.1016/j.ajhg.2011.12.009.

Forsberg L.A., Rasi C., Malmqvist N., Davies H., Pasupulati S., Pakalapati G., Sandgren J., de Ståhl T.D., Zaghlool A., Giedraitis V., Lannfelt L., Score J., Cross N.C.P., Absher D., Janson E.T., Lindgren C.M., Morris A.P., Ingelsson E., Lind L., Dumanski J.P. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat. Genet*. 2014;46(6):624-628. DOI 10.1038/ng.2966.

Forsberg L.A., Halvardson J., Rychlicka-Buniowska E., Danielsson M., Moghadam B.T., Mattisson J., Rasi C., Davies H., Lind L., Giedraitis V., Lannfelt L., Kilander L., Ingelsson M., Dumanski J.P. Mosaic loss of chromosome Y in leukocytes matters. *Nat. Genet*. 2019;51(1):4-7. DOI 10.1038/s41588-018-0267-9.

Ganster C., Kämpfe D., Jung K., Braulke F., Shirneshan K., Machherndl-Spandl S., Suessner S., Bramlage C.P., Legler T.J., Koziolek M.J., Haase D., Schanz J. New data shed light on Y-loss-related pathogenesis in myelodysplastic syndromes. *Genes Chromosom. Cancer*. 2015;54(12):717-724. DOI 10.1002/gcc.22282.

García-Isidoro M., Tabernero M.D., Najera M.L., García J.L., Hernandez J.M., Duran A., San Miguel J.F., Orfao A. Clinical and biological characteristics of myelodysplastic syndromes with nulisomy Y by fish. *Haematologica*. 1997;82(5):537-541.

Genovese G., Kähler A.K., Handsaker R.E., Lindberg J., Rose S.A., Bakhoum S.F., Chambert K., Mick E., Neale B.M., Fromer M., Purcell S.M., Svantesson O., Landén M., Höglund M., Lehmann S., Gabriel S.B., Moran J.L., Lander E.S., Sullivan P.F., Sklar P., Grönberg H., Hultman C.M., McCarroll S.A. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med*. 2014;371(26):2477-2487. DOI 10.1056/NEJMoa1409405.

Gómez-Isla T., Price J.L., McKeel Jr. D.W., Morris J.C., Growdon J.H., Hyman B.T. Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *J. Neurosci*. 1996;16(14): 4491-4500. DOI 10.1523/JNEUROSCI.16-14-04491.1996.

Graham E.J., Vermeulen M., Vardarajan B., Bennett D., De Jager P., Pearse R.V., Young-Pearse T.L., Mostafavi S. Somatic mosaicism

of sex chromosomes in the blood and brain. *Brain Res*. 2019;1721: 146345. DOI 10.1016/j.brainres.2019.146345.

Guo X., Dai X., Zhou T., Wang H., Ni J., Xue J., Wang X. Mosaic loss of human Y chromosome: what, how and why. *Hum. Genet*. 2020;139(4):421-446. DOI 10.1007/s00439-020-02114-w.

Guttenbach M., Schakowski R., Schmid M. Aneuploidy and ageing: sex chromosome exclusion into micronuclei. *Hum. Genet*. 1994; 94(3):295-298. DOI 10.1007/BF00208287.

Hirata T., Hishimoto A., Otsuka I., Okazaki S., Boku S., Kimura A., Horai T., Sora I. Investigation of chromosome Y loss in men with schizophrenia. *Neuropsychiatr. Dis. Treat.* 2018;14:2115-2122. DOI 10.2147/NDT.S172886.

Hollows R., Wei W., Cazier J., Mehanna H., Parry G., Halford G., Murray P. Association between loss of Y chromosome and poor prognosis in male head and neck squamous cell carcinoma. *Head Neck*. 2019;41(4):993-1006. DOI 10.1002/hed.25537.

Holmes R.I., Keating M.J., Cork A., Trujillo J.M., McCredie K.B., Freireich E.J. Loss of the Y chromosome in acute myelogenous leukemia: A report of 13 patients. *Cancer Genet. Cytogenet*. 1985; 17(3):269-278. DOI 10.1016/0165-4608(85)90018-4.

Hunter S., Gramlich T., Abbott K., Varma V. Y chromosome loss in esophageal carcinoma: An in situ hybridization study. *Genes Chromosomes Cancer*. 1993;8(3):172-177. DOI 10.1002/gcc.2870080306.

Jacobs P.A., Brunton M., Court Brown W.M., Doll R., Goldstein H. Change of human chromosome count distributions with age: evidence for a sex difference. *Nature*. 1963;197(4872):1080-1081. DOI 10.1038/1971080a0.

Jaiswal S., Fontanillas P., Flannick J., Manning A., Grauman P.V., Mar B.G., Lindsley R.C., Mermel C.H., Burtt N., Chavez A., Higgins J.M., Moltchanov V., Kuo F.C., Kluk M.J., Henderson B., Kinnunen L., Koistinen H.A., Ladenvall C., Getz G., Correa A., Banahan B.F., Gabriel S., Kathiresan S., Stringham H.M., McCarthy M.I., Boehnke M., Tuomilehto J., Haiman C., Groop L., Atzmon G., Wilson J.G., Neuberg D., Altshuler D., Ebert B.L. Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med*. 2014;371(26):2488-2498. DOI 10.1056/NEJMoa1408617.

Jaiswal S., Natarajan P., Silver A.J., Gibson C.J., Bick A.G., Shvartz E., McConkey M., Gupta N., Gabriel S., Ardissino D., Baber U., Mehran R., Fuster V., Danesh J., Frossard P., Saleheen D., Melander O., Sukhova G.K., Neuberg D., Libby P., Kathiresan S., Ebert B.L. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N. Engl. J. Med*. 2017;377(2):111-121. DOI 10.1056/NEJMoa1701719.

Jobling M.A., Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet*. 2003;4(8):598-612. DOI 10.1038/nrg1124.

Jobling M.A., Tyler-Smith C. Human Y-chromosome variation in the genome-sequencing era. *Nat. Rev. Genet*. 2017;18(8):485-497. DOI 10.1038/nrg.2017.36.

Kimura A., Hishimoto A., Otsuka I., Okazaki S., Boku S., Horai T., Izumi T., Takahashi M., Ueno Y., Shirakawa O., Sora I. Loss of chromosome Y in blood, but not in brain, of suicide completers. *PLoS One*. 2018;13(1):e0190667. DOI 10.1371/journal.pone.0190667.

Kobro-Flatmoen A., Lagartos-Donate M.J., Aman Y., Edison P., Witter M.P., Fang E.F. Re-emphasizing early Alzheimer's disease pathology starting in select entorhinal neurons, with a special focus on mitophagy. *Ageing Res. Rev*. 2021;67:101307. DOI 10.1016/j.arr.2021.101307.

König J.J., Teubel W., Van Dongen J.W., Romijn J.C., Hagemeijer A., Schröder F.H. Loss and gain of chromosomes 1, 18, and Y in prostate cancer. *Prostate*. 1994;25(6):281-291. DOI 10.1002/pros.2990250602.

Kowalski J., Morsberger L.A., Blackford A., Hawkins A., Yeo C.J., Hruban R.H., Griffin C.A. Chromosomal abnormalities of adenocarcinoma of the pancreas: identifying early and late changes. *Cancer Genet. Cytogenet*. 2007;178(1):26-35. DOI 10.1016/j.cancergencyto.2007.06.004.

Kuznetsova I.L., Ponomareva N.V., Alemastseva E.A., Manakhov A.D., Andreeva T.V., Gusev F.E., Rogaev E.I. The interactive effect of genetic and epigenetic variations in *FKBP5* and *ApoE* genes on anxiety and brain EEG parameters. *Genes*. 2022;13(2):164. DOI 10.3390/genes13020164.

Laine J., Künstle G., Obata T., Sha M., Noguchi M. The protooncogene *TCL1* is an akt kinase coactivator. *Mol. Cell*. 2000;6(2):395-407. DOI 10.1016/S1097-2765(00)00039-3.

Lanni C., Masi M., Racchi M., Govoni S. Cancer and Alzheimer's disease inverse relationship: an age-associated diverging derailment of shared pathways. *Mol. Psychiatry*. 2021;26(1):280-295. DOI 10.1038/s41380-020-0760-2.

Lippert E., Etienne G., Mozziconacci M.-J., Laibe S., Gervais C., Girault S., Gachard N., Tigaud I., Dastugue N., Huguet F., Fort M.-P., Legros L., Eclache V., Mahon F.-X. Loss of the Y chromosome in Philadelphia-positive cells predicts a poor response of chronic myeloid leukemia patients to imatinib mesylate therapy. *Haematologica*. 2010;95(9):1604-1607. DOI 10.3324/haematol.2009.019109.

Ljungström V., Mattisson J., Halvardson J., Pandzic T., Davies H., Rychlicka-Buniowska E., Danielsson M., Lacaze P., Cavelier L., Dumanski J.P., Baliakas P., Forsberg L.A. Loss of Y and clonal hematopoiesis in blood – two sides of the same coin? *Leukemia*. 2022; 36(3):889-891. DOI 10.1038/s41375-021-01456-2.

Lleo A., Oertelt-Prigione S., Bianchi I., Caliari L., Finelli P., Miozzo M., Lazzari R., Floreani A., Donato F., Colombo M., Gershwin M.E., Podda M., Invernizzi P. Y chromosome loss in male patients with primary biliary cirrhosis. *J. Autoimmun*. 2013;41:87-91. DOI 10.1016/j.jaut.2012.12.008.

Loftfield E., Zhou W., Graubard B.I., Yeager M., Chanock S.J., Freedman N.D., Machiela M.J. Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank. *Sci. Rep*. 2018; 8(1):12316. DOI 10.1038/s41598-018-30759-1.

Loftfield E., Zhou W., Yeager M., Chanock S.J., Freedman N.D., Machiela M.J. Mosaic Y loss is moderately associated with solid tumor risk. *Cancer Res*. 2019;79(3):461-466. DOI 10.1158/0008-5472.CAN-18-2566.

Loss of the Y chromosome from normal and neoplastic bone marrows. United Kingdom Cancer Cytogenetics Group (UKCCG). *Genes Chromosom. Cancer*. 1992;5(1):83-88. DOI 10.1002/gcc.2870050112.

Ly P., Brunner S.F., Shoshani O., Kim D.H., Lan W., Pyntikova T., Flanagan A.M., Behjati S., Page D.C., Campbell P.J., Cleveland D.W. Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. *Nat. Genet*. 2019; 51(4):705-715. DOI 10.1038/s41588-019-0360-8.

Maan A.A., Eales J., Akbarov A., Rowland J., Xu X., Jobling M.A., Charchar F.J., Tomaszewski M. The Y chromosome: a blueprint for men's health? *Eur. J. Hum. Genet*. 2017;25(11):1181-1188. DOI 10.1038/ejhg.2017.128.

Masters C.L., Bateman R., Blennow K., Rowe C.C., Sperling R.A., Cummings J.L. Alzheimer's disease. *Nat. Rev. Dis. Primers*. 2015; 1(1):15056. DOI 10.1038/nrdp.2015.56.

Mattisson J., Danielsson M., Hammond M., Davies H., Gallant C.J., Nordlund J., Raine A., Edén M., Kilander L., Ingelsson M., Dumanski J.P., Halvardson J., Forsberg L.A. Leukocytes with chromosome Y loss have reduced abundance of the cell surface immunoprotein CD99. *Sci. Rep*. 2021;11(1):15160. DOI 10.1038/s41598-021-94588-5.

Mendivil-Perez M., Velez-Pardo C., Kosik K.S., Lopera F., Jimenez-Del-Rio M. iPSCs-derived nerve-like cells from familial Alzheimer's disease PSEN 1 E280A reveal increased amyloid-beta levels and loss of the Y chromosome. *Neurosci. Lett*. 2019;703:111-118. DOI 10.1016/j.neulet.2019.03.032.

Minner S., Kilgué A., Stahl P., Weikert S., Rink M., Dahlem R., Fisch M., Höppner W., Wagner W., Bokemeyer C., Terracciano L., Simon R., Sauter G., Wilczak W. Y chromosome loss is a frequent early event in urothelial bladder cancer. *Pathology*. 2010;42(4):356-359. DOI 10.3109/00313021003767298.

Murakami S., Chishima S., Uemoto H., Sakamoto E., Sato T., Kurabe N., Kawasaki Y., Shibata T., Akiyama H., Tashiro F. The

male-specific factor Sry harbors an oncogenic function. *Oncogene*. 2014;33(23):2978-2986. DOI 10.1038/onc.2013.262.

Nebel A., Kleindorp R., Caliebe A., Nothnagel M., Blanché H., Junge O., Wittig M., Ellinghaus D., Flachsbart F., Wichmann H.-E., Meitinger T., Nikolaus S., Franke A., Krawczak M., Lathrop M., Schreiber S. A genome-wide association study confirms *APOE* as the major gene influencing survival in long-lived individuals. *Mech. Ageing Dev*. 2011;132(6-7):324-330. DOI 10.1016/j.mad.2011.06.008.

Noveski P., Madjunkova S., Sukarova Stefanovska E., Matevska Geshkovska N., Kuzmanovska M., Dimovski A., Plaseska-Karanfilska D. Loss of Y chromosome in peripheral blood of colorectal and prostate cancer patients. *PLoS One*. 2016;11(1):e0146264. DOI 10.1371/journal.pone.0146264.

Ouseph M.M., Hasserjian R.P., Dal Cin P., Lovitch S.B., Steensma D.P., Nardi V., Weinberg O.K. Genomic alterations in patients with somatic loss of the Y chromosome as the sole cytogenetic finding in bone marrow cells. *Haematologica*. 2021;106(2):555-564. DOI 10.3324/haematol.2019.240689.

Parker K., Erzurumluoglu A.M., Rodriguez S. The Y chromosome: a complex locus for genetic analyses of complex human traits. *Genes*. 2020;11(11):1273. DOI 10.3390/genes11111273.

Pasello M., Manara M.C., Scotlandi K. CD99 at the crossroads of physiology and pathology. *J. Cell Commun. Signal*. 2018;12(1):55-68. DOI 10.1007/s12079-017-0445-z.

Persani L., Bonomi M., Lleo A., Pasini S., Civardi F., Bianchi I., Campi I., Finelli P., Miozzo M., Castronovo C., Sirchia S., Gershwin M.E., Invernizzi P. Increased loss of the Y chromosome in peripheral blood cells in male patients with autoimmune thyroiditis. *J. Autoimmun*. 2012;38(2-3):J193-J196. DOI 10.1016/j.jaut.2011.11.011.

Pierre R.V., Hoagland H.C. Age-associated aneuploidy: Loss of Y chromosome from human bone marrow cells with aging. *Cancer*. 1972; 30(4):889-894. DOI 10.1002/1097-0142(197210)30:4<889::aid-cncr2820300405>3.0.co;2-1.

Qin N., Li N., Wang C., Pu Z., Ma Z., Jin G., Zhu M., Dai M., Hu Z., Ma H., Shen H. Association of mosaic loss of chromosome Y with lung cancer risk and prognosis in a Chinese population. *J. Thorac. Oncol*. 2019;14(1):37-44. DOI 10.1016/j.jtho.2018.09.013.

Rhie A., Nurk S., Cechova M., Hoyt S.J., Taylor D.J., Altemose N., Hook P.W., … Zook J.M., Eichler E.E., O'Neill R., Schatz M.C., Miga K.H., Makova K.D., Phillippy A.M. The complete sequence of a human Y chromosome. *bioRxiv*. 2022. DOI 10.1101/2022.12.01.518724.

Riaz M., Mattisson J., Polekhina G., Bakshi A., Halvardson J., Danielsson M., Ameur A., McNeil J., Forsberg L.A., Lacaze P. A polygenic risk score predicts mosaic loss of chromosome Y in circulating blood cells. *Cell Biosci*. 2021;11(1):205. DOI 10.1186/s13578-021-00716-z.

Sandberg A.A. Chromosome changes in early bladder neoplasms. *J. Cell. Biochem*. 1992;50(S16I):76-79. DOI 10.1002/jcb.240501317.

Sano S., Horitani K., Ogawa H., Halvardson J., Chavkin N.W., Wang Y., Sano M., Mattisson J., Hata A., Danielsson M., Miura-Yura E., Zaghlool A., Evans M.A., Fall T., De Hoyos H.N., Sundström J., Yura Y., Kour A., Arai Yohei, Thel M.C., Arai Yuka, Mychaleckyj J.C., Hirschi K.K., Forsberg L.A., Walsh K. Hematopoietic loss of Y chromosome leads to cardiac fibrosis and heart failure mortality. *Science*. 2022;377(6603):292-297. DOI 10.1126/science.abn3100.

Sauter G., Moch H., Wagner U., Novotna H., Gasser T.C., Mattarelli G., Mihatsch M.J., Waldman F.M. Y chromosome loss detected by FISH in bladder cancer. *Cancer Genet. Cytogenet*. 1995;82(2):163-169. DOI 10.1016/0165-4608(95)00030-S.

Schenkel A.R., Mamdouh Z., Chen X., Liebman R.M., Muller W.A. CD99 plays a major role in the migration of monocytes through endothelial junctions. *Nat. Immunol*. 2002;3(2):143-150. DOI 10.1038/ni749.

Sharp A.J., Stathaki E., Migliavacca E., Brahmachary M., Montgomery S.B., Dupre Y., Antonarakis S.E. DNA methylation profiles of human active and inactive X chromosomes. *Genome Res*. 2011; 21(10):1592-1600. DOI 10.1101/gr.112680.110.

Sherva R., Tripodis Y., Bennett D.A., Chibnik L.B., Crane P.K., de Jager P.L., Farrer L.A., Saykin A.J., Shulman J.M., Naj A., Green R.C. Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimers Dement*. 2014;10(1):45-52. DOI 10.1016/j.jalz.2013.01.008.

Silva Veiga L.C., Bérgamo N.A., Reis P.P., Kowalski L.P., Rogatto S.R. Loss of Y-chromosome does not correlate with age at onset of head and neck carcinoma: a case-control study. *Braz. J. Med. Biol. Res*. 2012;45(2):172-178. DOI 10.1590/S0100-879X2012007500004.

Sohn H.W., Shin Y.K., Lee I.-S., Bae Y.M., Suh Y.H., Kim M.K., Kim T.J., Jung K.C., Park W.S., Park C.-S., Chung D.H., Ahn K., Kim I.S., Ko Y.H., Bang Y.J., Kim C.W., Park S.H. CD99 regulates the transport of MHC class I molecules from the Golgi complex to the cell surface. *J. Immunol*. 2001;166(2):787-794. DOI 10.4049/jimmunol.166.2.787.

Sorbi S., Nacmias B., Forleo P., Piacentini S., Sherrington R., Rogaev E., Amaducci L. Misserise mutation of S182 gene in Italian families with early-onset Alzheimer's disease. *Lancet*. 1995;346(8972):439-440. DOI 10.1016/S0140-6736(95)92809-X.

Stahl P.R., Kilgué A., Tennstedt P., Minner S., Krohn A., Simon R., Krause G.V., Izbicki J., Graefen M., Sauter G., Schlomm T., Wilczak W. Y chromosome losses are exceedingly rare in prostate cancer and unrelated to patient age. *Prostate*. 2012;72(8):898-903. DOI 10.1002/pros.21492.

Takahashi S., Qian J., Brown J.A., Alcaraz A., Bostwick D.G., Lieber M.M., Jenkins R.B. Potential markers of prostate cancer aggressiveness detected by fluorescence *in situ* hybridization in needle biopsies. *Cancer Res*. 1994;54(13):3574-3579.

Terao C., Momozawa Y., Ishigaki K., Kawakami E., Akiyama M., Loh P.-R., Genovese G., Sugishita H., Ohta T., Hirata M., Perry J.R.B., Matsuda K., Murakami Y., Kubo M., Kamatani Y. GWAS of mosaic loss of chromosome Y highlights genetic effects on blood cell differentiation. *Nat. Commun*. 2019;10(1):4719. DOI 10.1038/s41467-019-12705-5.

Thompson D.J., Genovese G., Halvardson J., Ulirsch J.C., Wright D.J., Terao C., Davidsson O.B., … Ong K.K., Machiela M.J., Loh P.-R., Dumanski J.P., Chanock S.J., Forsberg L.A., Perry J.R.B. Genetic predisposition to mosaic Y chromosome loss in blood. *Nature*. 2019; 575(7784):652-657. DOI 10.1038/s41586-019-1765-3.

Vakilian H., Mirzaei M., Sharifi Tabar M., Pooyan P., Habibi Rezaee L., Parker L., Haynes P.A., Gourabi H., Baharvand H., Salekdeh G.H. *DDX3Y*, a male-specific region of Y chromosome gene, may modulate neuronal differentiation. *J. Proteome Res*. 2015;14(9):3474-3483. DOI 10.1021/acs.jproteome.5b00512.

van Dekken H., Pizzolo J.G., Kelsen D.P., Melamed M.R. Targeted cytogenetic analysis of gastric tumors byin situhybridization with a set of chromosome-specific DNA probes. *Cancer*. 1990;66(3):491-497. DOI 10.1002/1097-0142(19900801)66:3<491::aid-cncr2820660315>3.0.co;2-Q.

Vermeulen M.C., Pearse R., Young-Pearse T., Mostafavi S. Mosaic loss of chromosome Y in aged human microglia. *Genome Res*. 2022; 32(10):1795-1807. DOI 10.1101/gr.276409.121.

Wiktor A., Rybicki B.A., Piao Z.S., Shurafa M., Barthel B., Maeda K., Van Dyke D.L. Clinical significance of Y chromosome loss in hematologic disease. *Genes Chromosomes Cancer*. 2000;27(1):11-16. DOI 10.1002/(SICI)1098-2264(200001)27:1<11::aid-gcc2>3.0.co;2-I.

Willis-Owen S.A.G., Domingo-Sabugo C., Starren E., Liang L., Freidin M.B., Arseneault M., Zhang Y., Lu S.K., Popat S., Lim E., Nicholson A.G., Riazalhosseini Y., Lathrop M., Cookson W.O.C., Moffatt M.F. Y disruption, autosomal hypomethylation and poor male lung cancer survival. *Sci. Rep*. 2021;11(1):12453. DOI 10.1038/s41598-021-91907-8.

Wright D.J., Day F.R., Kerrison N.D., Zink F., Cardona A., Sulem P., Thompson D.J., Sigurjonsdottir S., Gudbjartsson D.F., Helgason A., Chapman J.R., Jackson S.P., Langenberg C., Wareham N.J., Scott R.A., Thorsteindottir U., Ong K.K., Stefansson K., Perry J.R.B. Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility. *Nat. Genet*. 2017;49(5):674-679. DOI 10.1038/ng.3821.

Xue Y., Tyler-Smith C. Past successes and future opportunities for the genetics of the human Y chromosome. *Hum. Genet*. 2017;136(5):481-483. DOI 10.1007/s00439-017-1806-7.

Yamazaki Y., Zhao N., Caulfield T.R., Liu C.-C., Bu G. Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. *Nat. Rev. Neurol*. 2019;15(9):501-518. DOI 10.1038/s41582-019-0228-7.

Yan L., Cogan J.D., Hedges L.K., Nunley B., Hamid R., Austin E.D. The Y chromosome regulates BMPR2 expression via SRY: A possible reason "Why" fewer males develop pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med*. 2018;198(12):1581-1583. DOI 10.1164/rccm.201802-0308LE.

Zhang L., Shin E.S., Yu Z., Li S. Molecular genetic evidence of Y chromosome loss in male patients with hematological disorders. *Chin. Med. J*. 2007;120(22):2002-2005.

Zhou W., Machiela M.J., Freedman N.D., Rothman N., Malats N., Dagnall C., Caporaso N., Teras L.T., Gaudet M.M., Gapstur S.M., Stevens V.L., Jacobs K.B., Sampson J., Albanes D., Weinstein S., Virtamo J., Berndt S., Hoover R.N., Black A., Silverman D., Figueroa J., Garcia-Closas M., Real F.X., Earl J., Marenne G., Rodriguez-Santiago B., Karagas M., Johnson A., Schwenn M., Wu X., Gu J., Ye Y., Hutchinson A., Tucker M., Perez-Jurado L.A., Dean M., Yeager M., Chanock S.J. Mosaic loss of chromosome Y is associated with common variation near *TCL1A*. *Nat. Genet*. 2016;48(5):563-568. DOI 10.1038/ng.3545.

ORCID ID

I.L. Kuznetsova orcid.org/0000-0002-7785-734X
L.I. Uralsky orcid.org/0000-0002-5565-7961
T.V. Tyazhelova orcid.org/0000-0002-0594-0654
E.I. Rogaev orcid.org/0000-0003-0594-4767

Acknowledgements. The study was supported by Russian Science Foundation, Research Project No. 19-75-30039.

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

Received December 13, 2022. Revised February 14, 2023. Accepted February 14, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):512-521
DOI 10.18699/VJGB-23-62

Original Russian text https://vavilovj-icg.ru/

Role of PI3K/AKT/mTOR signaling pathway and sirtuin genes in chronic obstructive pulmonary disease development

G.F. Korytina[1, 2] ✉, L.Z. Akhmadishina[1], V.A. Markelov[1, 2], Y.G. Aznabaeva[2], O.V. Kochetova[1], T.R. Nasibullin[1], A.P. Larkina[1], N.N. Khusnutdinova[1], N.Sh. Zagidullin[2], T.V. Victorova[2]

[1] Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
[2] Bashkir State Medical University, Ufa, Russia
✉ guly_kory@mail.ru

Abstract. Chronic obstructive pulmonary disease (COPD) is a multifactorial disease of the respiratory system which develops as a result of a complex interaction of genetic and environmental factors closely related to lifestyle. We aimed to assess the combined effect of the PI3K/AKT/mTOR signaling pathway (*PIK3R1*, *AKT1*, *MTOR*, *PTEN*) and sirtuin (*SIRT1*, *SIRT3*, *SIRT6*) genes to COPD risk. SNPs of *SIRT1* (rs3758391, rs3818292), *SIRT3* (rs3782116, rs536715), *SIRT6* (rs107251), *AKT1* (rs2494732), *PIK3R1* (rs10515070, rs831125, rs3730089), *MTOR* (rs2295080, rs2536), *PTEN* (rs701848, rs2735343) genes were genotyped by real-time polymerase chain reaction (PCR) among 1245 case and control samples. Logistic regression was used to detect the association of SNPs in different models. Linear regression analyses were performed to estimate the relationship between SNPs and lung function parameters and smoking pack-years. Significant associations with COPD were identified for *SIRT1* (rs3818292) ($P = 0.001$, OR = 1.51 for AG), *SIRT3* (rs3782116) ($P = 0.0055$, OR = 0.69) and *SIRT3* (rs536715) ($P = 0.00001$, OR = 0.50) under the dominant model, *SIRT6* (rs107251) ($P = 0.00001$, OR = 0.55 for CT), *PIK3R1*: (rs10515070 ($P = 0.0023$, OR = 1.47 for AT), rs831125 ($P = 0.00001$, OR = 2.28 for AG), rs3730089 ($P = 0.0007$, OR = 1.73 for GG)), *PTEN*: (rs701848 ($P = 0.0015$, OR = 1.35 under the log-additive model), and rs2735343 ($P = 0.0001$, OR = 1.64 for GC)). A significant genotype-dependent variation of lung function parameters was observed for *SIRT1* (rs3818292), *SIRT3* (rs3782116), *PIK3R1* (rs3730089), and *MTOR* (rs2536). Gene-gene combinations that remained significantly associated with COPD were obtained; the highest risk of COPD was conferred by a combination of G allele of the *PIK3R1* (rs831125) gene and GG of *SIRT3* (rs536715) (OR = 3.45). The obtained results of polygenic analysis indicate the interaction of genes encoding sirtuins *SIRT3*, *SIRT2*, *SIRT6* and *PI3KR1*, *PTEN*, *MTOR* and confirm the functional relationship between sirtuins and the PI3K/AKT/mTOR signaling pathway.
Key words: chronic obstructive pulmonary disease; PI3K/AKT/mTOR signaling pathway; sirtuins; cellular senescence; oxidative stress.

For citation: Korytina G.F., Akhmadishina L.Z., Markelov V.A., Aznabaeva Y.G., Kochetova O.V., Nasibullin T.R., Larkina A.P., Khusnutdinova N.N., Zagidullin N.Sh., Victorova T.V. Role of PI3K/AKT/mTOR signaling pathway and sirtuin genes in chronic obstructive pulmonary disease development. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):512-521. DOI 10.18699/VJGB-23-62

Роль генов PI3K/AKT/mTOR-сигнального каскада и сиртуинов в развитии хронической обструктивной болезни легких

Г.Ф. Корытина[1, 2] ✉, Л.З. Ахмадишина[1], В.А. Маркелов[1, 2], Ю.Г. Азнабаева[2], О.В. Кочетова[1], Т.Р. Насибуллин[1], А.П. Ларкина[1], Н.Н. Хуснутдинова[1], Н.Ш. Загидуллин[2], Т.В. Викторова[2]

[1] Институт биохимии и генетики – обособленное структурное подразделение Уфимского федерального исследовательского центра Российской академии наук, Уфа, Россия
[2] Башкирский государственный медицинский университет, Уфа, Россия
✉ guly_kory@mail.ru

Аннотация. Хроническая обструктивная болезнь легких (ХОБЛ) – многофакторное заболевание дыхательной системы, развивающееся в результате комплексного взаимодействия молекулярно-генетических и средовых факторов, тесно связанных с образом жизни. Цель исследования – анализ комбинированного вклада генов PI3K/AKT/mTOR-сигнального каскада (*PIK3R1*, *AKT1*, *MTOR*, *PTEN*) и сиртуинов (*SIRT1*, *SIRT3*, *SIRT6*) в риск развития хронической обструктивной болезни легких. В работе использованы образцы ДНК 1245 индивидов. Полиморфные варианты генов *SIRT1* (rs3758391, rs3818292), *SIRT3* (rs3782116, rs536715), *SIRT6* (rs107251), *AKT1* (rs2494732), *PIK3R1* (rs10515070, rs831125, rs3730089), *MTOR* (rs2295080, rs2536), *PTEN* (rs701848, rs2735343) анализировали методом полимеразной цепной реакции в реальном времени. Логистическую ре-

грессию использовали для выявления ассоциации полиморфных локусов в различных моделях. Проводили линейный регрессионный анализ для оценки вклада генотипов изучаемых локусов в вариабельность показателей функции внешнего дыхания и индекса курения. Установлена ассоциация генов: *SIRT1* (rs3818292) (*P* = 0.001, OR = 1.51 для генотипа AG), *SIRT3* (rs3782116) (*P* = 0.0055, OR = 0.69) и *SIRT3* (rs536715) (*P* = 0.00001, OR = 0.50) в доминантной модели; *SIRT6* (rs107251) (*P* = 0.00001, OR = 0.55 для генотипа CT), *PIK3R1*: (rs10515070 (*P* = 0.0023, OR = 1.47 для генотипа AT), rs831125 (*P* = 0.00001, OR = 2.28 для генотипа AG), rs3730089 (*P* = 0.0007, OR = 1.73 для генотипа GG)) и *PTEN*: (rs701848 (*P* = 0.0015, OR = 1.35 в лог-аддитивной модели) и rs2735343 (*P* = 0.0001, OR = 1.64 для генотипа GC)). Обнаружена вариабельность показателей функции легочного дыхания в зависимости от полиморфных вариантов генов *SIRT1* (rs3818292), *SIRT3* (rs3782116), *PIK3R1* (rs3730089) и *MTOR* (rs2536). Идентифицированы ген-генные сочетания, ассоциированные с ХОБЛ; наибольший риск развития ХОБЛ определялся сочетанием аллеля G гена *PIK3R1* (rs831125) с генотипом GG гена *SIRT3* (rs536715) (OR = 3.45). Полученные результаты полигенного анализа указывают на взаимодействие генов, кодирующих сиртуины *SIRT3*, *SIRT2*, *SIRT6* и *PI3KR1*, *PTEN*, *MTOR*, и находят подтверждение в функциональной взаимосвязи сиртуинов и PI3K/AKT/mTOR-сигнального каскада.
Ключевые слова: хроническая обструктивная болезнь легких; PI3K/AKT/mTOR-сигнальный каскад; сиртуины; клеточное старение; окислительный стресс.

Introduction

Chronic obstructive pulmonary disease (COPD) is a multifactorial respiratory system disease that affects the distal parts of the respiratory tract (bronchi, bronchioles) and lung parenchyma with lung emphysema development (Chuchalin et al., 2022). Chronic obstructive pulmonary disease develops as a result of complex interaction between molecular genetic and environmental factors, closely related to lifestyle, and smoking is considered the main cause of COPD (Ragland et al., 2019). Published data suggest that the COPD pathogenesis may involve dysregulation of stress responses that inhibit cellular senescence, which includes a wide range of signaling cascades and their regulators (Ryter et al., 2018).

The PI3K/AKT/mTOR intracellular signaling pathway is a universal pathway controlling cell growth, metabolism, and proliferation (Ersahin et al., 2015). The key components of this signaling pathway are phosphatidylinositol-3 kinase (PI3K), serine/threonine protein kinase (AKT), and serine/threonine kinase (mammalian target of rapamycin, mTOR) (Ersahin et al., 2015). Signal transduction through the PI3K/AKT/mTOR signaling cascade is essential for cellular senescence. This signaling pathway is inhibited by the tyrosine phosphatases PTEN (phosphatase and tensin homolog) and SHIP-1 (inositol polyphosphate-5-phosphatase D). Both enzymes have oxidation-sensitive cysteine residue in the active region (Worby, Dixon, 2014).

Oxidative stress is the main mechanism that causes accelerated senescence through its damaging effects on DNA and the PI3K/AKT/mTOR signaling pathway activation (Wang et al., 2013). In COPD and other age-associated diseases, the expression of genes encoding endogenous antioxidant molecules is reduced, which leads to an increased level of oxidative stress and cellular senescence activation (Kirkham, Barnes, 2013). NAD-dependent protein deacetylases from the sirtuins family are considered as potential factors that decrease senescence (Ito, Barnes, 2009).

We aimed to assess the combined effect of the PI3K/AKT/mTOR signaling pathway (*PIK3R1*, *AKT1*, *MTOR*, *PTEN*) and sirtuins (*SIRT1*, *SIRT3*, *SIRT6*) genes on COPD risk.

Materials and methods

DNA samples were collected from unrelated subjects who were Tatars in ethnicity and resided in the Republic of Bashkortostan. The study was approved by the Ethics Committee at the Institute of Biochemistry and Genetics (Protocol No 17, December 7, 2010). All participants of this study provided written informed consent. The COPD group included 621 patients (539 (86.79 %) males and 82 (13.21 %) females) with a mean age of 64.42 ± 10.71 years. There were 510 (82.13 %) smokers and former smokers and 111 (17.87 %) nonsmokers in the COPD group. The smoking index was 45.34 ± 23.84 pack years in the smokers and former smokers. The control group included 624 subjects (555 (88.94 %) males and 69 (11.06 %) females) with a mean age of 59.67 ± 12.31. There were 526 (84.29 %) smokers and former smokers and 98 (15.71 %) nonsmokers in the group; the smoking index was 38.75 ± 24.87 pack years in the smokers.

In all patients, spirometry was performed to assess lung function, including vital capacity (VC), forced vital capacity (FVC), forced expiration volume in the first second (FEV1), and the FEV1/FVC ratio. The group of patients had the following parameters (% of normal levels): FEV1 = 40.75 ± 18.33; FVC = 45.01 ± 18.22; VC = 49.32 ± 14.34; FEV1/FVC = 59.5 ± 12.34. Inclusion and exclusion criteria for the COPD and control have been previously described (Korytina et al., 2019).

Genotyping. DNA was isolated from peripheral blood leukocytes by phenol-chloroform extraction. The set included SNPs of the following genes: *SIRT1* (rs3758391, rs3818292), *SIRT3* (rs3782116, rs536715), *SIRT6* (rs107251), *AKT1* (rs2494732), *PIK3R1* (rs10515070, rs831125, rs3730089), *MTOR* (rs2295080, rs2536), *PTEN* (rs701848, rs2735343) (Supplementary Material 1)[1].

The SNPs were selected for the study based on the following criteria: functional effect of SNP on gene expression or relation to non-synonymous substitutions, and/or associations with complex human diseases; minor allele frequency

[1] Supplementary Materials 1 and 2 are available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/Suppl_Korytina_Engl_27_5.pdf

(MAF) of ≥ 5 % in the European populations according to the National Center for Biotechnology Information database (http://www.ncbi.nlm.nih.gov/projects/SNP/). The functional significance of the SNPs was verified using RegulomeDB Version 1.1 (https://regulomedb.org), SNPinfo Web Server (https://snpinfo.niehs.nih.gov), and HaploReg v3 (Ward, Kellis, 2016). Data were presented in Supplementary Material 2. SNP genotyping was performed by real-time polymerase chain reaction (PCR) using commercial kits for fluorescence detection (https://www.oligos.ru, DNK-Sintez, Russia) and a BioRad CFX96™ instrument (Bio-Rad Laboratories, United States). The methods of analysis were described in detail previously (Korytina et al., 2019).

Statistical analyses. Statistical analyses of the results were performed using the software packages IBM SPSS Statistics 22.0 and PLINK v. 1.07 (Purcell et al., 2007). The methods were described in detail previously (Korytina et al., 2019). Association analyses of allele or genotype combinations with COPD were carried out using the APSampler 3.6.1 program (http://sourceforge.net/projects/apsampler/). The Benjiamini–Hochberg correction for multiple testing was performed using special software (http://www.sdmproject.com/utilinies/?show=FDR) to decrease the false discovery rate (FDR) and to obtain $P_{cor\text{-}FDR}$. The linkage disequilibrium structure LD (D′) and haplotype frequencies were calculated with Haploview 4.2.

Results

Before analyzing the association of candidate gene alleles with COPD, we verified whether the genotype frequency distributions corresponded to the Hardy–Weinberg equilibrium and evaluated minor allele frequencies (MAF) both in the combined group of patients and healthy subjects and in either group individually (see Supplementary Material 1). All studied SNPs were in Hardy–Weinberg equilibrium in the control group: *SIRT1* (rs3758391) ($P_{X\text{-}B}$ = 0.24), *SIRT1* (rs3818292) ($P_{X\text{-}B}$ = 0.47), *SIRT3* (rs3782116) ($P_{X\text{-}B}$ = 0.5), *SIRT3* (rs536715) ($P_{X\text{-}B}$ = 0.75), *SIRT6* (rs107251) ($P_{X\text{-}B}$ = 0.67), *AKT1* (rs2494732) ($P_{X\text{-}B}$ = 0.2), *PIK3R1* (rs10515070) ($P_{X\text{-}B}$ = 0.65), *PIK3R1* (rs831125) ($P_{X\text{-}B}$ = 0.25), *PIK3R1* (rs3730089) ($P_{X\text{-}B}$ = 0.22), *MTOR* (rs2295080) ($P_{X\text{-}B}$ = 0.15), *MTOR* (rs2536) ($P_{X\text{-}B}$ = 0.24), *PTEN* (rs701848) ($P_{X\text{-}B}$ = 0.85), *PTEN* (rs2735343) ($P_{X\text{-}B}$ = 0.06).

The groups of COPD patients and healthy controls differed significantly in the genotypes and/or alleles frequency distributions of *SIRT1* (rs3818292), *SIRT3* (rs3782116, rs536715), *SIRT6* (rs107251), *AKT1* (rs2494732), *PIK3R1* (rs10515070, rs831125, rs3730089), and *PTEN* (rs701848, rs2735343) (Table 1). Statistically significant results of association analysis of the studied gene loci and COPD are shown in Table 2.

An association of *SIRT1* (rs3818292) with COPD was established in the dominant model (P_{adj} = 0.0066, OR = 1.40). The risk of COPD was increased in heterozygous individuals (P_{adj} = 0.001, OR = 1.51). An association of COPD with the heterozygous genotype (P_{adj} = 0.0052, OR = 0.69) and in the dominant model of *SIRT3* (rs3782116) and in the dominant (P_{adj} = 0.00001, OR = 0.50), log-additive (P_{adj} = 0.00001, OR = 0.66) models and with the heterozygous genotype (P_{adj} = 0.00001, OR = 0.48) of *SIRT3* (rs536715). It should be noted that in both cases the COPD risk was associated with the frequent G allele (rs3782116 – OR = 1.21 95 % CI 1.03–1.43 и rs536715 – OR = 1.58 95 % CI 1.32–1.91) and the GG genotype (rs3782116 – OR = 1.44 95 % CI 1.16–1.81 и rs536715 – OR = 1.99 95 % CI 1.58–2.51).

We carried out a linkage disequilibrium analysis of the rs3758391 and rs3818292 loci of the *SIRT1*, rs3782116 and rs536715 of the *SIRT3*, which showed the absence of linkage disequilibrium between the loci of the *SIRT1* gene (D′ = 0.168, r^2 = 0.097) and the *SIRT3* gene (D′ = 0.28, r^2 = 0.011). Based on the obtained data, haplotypes association analysis was not performed. Association of *SIRT6* (rs107251) with the development of COPD was detected in the dominant model (P_{adj} = 0.0005, OR = 0.65), but the association with the heterozygous CT genotype was more significant (P_{adj} = 0.00001, OR = 0.55). The risk of COPD was associated with the CC genotype of *SIRT6* (rs107251) (OR = 1.54 95 % CI 1.23–1.93).

We have identified the association SNPs of the *PIK3R1* gene (rs10515070, rs831125, rs3730089) with COPD. The risk of COPD was associated with heterozygous genotypes of the studied SNPs of the *PIK3R1* gene: rs10515070 (P_{adj} = 0.0023, OR = 1.47), rs831125 (P_{adj} = 0.00001, OR = 2.28) and with the GG genotype of rs3730089 (P_{adj} = 0.0007, OR = 1.73). We showed the absence of linkage disequilibrium between rs10515070, rs831125, rs3730089 of the *PIK3R1* gene: for rs10515070 and rs831125 (D′ = 0.02, r^2 = 0.00), for rs10515070 and rs3730089 (D′ = 0.127, r^2 = 0.008), for rs831125 and rs3730089 (D′ = 0.155, r^2 = 0.005).

Based on the obtained data, haplotypes association analysis was not performed. Association of *PTEN* (rs701848) and COPD was established in the dominant (P_{adj} = 0.0035, OR = 1.52), recessive (P_{adj} = 0.028, OR = 1.44) and log-additive models P_{adj} = 0.0015, OR = 1.35). We have identified the association of *PTEN* (rs2735343) with COPD in the log-dominant model (P_{adj} = 0.01, OR = 1.42) and for the heterozygous genotype (P_{adj} = 0.0001, OR = 1.64). Linkage disequilibrium between the rs701848 and rs2735343 was not detected (D′ = 0.234, r^2 = 0.035), thus, haplotype association analysis was not performed.

Association of the studied genes loci and quantitative phenotypes with lung function parameters and smoking index

Lung function decline is a key clinical feature of airway obstruction in COPD that indicates progression of the disease. We investigated the relationship between the studied genes loci and lung function parameters: Forced Vital Capacity (FVC), Forced Expiration Volume in 1 s (FEV1), and FEV1/FVC ratio in COPD patients (Table 3). The heterozygous genotype of *PIK3R1* (rs3730089) (P = 0.013) and the TT genotype of *MTOR* (rs2536) (P = 0.013) were associated with a decrease in the FVC value. Carriers of the *SIRT3* (rs3782116) AA genotype exhibited a higher FVC value (P = 0.0015).

Individuals that presented the AA genotype of *SIRT1* (rs3818292) (P = 0.017), the GG genotype of *PIK3R1* (rs3730089) (P = 0.025), and the AA genotype of *SIRT3* (rs3782116) (P = 0.0028) showed a significant increase in their FVC. Meanwhile, carriers of the heterozygous genotypes of *SIRT1* (rs3818292) (P = 0.04), *MTOR* (rs2295080) (P = 0.025), and *SIRT3* (rs3782116) (P = 0.016) exhibited a lower FVC

Table 1. Genotypes and alleles distribution of the studied PI3K/AKT/mTOR signaling pathway and sirtuins genes loci in COPD and control

Gene, SNP	Rare allele	Genotypes and alleles	COPD, *n* (%) (*N* = 621)	Control, *n* (%) (*N* = 624)	*P*
SIRT1 rs3758391 T>C	C	TT/TC/CC	184/265/172 (29.63/42.67/27.70)	168/294/162 (26.92/47.12/25.96)	0.283
		T/C	633/609 (50.97/49.03)	630/618 (50.48/49.52)	0.84
SIRT1 rs3818292 A>G	G	AA/AG/GG	322/273/26 (51.85/43.96/4.19)	375/213/36 (60.10/34.13/5.77)	0.001
		A/G	917/325 (73.83/26.17)	963/285 (77.16/22.84)	0.059
SIRT3 rs3782116 G>A	A	GG/GA/AA	294/230/97 (47.34/37.04/15.62)	239/287/98 (38.30/45.99/15.71)	0.003
		G/A	818/424 (65.86/34.14)	765/483 (61.30/38.70)	0.02
SIRT3 rs536715 G>A	A	GG/GA/AA	424/148/49 (68.28/23.83/7.89)	324/249/51 (51.92/39.90/8.17)	0.00001
		G/A	996/246 (80.19/19.81)	897/351 (71.88/28.13)	0.00001
SIRT6 rs107251 C>T	T	CC/CT/TT	396/160/65 (63.77/25.76/10.47)	333/243/48 (53.37/38.94/7.69)	0.0001
		C/T	952/290 (76.65/23.35)	909/339 (72.84/27.16)	0.032
AKT1 rs2494732 T>C	C	TT/CT/CC	193/313/115 (31.08/50.40/18.52)	159/329/136 (25.48/52.72/21.79)	0.062
		T/C	699/543 (56.28/43.72)	647/601 (51.84/48.16)	0.029
PIK3R1 rs10515070 T>A	A	TT/TA/AA	175/351/95 (28.18/56.52/15.30)	228/293/103 (36.54/46.96/16.51)	0.001
		T/A	701/541 (56.44/43.56)	749/499 (60.02/39.98)	0.077
PIK3R1 rs831125 A>G	G	AA/AG/GG	299/274/48 (48.15/44.12/7.73)	441/161/22 (70.67/25.80/3.53)	0.00001
		A/G	872/370 (70.21/29.79)	1 043/205 (83.57/16.43)	0.00001
PIK3R1 rs3730089 G>A	G	AA/AG/GG	230/257/134 (37.04/41.38/21.58)	270/269/85 (43.27/43.11/13.62)	0.0001
		A/G	717/525 (57.73/42.27)	809/439 (64.82/35.18)	0.0001
MTOR rs2295080 T>G	G	TT/TG/GG	220/304/97 (35.43/48.95/15.62)	218/318/88 (34.94/50.96/14.10)	0.686
		T/G	744/498 (59.90/40.10)	754/494 (60.42/39.58)	0.825
MTOR rs2536 T>C	C	TT/TC/CC	580/41/0 (93.40/6.60/0)	567/55/2 (90.87/8.81/0.32)	0.124
		T/C	1 201/41 (96.70/3.30)	1 189/59 (95.27/4.73)	0.087
PTEN rs701848 T>C	C	TT/CT/CC	156/324/141 (25.12/52.17/22.71)	212/306/106 (33.97/49.04/16.99)	0.0001
		T/C	636/606 (51.21/48.79)	730/518 (58.49/41.51)	0.0001
PTEN rs2735343 G>C	C	GG/GC/CC	169/355/97 (27.21/57.17/15.62)	217/280/127 (34.78/44.87/20.35)	0.0001
		G/C	693/549 (55.80/44.20)	714/534 (57.21/42.79)	0.52

Note. *P* is the significance of group differences in allele and genotype frequencies (sample χ^2 homogeneity test).

Table 2. Significant association of the studied PI3K/AKT/mTOR signaling pathway and sirtuins genes loci with COPD

Gene, SNP	Rare allele	*N*	Genotype/model	OR_{adj} (CI 95 %)	P_{adj}	$P_{cor\text{-}FDR}$
SIRT1 rs3818292 A>G	G	1245	AA AG+GG dominant	1.00 1.40 (1.10–1.78)	0.0066	0.0093
			AA+GG AG	1.00 1.51 (1.18–1.94)	0.001	0.0024
			Log-additive	1.21 (0.99–1.49)	0.063	0.068
			G	1.19 (0.99–1.44)	0.059	0.062
SIRT3 rs3782116 G>A	A	1245	GG GA+AA dominant	1.00 0.69 (0.53–0.90)	0.0055	0.008
			AA+GG AG	1.00 0.69 (0.53–1.90)	0.0052	0.008
			Log-additive	0.84 (0.7–1.00)	0.055	0.061
			A	0.82 (0.69–0.96)	0.02	0.03
SIRT3 rs536715 G>A	A	1245	GG GA+AA dominant	1.00 0.50 (0.39–0.65)	0.00001	3.85e-05
			AA+GG AG	1.00 0.48 (0.36–0.62)	0.00001	3.85e-05
			Log-additive	0.66 (0.55–0.80)	0.00001	3.85e-05
			A	0.63 (0.52–0.76)	0.00001	3.85e-05
SIRT6 rs107251 C>T	T	1245	CC CT+TT dominant	1.00 0.65 (0.51–0.83)	0.0005	0.0015
			CC+TT CT	1.00 0.55 (0.42–0.71)	0.00001	3.85e-05
			Log-additive	0.84 (0.70–1.00)	0.051	0.059
			T	0.82 (0.68–0.97)	0.032	0.039
PIK3R1 rs10515070 T>A	A	1245	TT TA+AA dominant	1.00 1.47 (1.13–1.92)	0.0042	0.007
			TT+AA TA	1.00 1.47 (1.15–1.88)	0.0023	0.0044
			Log-additive	1.17 (0.98–1.41)	0.087	0.09
			A	1.15 (0.98–1.36)	0077	0.088
PIK3R1 rs831125 A>G	G	1245	AA AG+GG dominant	1.00 2.61 (1.99–3.42)	0.00001	3.85e-05
			AA+GG AG	1.00 2.28 (1.73–3.00)	0.00001	3.85e-05
			Log-additive	2.19 (1.74–2.75)	0.00001	3.85e-05
			G	2.16 (1.78–2.62)	0.00001	3.85e-05
PIK3R1 rs3730089 G>A	G	1245	AA AG+GG dominant	1.00 1.30 (1.02–1.66)	0.037	0.045
			AA+AG GG recessive	1.00 1.73 (1.26–2.39)	0.0007	0.0018
			Log-additive G	1.31 (1.11–1.55) 1.35 (1.15–1.58)	0.0014 0.0001	0.003 0.00033
PTEN rs701848 T>C	C	1245	TT TC+CC dominant	1.00 1.52 (1.15–2.02)	0.0035	0.006
			TT+TC CC recessive	1.00 1.44 (1.04–1.99)	0.028	0.036
			Log-additive	1.35 (1.12–1.63)	0.0015	0.003
			C	1.34 (1.14–1.57)	0.0001	0.00033
PTEN rs2735343 G>C	C	1245	GG GC+CC dominant	1.00 1.42 (1.09–1.87)	0.01	0.013
			GG+CC GC	1.00 1.64 (1.28–2.12)	0.0001	0.00033
			Log-additive	1.06 (0.88–1.27)	0.53	0.53
			C	1.06 (0.9–1.24)	0.52	0.522

Note. N is the number of individuals included in the analysis; P_{adj} – significance in the likelihood ratio test for the regression model adjusted for age, sex, smoking status and pack-years; OR_{adj} – adjusted odds ratio and CI 95 % – confidence interval; $P_{cor\text{-}FDR}$ – significance after the FDR correction; in the log-additive model per rare allele dosage, the rare allele dosage increases in the following order: homozygote for the common allele (0) – heterozygote, (1) – homozygote for the rare allele (2).

Table 3. Association of the PI3K/AKT/mTOR signaling pathway and sirtuins gene loci with lung function parameters and smoking index

Gene, SNP	Genotypes	M ± SE	*P*	beta (CI 95 %)
FVC				
MTOR rs2536 T>C	TT CT	53.76 (1.03) 64.79 (4.02)	0.013	0.00 11.03 (2.38–19.67)
PIK3R1 rs3730089 G>A	AA+GG AG	56.5 (1.31) 51.4 (1.51)	0.013	0.00 –5.10 (–9.09…–1.11)
SIRT3 rs3782116 G>A	GG+AG AA	53.31 (1.1) 62.59 (3.06)	0.0015	0.00 9.28 (3.61–14.95)
FEV1				
SIRT1 rs3818292 A>G	AA AG+GG	56.06 (1.45) 51.06 (1.51)	0.017	0.00 –5.00 (–9.10…–0.90)
	AA+GG AG	55.52 (1.39) 51.35 (1.59)	0.04	0.00 –4.17 (–8.31…–0.03)
MTOR rs2295080 T>G	TT+GG GT	57.42 (1.61) 52.09 (1.75)	0.025	0.00 –5.34 (–9.99…–0.68)
PIK3R1 rs3730089 G>A	AA+AG GG	53.15 (1.13) 61.18 (2.57)	0.0026	0.00 8.03 (2.83–13.23)
SIRT3 rs3782116 G>A	GG+AG AA	54.6 (1.2) 63.95 (3.23)	0.0028	0.00 9.36 (3.27–5.45)
	GG+AA AG	58.28 (1.5) 52.61 (1.73)	0.016	0.00 –5.67 (–10.24…–1.10)
Smoking index in pack-years				
PIK3R1 rs10515070 T>A	TT+AA AT	31.61 (1.3) 31.00 (1.11)	0.025	0.00 –3.30 (–6.17…–0.42)
PIK3R1 rs831125 A>G	AA+GG AG	30.85 (1.08) 32.48 (1.61)	0.0082	0.00 –4.37 (–7.60…–1.14)
SIRT1 rs3818292 A>G	AA AG+GG	32.00 (1.13) 31.03 (1.25)	0.036	0.00 –3.04 (–5.88…–0.20)

Note. M ± SE is the mean ± standard error of the mean; *P* is the significance level for the regression equation; beta (95 % CI) is the regression coefficient (95 % confidence interval of the coefficient).

value (see Table 3). The carriers of the heterozygous genotype of *PIK3R1 (*rs831125) (*P* = 0.0082), the AA genotype of *SIRT1* (rs3818292) (*P* = 0.036) had a significantly higher smoking index.

Analysis of gene-gene interactions

Using the APSampler algorithm, we have identified gene-gene combinations significantly associated with chronic obstructive pulmonary disease. In order to identify significant interactions of functionally related sirtuins genes, *SIRT2* (rs10410544) was included in the analysis (Korytina et al., 2019). We obtained 2324 patterns associated with chronic obstructive pulmonary disease. Table 4 shows the results of the most significant gene-gene combinations with P_{FDR} less than 0.05 and OR more than 2.0 for risk combinations) or less than 0.35 for protective combinations. A total of 19 gene-gene combinations fulfilled the above-mentioned criteria. Nine patterns were associated with an increased risk of COPD; ten were protective. Allele G and genotype GG of *PIK3R1* (rs831125) contributed to the most significant combinations associated with COPD risk (four patters).

The highest risk of COPD was conferred by the combination of these variants of *PIK3R1* (rs831125) with the GG genotype of *SIRT3* (rs536715) (OR = 3.45); and with the C allele of *PTEN* (rs2735343) (OR = 3.06) and their combination: genotype GA of *PIK3R1* (rs831125) together with the G allele of *SIRT3* (rs536715) and the C allele of *PTEN* (rs2735343) (OR = 2.86). The analysis on the gene-gene interaction of the studied gene loci established an association of the T allele of *MTOR* (rs2536) only in combinations with the G allele of *PIK3R1* (rs831125) (OR = 2.71). Three of the identified patterns included the G allele of *PIK3R1* (rs3730089) in

Table 4. Gene-gene combinations of the PI3K/AKT/mTOR signaling pathway and sirtuins gene loci associated with COPD

Pattern	COPD patients	Healthy individuals	*P*	P_{FDR}	OR	CI (95 %)
Risk patterns						
PIK3R1 (rs831125) G + *SIRT3* (rs536715) GG	0.36	0.12	1.06e-14	2.45e-11	3.45	2.49–4.78
PIK3R1 (rs831125) G + *MTOR* (rs2536) T	0.53	0.29	7.65e-13	1.77e-09	2.71	2.05–3.57
PIK3R1 (rs831125) G + *PTEN* (rs2735343) C	0.39	0.17	7.7e-13	1.79e-09	3.06	2.23–4.18
PIK3R1 (rs831125) GA + *SIRT3* (rs536715) G + *PTEN* (rs2735343) C	0.31	0.14	9.07e-10	2.10e-06	2.86	2.02–4.04
SIRT3 (rs536715) GG + *PTEN* (rs2735343) CG	0.61	0.42	2.98e-09	6.92e-06	2.17	1.67–2.83
PIK3R1 (rs3730089) G + *SIRT3* (rs536715) GG	0.46	0.29	1.14e-08	2.64e-05	2.09	1.62–2.72
PIK3R1 (rs3730089) G + *SIRT6* (rs107251) CC	0.45	0.29	4.13e-08	9.60e-05	2.01	1.56–2.59
PIK3R1 (rs10515070) A + *PTEN* (rs701848) C + *PTEN* (rs2735343) C	0.39	0.23	1.11e-06	0.0025	2.10	1.54–2.85
Protective patterns						
PIK3R1 (rs831125) AA + *PIK3R1* (rs3730089) A	0.34	0.61	1.11e-16	2.58e-13	0.32	0.24–0.42
PIK3R1 (rs831125) A + *PIK3R1* (rs3730089) A + *SIRT3* (rs3782116) A + *SIRT3* (rs536715) A	0.03	0.22	2.7e-15	6.27e-12	0.13	0.07–0.24
PIK3R1 (rs831125) A + *SIRT3* (rs3782116) A + *SIRT3* (rs536715) A	0.06	0.25	1.07e-14	2.49e-11	0.18	0.11–0.29
PIK3R1 (rs831125) A + *PIK3R1* (rs3730089) A + *SIRT3* (rs3782116) AG + *SIRT3* (rs536715) A	0.02	0.17	3.82e-14	8.87e-11	0.09	0.04–0.20
PIK3R1 (rs3730089) A + *SIRT3* (rs3782116) A + *SIRT3* (rs536715) A	0.06	0.22	2.3e-13	5.36e-10	0.21	0.13–0.32
SIRT3 (rs3782116) A + *SIRT3* (rs536715) A	0.08	0.26	6.52e-13	1.52e-09	0.25	0.17–0.38
SIRT3 (rs3782116) AG + *SIRT3* (rs536715) GA	0.04	0.18	4.0e-11	9.29e-08	0.20	0.12–0.35
SIRT3 (rs536715) A + *SIRT6* (rs107251) T	0.10	0.25	8.12e-11	1.89e-07	0.33	0.23–0.47
SIRT3 (rs536715) A + *PTEN* (rs2735343) GG	0.06	0.18	4.73e-09	1.09e-05	0.27	0.17–0.44
SIRT3 (rs536715) GA + *SIRT6* (rs107251) TC	0.06	0.17	9.75e-09	2.26e-05	0.30	0.19–0.47

Notes. *P*-value is the significance level for Fisher's test, P_{FDR} – is the FDR value after FDR correction; OR is odds ratio, 95 % CI is the 95 % confidence interval for the OR.

combination with the GG genotype of *SIRT3* (rs536715) (OR = 2.09), or with the CC genotype of *SIRT6* (rs107251) (OR = 2.01).

The most significant protective patterns included the A allele or the AA genotype of *PIK3R1* (rs831125) and the A allele of *PIK3R1* (rs3730089) in combination with the A allele or the AG genotype of *SIRT3* (rs3782116) and the A allele of *SIRT3* (rs536715) (see Table 4). Thus, the *PIK3R1* (rs831125), *PIK3R1* (rs3730089), and *SIRT3* (rs536715) loci exhibited an allele-specific effect, when the G allele of *PIK3R1* (rs831125), the G allele of *PIK3R1* (rs3730089), and the G allele and the GG genotype of *SIRT3* (rs536715) were part of risk patterns, and alternative alleles of the same polymorphic loci were present in protective patterns.

Discussion

As a result our study, significant associations between the *SIRT1* (rs3818292), *SIRT3* (rs3782116, rs536715), *SIRT6* (rs107251) polymorphic variants and COPD were found. SIRT1 is the most studied member of the mammalian sirtuin family. It has been shown that SIRT1 plays an important role in signaling pathways involved in cellular senescence and cell death (Finkel et al., 2009). SIRT1 deacetylates many key regulatory proteins and transcription factors involved in DNA repair, inflammation, expression of antioxidant genes, and cellular senescence, including the PI3K/AKT/mTOR signaling pathway genes, transcription factor FOXO3a, p21, p16, Klotho proteins (Cao et al., 2013).

Increased expression of SIRT1 inhibits the TGF-β1/SMAD3 signaling pathway and impairs epithelial-mesenchymal transformation, which leads to a decrease in COPD-associated airway remodeling (Zhang et al., 2022). SIRT1 levels are reduced in peripheral pulmonary and peripheral blood mononuclear cells of patients with COPD (Rajendrasozhan et al., 2008).

We found that the COPD risk is higher in heterozygous carriers of *SIRT1* (rs3818292). Moreover, this polymorphic variant demonstrates an association with a decrease in FVC1,

which reflects the progression of the disease. Functional analysis showed that *SIRT1* (rs3818292) is in linkage with a SNP in the 5′-untranslated DNA region (rs3740051) that changes the NFKB transcription factor binding site. We did not associate *SIRT1* (rs3758391) with COPD. The results of our study are in agreement with the previously published data reported for the Han Chinese population (Gao et al., 2018). According to functional analysis, rs3758391 is located in the promoter region of the *SIRT1* gene, and the C variant disrupts binding sites for several transcription factors and regulatory proteins, affecting gene expression. The role of rs3758391 in the development of age-associated diseases is well known (Wu et al., 2022).

Mitochondrial dysfunction in respiratory epithelial cells plays an important role in the pathogenesis of COPD (Zhang et al., 2022). SIRT3 is the main mitochondrial deacetylase regulating many enzymes involved in energy metabolism, respiratory chain components, the tricarboxylic acid cycle, ketogenesis, and fatty acid beta-oxidation (Wu et al., 2022).

SIRT3 can directly control the production of reactive oxygen species (ROS) by deacetylating manganese superoxide dismutase (SOD2), the main mitochondrial antioxidant enzyme (Dikalova et al., 2017). SIRT3 is involved in regulating the activity of the DNA repair enzyme OGG1, which leads to increased damage to mtDNA and apoptosis of alveolar epithelial cells (Sun et al., 2018). A number of studies have shown the SIRT3 association with various complex diseases (Wu et al., 2022).

We studied the association between two *SIRT3* gene functional polymorphisms (rs3782116 and rs536715) and chronic obstructive pulmonary disease. Functional analysis showed that rs3782116 is located in the region of binding sites for hsa-miR-328; polymorphic loci rs3782116 and rs536715 are located in DNA regions that bind regulatory proteins. Both polymorphic loci were associated with COPD in our cohort; the disease development risk was associated with the G alleles of rs3782116 and rs536715. It should be noted that the carriers of the homozygous rare allele A of rs3782116 of the *SIRT3* gene had higher values of VC and FVC1. The contribution of *SIRT3* variants to COPD has not been studied previously. At the same time, effects of the *SIRT3* gene SNPs have been fairly extensively investigated in age-associated diseases in which oxidative stress and cellular senescence play a key role (Song et al., 2022).

We obtained significant associations of *SIRT6* (rs107251) with COPD; the frequent C allele is associated with COPD risk, while the heterozygous genotype has a protective effect on the disease development. rs107251 is located in the DNA region that binds to the SOX8 regulatory protein and it is in close linkage with rs350846, localized in the 3-non-translational region of the *SIRT6* gene – a binding site for several miRNAs (hsa-miR-1207-5p, hsa-miR-24, hsa-miR-34a, hsa-miR-644, hsa-miR-940). SIRT6 participates in the regulation of genome stability, NF-kB signaling, glucose homeostasis, exhibits ADP-ribosyltransferase and histone deacetylase activity, plays a role in DNA repair and maintenance of telomeric chromatin integrity (Kugel, Mostoslavsky, 2014). In the study by N. Takasaka et al. (2014), a decrease in SIRT6 levels was shown in respiratory epithelial cells of COPD patients due to cigarette smoke exposure, leading to cellular senescence and disruption of autophagy processes. Association of the *SIRT6* gene SNPs with COPD has not been studied previously, but there is evidence of their association with cardiovascular diseases, which are often comorbid pathologies in COPD (Song et al., 2022).

The *PIK3R1* gene encodes regulatory subunit 1 of phosphoinositide 3-kinase, a key element of the PI3K/AKT/mTOR signaling cascade (Ersahin et al., 2015). We investigated three *PIK3R1* gene polymorphic loci (rs10515070, rs831125, and rs3730089), which showed a significant association with COPD in our studied group. Carriers of rare alleles of these polymorphic loci had a high risk of COPD. In addition, we investigated the relationship between rs3730089 genotypes and VC and FVC1 values; thus, heterozygotes have lower values, and these results are in agreement with association analysis. Functional analysis showed that an intronic variant rs831125 is located in a binding site for regulatory proteins; rs3730089 is a missense variant with a "benign" effect according to the PolyPhen-2 database (http://genetics.bwh.harvard.edu/pph2/), located in a binding site for regulatory proteins and affecting splice sites. SNPs of the *PIK3R1* gene have not been evaluated for COPD before. Previously, an association has been observed between rs3730089 and type 2 diabetes (Karadoğan et al., 2018).

The phosphatase PTEN regulates the activity of phosphoinositide-3 kinase (PI3K) (Worby Dixon, 2014). Smoking as a major risk factor for COPD provokes oxidative stress, which, in turn, affects *PTEN* expression (Cai et al., 2022). We investigated two *PTEN* gene functional polymorphic loci; rs70184 is located in *PTEN* gene 3′ region and changes binding sites for hsa-miR-1252 and hsa-miR-1304 miRNA; rs2735343, located in the intronic region, affects binding sites for several regulatory proteins.

Significant associations with COPD in our sample were found with *PTEN* gene loci; thus, homozygous carriers of the rare C allele of rs701848 and heterozygous carriers of rs2735343 had a significant risk of COPD development.

Published data have demonstrated an association between *PTEN* (rs701848) and COPD; the risk was significantly reduced for homozygous T allele carriers, which is consistent with the data obtained for our sample (Hosgood et al., 2009). PTEN participates in the regulation of various biological processes, including cell proliferation, apoptosis, inflammatory reactions, transcription, and genomic stability (Cai et al., 2022). Decreased levels of PTEN lead to activation of PI3K signaling and increased cell senescence in COPD (Barnes et al., 2019). It has been shown that decreased PTEN activity in COPD increases the activity of matrix metalloprotease MMP9 in bronchial epithelial cells, which consequently contributes to the progression of inflammation and extracellular matrix degradation (Vannitamby et al., 2017).

The analysis of gene-gene interactions revealed significant synergy between polymorphic loci of genes encoding phosphoinositide-3-kinase (PIK3R1) and mitochondrial deacetylase (SIRT3), which were present in most significant combinations associated with an increased risk of chronic obstructive pulmonary disease. The C allele in the *PTEN* (rs2735343) was part of four informative combinations associated with a high risk of chronic obstructive pulmonary disease.

The results of polygenic analysis indicate the interaction of genes encoding sirtuins *SIRT3*, *SIRT2*, *SIRT6* and *PI3KR1*, *PTEN*, *MTOR* and confirm the functional relationship between sirtuins and the PI3K/AKT/mTOR signaling pathway.

Conclusion

The obtained results of single locus and polygenic analysis indicate the contribution of *SIRT3* (rs3782116, rs536715), *SIRT6* (rs107251) and *PIK3R1* (rs10515070, rs831125, rs3730089) polymorphisms to COPD and interaction of genes encoding the key components of the PI3K/AKT/mTOR signalling pathway and sirtuins, and confirm the involvement of cellular senescence mechanisms with COPD development.

References

Barnes P.J., Baker J., Donnelly L.E. Cellular senescence as a mechanism and target in chronic lung diseases. *Am. J. Respir. Crit. Care Med.* 2019;200(5):556-564. DOI 10.1164/rccm.201810-1975TR.

Cai B., Yang L., Do Jung Y., Zhang Y., Liu X., Zhao P., Li J. PTEN: An emerging potential target for therapeutic intervention in respiratory diseases. *Oxid. Med. Cell Longev*. 2022;2022:4512503. DOI 10.1155/2022/4512503.

Cao L., Liu C., Wang F., Wang H. SIRT1 negatively regulates amyloid-beta-induced inflammation via the NF-κB pathway. *Braz. J. Med. Biol. Res.* 2013;46(8):659-669. DOI 10.1590/1414-431X20132903.

Chuchalin A.G., Avdeev S.N., Aisanov Z.R., Belevskiy A.S., Leshchenko I.V., Ovcharenko S.I., Shmelev E.I. Federal guidelines on diagnosis and treatment of chronic obstructive pulmonary disease. *Pul'monologiya = Pulmonology*. 2022;32(3):356-392. DOI 10.18093/0869-0189-2022-32-3-356-392. (in Russian)

Dikalova A.E., Itani H.A., Nazarewicz R.R., McMaster W.G., Flynn C.R., Uzhachenko R., Fessel J.P., Gamboa J.L., Harrison D.G., Dikalov S.I. Sirt3 impairment and SOD2 hyperacetylation in vascular oxidative stress and hypertension. *Circ. Res*. 2017;121(5):564-574. DOI 10.1161/CIRCRESAHA.117.310933.

Ersahin T., Tuncbag N., Cetin-Atalay R. The PI3K/AKT/mTOR interactive pathway. *Mol. Biosyst*. 2015;11(7):1946-1954. DOI 10.1039/c5mb00101c.

Finkel T., Deng C.X., Mostoslavsky R. Recent progress in the biology and physiology of sirtuins. *Nature.* 2009;460(7255):587-591. DOI 10.1038/nature08197.

Gao S.L., Wang Y.H., Li C.Y., Yang L.W., Zou B.A., Chen Z.G., Yao W.M., Song Z.Q., Cheng J.F., Lin Z.Y., Liu G. A highly significant association between Cathepsin S gene polymorphisms rs12068264 and chronic obstructive pulmonary disease susceptibility in Han Chinese population. *Biosci. Rep.* 2018;38(4):BSR20180410. DOI 10.1042/BSR20180410.

Hosgood H.D., Menashe I., He X., Chanock S., Lan Q. *PTEN* identified as important risk factor of chronic obstructive pulmonary disease. *Respir. Med*. 2009;103(12):1866-1870. DOI 10.1016/j.rmed.2009. 06.016.

Ito K., Barnes P.J. COPD as a disease of accelerated lung aging. *Chest.* 2009;135(1):173-180. DOI 10.1378/chest.08-1419.

Karadoğan A.H., Arikoglu H., Göktürk F., İşçioğlu F., İpekçi S.H. PIK3R1 gene polymorphisms are associated with type 2 diabetes and related features in the Turkish population. *Adv. Clin. Exp. Med.* 2018;27(7):921-927. DOI 10.17219/acem/68985.

Kirkham P.A., Barnes P.J. Oxidative stress in COPD. *Chest.* 2013; 144(1):266-273. DOI 10.1378/chest.12-2664.

Korytina G.F., Akhmadishina L.Z., Aznabaeva Y.G., Kochetova O.V., Zagidullin N.S., Kzhyshkowska J.G., Zagidullin S.Z., Viktorova T.V. Associations of the NRF2/KEAP1 pathway and antioxidant defense gene polymorphisms with chronic obstructive pulmonary disease. *Gene.* 2019;692:102-112. DOI 10.1016/j.gene.2018.12.061.

Kugel S., Mostoslavsky R. Chromatin and beyond: the multitasking roles for SIRT6. *Trends Biochem. Sci*. 2014;39(2):72-81. DOI 10.1016/j.tibs.2013.12.002.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., de Bakker P.I., Daly M.J., Sham P.C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet*. 2007;81(3):559-575. DOI 10.1086/519795.

Ragland M.F., Benway C.J., Lutz S.M., Bowler R.P., Hecker J., Hokanson J.E., Crapo J.D., Castaldi P.J., DeMeo D.L., Hersh C.P., Hobbs B.D., Lange C., Beaty T.H., Cho M.H., Silverman E.K. Genetic advances in chronic obstructive pulmonary disease. Insights from COPDGene. *Am J. Respir. Crit. Care. Med*. 2019;200(6):677-690. DOI 10.1164/rccm.201808-1455SO.

Rajendrasozhan S., Yang S.R., Kinnula V.L., Rahman I. SIRT1, an antiinflammatory and antiaging protein, is decreased in lungs of patients with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 2008;177(8):861-870. DOI 10.1164/rccm.200708-1269OC.

Ryter S.W., Rosas I.O., Owen C.A., Martinez F.J., Choi M.E., Lee C.G., Elias J.A., Choi A.M.K. Mitochondrial dysfunction as a pathogenic mediator of chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *Ann. Am. Thorac. Soc*. 2018;15(Suppl. 4):S266-S272. DOI 10.1513/AnnalsATS.201808-585MG.

Song X., Wang H., Wang C., Ji G., Jiang P., Liang D., Wang X. Association of sirtuin gene polymorphisms with susceptibility to coronary artery disease in a North Chinese population. *Biomed. Res. Int.* 2022;2022:4294008. DOI 10.1155/2022/4294008.

Sun W., Liu C., Chen Q., Liu N., Yan Y., Liu B. SIRT3: A new regulator of cardiovascular diseases. *Oxid. Med. Cell Longev*. 2018;2018: 7293861. DOI 10.1155/2018/7293861.

Takasaka N., Araya J., Hara H., Ito S., Kobayashi K., Kurita Y., Wakui H., Yoshii Y., Yumino Y., Fujii S., Minagawa S., Tsurushige C., Kojima J., Numata T., Shimizu K., Kawaishi M., Kaneko Y., Kamiya N., Hirano J., Odaka M., Morikawa T., Nishimura S.L., Nakayama K., Kuwano K. Autophagy induction by SIRT6 through attenuation of insulin-like growth factor signaling is involved in the regulation of human bronchial epithelial cell senescence. *J. Immunol.* 2014;192(3):958-968. DOI 10.4049/jimmunol.1302341.

Vannitamby A., Seow H.J., Anderson G., Vlahos R., Thompson M., Steinfort D., Irving L.B., Bozinovski S. Tumour-associated neutrophils and loss of epithelial PTEN can promote corticosteroid-insensitive MMP-9 expression in the chronically inflamed lung microenvironment. *Thorax.* 2017;72(12):1140-1143. DOI 10.1136/thoraxjnl-2016-209389.

Wang C.H., Wu S.B., Wu Y.T., Wei Y.H. Oxidative stress response elicited by mitochondrial dysfunction: implica-

tion in the pathophysiology of aging. *Exp. Biol. Med.* 2013;238(5):450-460. DOI 10.1177/1535370213493069.

Ward L.D., Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res*. 2016;44(D1):D877-D881. DOI 10.1093/nar/gkv1340.

Worby C.A., Dixon J.E. PTEN. *Annu. Rev. Biochem*. 2014;83:641-669. DOI 10.1146/annurev-biochem-082411-113907.

Wu Q.J., Zhang T.N., Chen H.H., Yu X.F., Lv J.L., Liu Y.Y., Liu Y.S., Zheng G., Zhao J.Q., Wei Y.F., Guo J.Y., Liu F.H., Chang Q., Zhang Y.X., Liu C.G., Zhao Y.H. The sirtuin family in health and disease. *Signal Transduct. Target. Ther*. 2022;7(1):402. DOI 10.1038/s41392-022-01257-8.

Zhang X.Y., Li W., Zhang J.R., Li C.Y., Zhang J., Lv X.J. Roles of sirtuin family members in chronic obstructive pulmonary disease. *Respir. Res*. 2022; 23(1):66. DOI 10.1186/s12931-022-01986-y.

ORCID ID

G.F. Korytina orcid.org/0000-0002-1695-5173
L.Z. Akhmadishina orcid.org/0000-0003-0043-5090
V.A. Markelov orcid.org/0000-0002-0663-7219
Y.G. Aznabaeva orcid.org/0000-0002-1518-774X
O.V. Kochetova orcid.org/0000-0002-2944-4428
T.R. Nasibullin orcid.org/0000-0001-8823-8678
A.P. Larkina orcid.org/0009-0003-0710-6705
N.N. Khusnutdinova orcid.org/0000-0003-4127-078X
N.Sh. Zagidullin orcid.org/0000-0003-2386-6707
T.V. Victorova orcid.org/0000-0001-8900-2480

Acknowledgements. The study was performed with the support of the grant of Russian Science Foundation 23-25-00019 (http://rscf.ru/project/23-25-00019/). The work was performed using the equipment of the Centre for Collective Use "Biomika" and the unique KODINK research facility (Institute of Biochemistry and Genetics, Ufa Federal Research Centre, Russian Academy of Sciences).

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

Received March 20, 2023. Revised April 19, 2023. Accepted April 19, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):522-529
DOI 10.18699/VJGB-23-63

Original Russian text https://vavilovj-icg.ru/

Searching for new genes associated with the familial hypercholesterolemia phenotype using whole-genome sequencing and machine learning

D.E. Ivanoshchuk[1, 3] ✉, A.B. Kolker[2], O.V. Timoshchenko[1], S.E. Semaev[1, 3], E.V. Shakhtshneider[1, 3]

[1] Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State Technical University, Novosibirsk, Russia
[3] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ dinara2084@mail.ru

Abstract. One of the most common congenital metabolic disorders is familial hypercholesterolemia. Familial hypercholesterolemia is a condition caused by a type of genetic defect leading to a decreased rate of removal of low-density lipoproteins from the bloodstream and a pronounced increase in the blood level of total cholesterol. This disease leads to the early development of cardiovascular diseases of atherosclerotic etiology. Familial hypercholesterolemia is a monogenic disease that is predominantly autosomal dominant. Rare pathogenic variants in the *LDLR* gene are present in 75–85 % of cases with an identified molecular genetic cause of the disease, and variants in other genes (*APOB*, *PCSK9*, *LDLRAP1*, *ABCG5*, *ABCG8*, and others) occur at a frequency of < 5 % in this group of patients. A negative result of genetic screening for pathogenic variants in genes of the low-density lipoprotein receptor and its ligands does not rule out a diagnosis of familial hypercholesterolemia. In 20–40 % of cases, molecular genetic testing fails to detect changes in the above genes. The aim of this work was to search for new genes associated with the familial hypercholesterolemia phenotype by modern high-tech methods of sequencing and machine learning. On the basis of a group of patients with familial hypercholesterolemia (enrolled according to the Dutch Lipid Clinic Network Criteria and including cases confirmed by molecular genetic analysis), decision trees were constructed, which made it possible to identify cases in the study population that require additional molecular genetic analysis. Five probands were identified as having the severest familial hypercholesterolemia without pathogenic variants in the studied genes and were analyzed by whole-genome sequencing on the HiSeq 1500 platform (Illumina). The whole-genome sequencing revealed rare variants in three out of five analyzed patients: a heterozygous variant (rs760657350) located in a splicing acceptor site in the *PLD1* gene (c.2430-1G>A), a previously undescribed single-nucleotide deletion in the *SIDT1* gene [c.2426del (p.Leu809CysfsTer2)], new missense variant c.10313C>G (p.Pro3438Arg) in the *LRP1B* gene, and single-nucleotide deletion variant rs753876598 [c.165del (p.Ser56AlafsTer11)] in the *CETP* gene. All these variants were found for the first time in patients with a clinical diagnosis of familial hypercholesterolemia. Variants were identified that may influence the formation of the familial hypercholesterolemia phenotype.
Key words: familial hypercholesterolemia; whole-genome sequencing; machine learning; *SIDT1*; *LRP1B*; *PLD1*; *CETP*.

For citation: Ivanoshchuk D.E., Kolker A.B., Timoshchenko O.V., Semaev S.E., Shakhtshneider E.V. Searching for new genes associated with the familial hypercholesterolemia phenotype using whole-genome sequencing and machine learning. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):522-529. DOI 10.18699/VJGB-23-63

Поиск новых генов, ассоциированных с фенотипом семейной гиперхолестеринемии, методами полногеномного секвенирования и машинного обучения

Д.Е. Иванощук[1, 3] ✉, А.Б. Колкер[2], О.В. Тимощенко[1], С.Е. Семаев[1, 3], Е.В. Шахтшнейдер[1, 3]

[1] Научно-исследовательский институт терапии и профилактической медицины – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский государственный технический университет, Новосибирск, Россия
[3] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ dinara2084@mail.ru

Аннотация. Одним из наиболее распространенных врожденных метаболических нарушений является семейная гиперхолестеринемия. Это заболевание приводит к раннему развитию сердечно-сосудистых заболеваний атеросклеротического генеза. Семейная гиперхолестеринемия относится к моногенным заболеваниям с пре-

имущественно аутосомно-доминантным типом наследования. Редкие патогенные варианты в гене *LDLR* определяются в 75–85 % случаев у пациентов с выявленной молекулярно-генетической причиной заболевания, варианты в других генах встречаются с частотой менее 5 % (*APOB*, *PCSK9*, *LDLRAP1*, *ABCG5*, *ABCG8* и др.). Отрицательный результат генетического скрининга патогенных вариантов генов рецептора липопротеинов низкой плотности и его лигандов не исключает диагноз «семейная гиперхолестеринемия». В 20–40 % случаев при молекулярно-генетическом исследовании не удается определить изменения в вышеназванных генах. Цель настоящей работы – поиск новых генов, ассоциированных с фенотипом семейной гиперхолестеринемии, с использованием современных высокотехнологичных методов секвенирования и машинного обучения. На основании выборки пациентов с семейной гиперхолестеринемией, сформированной по критериям Dutch Lipid Clinic Network и включающей случаи заболевания, подтвержденные молекулярно-генетическим анализом, построены решающие деревья, которые позволили выделить из выборки случаи, требующие дополнительного молекулярно-генетического анализа. Определены пять пробандов с наиболее тяжелым течением семейной гиперхолестеринемии без патогенных вариантов в изученных генах для проведения полногеномного секвенирования на платформе HiSeq 1500 (Illumina). При выполнении полногеномного секвенирования у трех из пяти обследованных пациентов найдены редкие варианты: гетерозиготный вариант (rs760657350), локализованный в акцепторном сайте сплайсинга гена *PLD1*: c.2430-1G>A, ранее не описанная однонуклеотидная делеция гена *SIDT1*: c.2426del (p.Leu809CysfsTer2), новый миссенс-вариант c.10313C>G (p.Pro3438Arg) гена *LRP1B* и вариант однонуклеотидной делеции rs753876598: c.165del (p.Ser56AlafsTer11) гена *CETP*. Все варианты впервые описаны у пробандов с клиническим диагнозом «семейная гиперхолестеринемия». Идентифицированы варианты, которые потенциально могут влиять на формирование фенотипа семейной гиперхолестеринемии.
Ключевые слова: семейная гиперхолестеринемия; полногеномное секвенирование; машинное обучение; *SIDT1*; *LRP1B*; *PLD1*; *CETP*.

Introduction

One of the most common congenital metabolic disorders is familial hypercholesterolemia (FH) (Ezhov et al., 2019). Familial hypercholesterolemia is a condition caused by a type of genetic defects leading to a decreased rate of removal of low-density lipoproteins from the bloodstream and a pronounced increase in the blood level of total cholesterol (Ezhov et al., 2019). This illness leads to early development of cardiovascular diseases of atherosclerotic origin (Wiegman et al., 2015; Santos et al., 2016; Borén et al., 2020). Familial hypercholesterolemia is a monogenic disease with predominantly autosomal dominant inheritance (Ezhov et al., 2019).

The prevalence of the heterozygous type of FH in white populations is 1 per 250 people (Ezhov et al., 2019). Rare pathogenic variants in the *LDLR* gene are present in 75–85 % of cases with an identified molecular genetic cause of the disease, and variants in other genes (*APOB*, *PCSK9*, *LDLRAP1*, *ABCG5*, *ABCG8*, and others) occur in this group of patients with a frequency of less than 5 % (Nordestgaard et al., 2013; Iacocca, Hegele, 2017; Vasilyev et al., 2020). Patients can be homozygous or heterozygous carriers of pathogenic variants, and this status determines the severity of the disease and onset age of manifestations of cardiovascular complications (Vaezi, Amini, 2022). A negative result of genetic screening for pathogenic variants does not rule out familial hypercholesterolemia. In 20–40 % of cases, changes in the above genes are absent according to molecular genetic analysis. The risk of coronary heart disease among patients with FH is 20 times higher in the absence of treatment (Khera et al., 2016); therefore, it is important and relevant to search for new approaches to identifying patients at early disease stages and to assessing predisposition to this disease in patients' families.

To search for new cases, some authors have proposed a classifier to identify potential patients with FH by means of electronic medical records. Using data from patients with confirmed FH (n = 197) and cases without FH (n = 6590), a decision tree classifier was trained in that study. The classifier showed a positive predictive value (PPV) of 0.88 and a sensitivity of 0.75 for long-term testing. This classifier proved to be effective at finding candidate patients for further screening for familial hypercholesterolemia. Such machine-learning-based strategies can result in efficient identification of patients having the highest risk of the disease (Banda et al., 2019).

For the diagnosis of FH, clinicians use the principle of cascade genetic screening. The latter is a step-by-step identification of patients with familial hypercholesterolemia. When elevated blood levels of total cholesterol and low-density lipoprotein cholesterol (LDL-C) are detected in a patient, his/her family history of health problems is collected and clinical manifestations are analyzed. In case of a diagnosis of "probable" or "definite" FH according to the Dutch Lipid Clinic Network Criteria (Geneva: World Health Organization), the patient is referred for molecular genetic testing. The cascade screening includes quantitation of blood lipids in all first-degree relatives of the proband. If the FH diagnosis is confirmed by the molecular genetic methods, then genetic screening is performed on his/her relatives. As new patients with FH are identified, their relatives are examined too. The cascade screening is the most effective way to detect previously undiagnosed FH (Nordestgaard et al., 2013; Ezhov et al., 2019).

The aim of the present work was to search for new genes associated with the FH phenotype using modern high-tech methods of sequencing and machine learning.

Materials and methods

A group of patients with FH (ICD-10 E78.0, E78.2, n = 102), was recruited from a clinical diagnostic department at the Institute of Internal and Preventive Medicine (IIPM) – a branch of the Institute of Cytology and Genetics, the Siberian Branch of the Russian Academy of Sciences (ICG SB RAS). The study protocol was approved by the Ethics Committee at the IIPM – a branch of the ICG SB RAS, decision No. 68 of June 4, 2019. Informed consent was obtained from each study participant.

The diagnosis of FH was made in accordance with the clinical lipid criteria of the Dutch Lipid Clinic Network (DLCN) (WHO-Human genetics DoNDP…, 1999). According to the criteria, a score was computed (see the Supplementary)[1] for patients with familial hypercholesterolemia. The patients underwent a medical examination, ultrasonographic diagnostics, and blood sampling for biochemical assays (lipid profiling and general analysis of biochemical parameters) and molecular genetic assays.

Blood samples for the biochemical analyses were collected once from the cubital vein in the morning on an empty stomach at 12 h after a meal. Serum levels of total cholesterol, triglycerides, LDL-C, high-density lipoprotein cholesterol (HDL-C), and glucose were determined by enzymatic methods on a KoneLab300i automatic biochemical analyzer (Finland) with reagents from TermoFisher (Finland). The LDL-C level was calculated via the Friedewald formula; at the LDL-C concentration of > 4.5 mmol/L, the method of direct quantification of LDL-C was used. Statistical analysis of the data was performed in the SPSS software for Windows, version 23.0.

Phenol-chloroform extraction was carried out to isolate DNA from blood (Sambrook, Russell, 2006). The quality of the extracted DNA was assessed with the help of an Agilent 2100 Bioanalyzer capillary electrophoresis system (Agilent Technologies Inc., USA).

Targeted DNA sequencing in patients with FH was performed on the MiSeq platform (Illumina) using a custom-designed panel of 43 genes: *LDLR*, *APOB*, *PCSK9*, *LDLRAP1*, *CETP*, *LPL*, *HMGCR*, *NPC1L1*, *PPARA*, *MTTP*, *LMF1*, *SAR1B*, *ABCA1*, *ABCG5*, *ABCG8*, *CYP7A1*, *STAP1*, *LIPA*, *PNPLA5*, *APOA1*, *APOA5*, *APOC2*, *APOE*, *LCAT*, *ANGPTL3*, *LIPC*, *APOA4*, *APOC3*, *SREBF1*, *LMNA*, *PPARG*, *PLIN1*, *POLD1*, *LPA*, *SMAD1*, *SMAD2*, *SMAD3*, *SMAD4*, *SMAD5*, *SMAD6*, *SMAD7*, *SMAD9*, and *LIPG* (NimbleGen SeqCap Target Enrichment, Roche, Switzerland).

At the next stage of this work, from the study population, 42 patients with FH were chosen who did not show pathogenic variants in the tested genes during the targeted sequencing analysis. These patients were subjected to multiplex ligation-dependent probe amplification (MLPA) analysis to identify possible sequence alterations (deletions or duplications) in the *LDLR* gene promoter and exons by means of SALSA MLPA KIT P062 (MRCHolland, Amsterdam, the Netherlands).

Using the group of patients with FH (compiled according to the DLCN criteria and including cases of the disease confirmed by molecular genetic analysis), decision trees were constructed, which enabled us to identify cases in this group that require additional molecular genetic analysis. Software was written in Python 3.9 for building a set of decision rules for predicting FH on the basis of machine learning with a limited training set. The decision rules were stored as data representation in the Predictive Model Markup Language. The decision rules were built by means of a labeled database of patients with a diagnosis of FH (Certificate of Database Registration: RU No. 2023660511; software registration application of May 2, 2023).

By machine learning methods, five probands with the severest FH without pathogenic variants in the tested genes were identified for subsequent whole-genome sequencing on the HiSeq 1500 platform (Illumina). Automated processing and annotation of the obtained sequencing data were conducted on the NGS Wizard platform (genomenal.ru). The sequence reads were mapped to the reference human genome (GRCh38/hg38). A potential effect of novel missense variants on protein function/structure was assessed using data from *in silico* prediction tools (CADD (https://cadd.gs.washington.edu/snv), PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/), and MutationTaster (https://www.mutationtaster.org/)) and data from gnomAD on the frequency of these variants in populations. In this way, variants (in genes associated with lipid metabolism) leading to a loss of protein function and missense variants with a frequency of less than 0.01 % were selected. Pathogenicity of new variants was evaluated according to the guidelines of the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (Richards et al., 2015). Analysis of protein-protein interaction networks was performed in STRING (Szklarczyk et al., 2019).

Results

By targeted sequencing and MLPA, "pathogenic" and "likely pathogenic" variants were detected in 47.5 % of the examined probands and in 85.7 % of the probands' children. Variants in the *LDLR* gene that were identified in patients with the FH phenotype in our study are presented in Table 1. All missense variants were found to be in a heterozygous state. Variants Cys352Tyr, Cys340Phe, and Leu401His have been previously described in patients with FH in Russia (Zakharova et al., 2005) and in other countries (Feussner et al., 1996; Torres et al., 2014).

In the study population, the heterozygous type of the disease was due to rare variants in the *LDLR* gene in 73 % of the cases. Two new variants – NM_000527.5:c.266G>C, NP_000518.1:p.Cys89Ser and NM_000527.5:c.1123T>G, NP_000518.1:p.Tyr375Asp – were identified in the *LDLR* gene. Two unrelated probands turned out to be carriers of compound heterozygous variants of the *LDLR* gene, whereas the clinical course of the disease in these patients corresponded to the homozygous type of familial hypercholesterolemia. In the first case, in a 28-year-old female patient with a diagnosis of definite FH, we detected rare variants NM_000527.5:c.796G>A, NP_000518.1:p.Asp266Asn, and NM_000527.5:c.1054T>A, NP_000518.1:p.Cys352Ser in exons 5 and 7 of *LDLR* (see Table 1). In the second case, a 35-year-old female patient with a diagnosis of definite FH has two missense variants in exons 3 and 8 of *LDLR*. One substitution is located in *LDLR* exon 3 (NM_000527.5:c.266G>C, NP_000518.1:p.Cys89Ser), in which at this position, rare likely pathogenic variant rs875989894 NM_000527.5:c.266G>A, NP_000518.1:p.Cys89Tyr has been previously described in patients with FH (Day et al., 1997; Graham et al., 1999; Fouchier et al., 2005).

The pathogenicity of the identified variant was confirmed by *in silico* analysis (Mutation Taster score: 112, CADD score: 24.8, PolyPhen-2 score: 1.000). The other new missense variant is situated in exon 8 of *LDLR*: NM_000527.5:c.1123T>G, NP_000518.1:p.Tyr375Asp (see Table 1). This missense variant causes an amino acid substitution at the same position where other missense variants have been described before as

[1] Supplementary Material is available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/Suppl_Ivanoshchuk_Engl_27_5.pdf

Table 1. Single-nucleotide variants in genes *LDLR*, *APOB*, and *LPL* of patients with the FH phenotype

Variant ID	Amino acid substitution	Minor allele frequency, gnomAD (v2.1.1)	Interpretation of the nucleotide sequence variant, according to database ClinVar or LOVD
		LDLR	
rs121908038	p.Leu401His	ND	Likely pathogenic
rs137853964	p.Val827Ile	A = 0.000919	
rs28942078	p.Val429Met	A = 0.000012	Pathogenic
rs539080792	p.Glu337Lys	A = 0.000106	
rs570942190	p.Arg416Trp	T = 0.000024	
rs755757866	p.Cys340Tyr	T = 0.000008	Likely pathogenic
rs761954844	p.Cys329Tyr	A = 0.000025	
rs879254566	p.Asp178Glu	ND	Pathogenic
rs879254721	p.Glu308Lys	ND	
rs879254980	p.Glu558Ter	ND	
rs879255191	–	ND	Likely pathogenic
rs875989907	p.Asp266Asn	A = 0.000012	Pathogenic
rs879254769	p.Cys352Ser	ND	Likely pathogenic
ND/rs875989894	p.Cys89Ser	ND	New variant
ND	p.Tyr375Asp	ND	
		APOB	
rs5742904	p.Arg3527Gln	T = 0.000294	Pathogenic
		LPL	
rs118204077	p.Arg270Cys	C = 0.000008	Pathogenic

Note. ND – no data; GenBank accession numbers of protein sequences that were used in variant annotation: *LDLR* (NP_000518.1), *APOB* (NP_000375.3), and *LPL* (NP_000228.1).

likely pathogenic in patients with FH (Assouline et al., 1997; García-García et al., 2001; Damgaard et al., 2005; Mollaki et al., 2014). The pathogenicity of the variant was also confirmed by *in silico* analysis (Mutation Taster score: 160, CADD score: 25.5, PolyPhen-2 score: 1.000). Both detected variants (NM_000527.5:c.266G>C, NP_000518.1:p.Cys89Ser and NM_000527.5:c.1123T>G, NP_000518.1:p.Tyr375Asp) are not annotated in the gnomAD database (v2.1.1). According to this evidence, both were assumed to be likely pathogenic variants.

Patients without functionally significant substitutions in lipid metabolism genes were analyzed by MLPA to determine sequence changes (deletions or duplications) in the promoter and exons of the *LDLR* gene. This assay revealed a deletion of a coding region in the *LDLR* gene [NM_000527.4:c.(2140+1_2141-1)_(2311+1_2312-1)del] in DNA samples from two unrelated patients.

In the molecular genetic analysis, in three patients from two unrelated families (a proband and a son of the proband from one family and a proband from another family), we identified variant rs5742904 (NM_000384.3:c.10580G>A, NP_000375.3:p.Arg3527Gln) (ClinVar Variation ID:17890) in the *APOB* gene (see Table 1).

Rare substitutions in the *APOB* gene region encoding the LDL receptor-inding site are associated with hypercholesterolemia. One of the variants in this region, NP_000375.3:p.Arg3527Gln, leads to hypercholesterolemia with reduced clearance of LDL-C owing to a defect in the structural motif of LDL that is responsible for affinity for LDL receptor (Pullinger et al., 1995).

Analysis of our data of targeted high-throughput sequencing revealed rare pathogenic variant rs118204077 (NM_000237.3:c.808C>T, NP_000228.1:p.Arg270Cys) in the *LPL* gene in a heterozygous state (ClinVar Variation ID: 1548) (see Table 1). This variant was found in a 45-year-old male patient with hypercholesterolemia (12.4 mmol/L) and hypertriglyceridemia (17.4 mmol/L; DLCN score: 5). Earlier, variants associated with hypertriglyceridemia have been described at this locus (Ma et al., 1994; Surendran et al., 2012) in patients with lipoprotein lipase deficiency (Hegele et al., 2018; Teramoto et al., 2018).

According to the results of the molecular genetic analyses, 52.5 % of our patients are not carriers of pathogenic variants in the studied lipid metabolism genes. Among these patients, by a machine learning algorithm, five subjects with the severest FH were chosen for whole-genome sequencing. As a result, in three patients, four variants with a minor allele frequency of less than 0.01 % were identified in genes related to lipid metabolism. Among these variants, two are single-nucleotide deletions, one affects a splicing acceptor site, and one is a missense variant. The findings are presented in Table 2.

In the *SIDT1* gene (encoding a protein called SID1 transmembrane family member 1), a new previously undescribed variant was detected that yields a frameshift starting with

Table 2. Rare variants identified in lipid metabolism genes by whole-genome sequencing in patients with the FH phenotype

Gene	Nucleotide sequence position (GRCh38/hg38)	Position in cDNA (transcript's GenBank accession No.)	Amino acid substitution (protein's GenBank accession No.)	Genotype	Minor allele frequency (gnomAD v2.1.1)	Minor allele frequency (RuSeq)	dbSNP ID
SIDT1	chr3:113627646 GT>G	(NM_017699.3) c.2426del	(NP_060169.2) p.Leu809CysfsTer2	Heterozygote	–	–	New
LRP1B	chr2:140442605 G>C	(NM_018557.3) c.10313C>G	(NP_061027.2) p.Pro3438Arg	Heterozygote	–	–	New
PLD1	chr3:171645024 G>A	(NM_002662.5) c.2430, –1G>A	–	Heterozygote	0.00001768	0.0002457	rs760657350
CETP	chr16:56963054 GC>G	(NM_000078.3) c.165del	(NP_000069.2) p.Ser56AlafsTer11	Heterozygote	0.00001415	0.0002081	rs753876598

codon 809 (NM_017699.3:c.2426de, NP_060169.2:p.Leu809CysfsTer2). The gene consists of 30 exons and is located in chromosomal region 3q13.2 (https://www.ncbi.nlm.nih.gov/gene/54847). According to gnomAD, loss-of-function variants in this gene have been documented, but according to ClinVar and the literature, none of them have been annotated as pathogenic or likely pathogenic. For this gene, the pLI score was found to be 0, indicating that the gene is resistant to loss-of-function variants. The weight of evidence suggested that this variant has uncertain clinical significance (pathogenicity criterion PM2).

A new missense variant was found in the *LRP1B* gene (LDL receptor-related protein 1B): c.10313C>G p.Pro3438Arg in a heterozygous state. The gene is located in chromosomal region 2q22.1-q22.2, consists of 92 exons, and encodes one of the receptors of the LDL receptor family (https://www.ncbi.nlm.nih.gov/gene/53353). In gnomAD, there are no data on the frequency of this variant. The pathogenicity of this variant was also corroborated by *in silico* analysis (Mutation Taster score: 103, CADD score: 33, PolyPhen-2 score: 1.000). Most variants in this gene either are benign (data from ClinVar, accessed in February 2023) or have uncertain clinical significance. The totality of the data indicated that this variant has uncertain clinical significance (pathogenicity criteria PM2, PP3, and BP1).

One of our patients proved to be a carrier of a rare heterozygous variant at a splicing acceptor site (NM_002662.5:c.2430, –1G>A) in the phospholipase D1 (*PLD1*) gene. This variant is registered in a control sample in gnomAD: five mutant alleles on 282,768 chromosomes (no homozygotes have been detected). The *PLD1* gene codes for a phosphatidylcholine-specific phospholipase that catalyzes the hydrolysis of phosphatidylcholine, thus yielding phosphatidic acid and choline (https://www.ncbi.nlm.nih.gov/gene/5337). The gene is situated in chromosomal region 3q26.31 and contains 35 exons. Phospholipase D (PLD) and its enzymatic reaction product, phosphatidic acid, regulate adhesion of immune cells (macrophages and neutrophils) to collagen (Speranza et al., 2014).

It is known that biallelic variants with loss of function of the *PLD1* gene cause neonatal cardiomyopathy and congenital malformations of the pulmonary valve and tricuspid valve, of the right ventricle of the heart, and of the outflow tract of the right ventricle (Ta-Shma et al., 2017; Lahrouchi et al., 2021). The weight of evidence suggested that this substitution is a likely pathogenic variant in relation to congenital heart malformations (pathogenicity criteria PM2 and PVS1). With respect to FH, we categorized the detected substitution as a variant of unknown clinical significance (pathogenicity criterion PM2).

One of the examined patients was found to have a heterozygous single-nucleotide deletion in *CETP*: rs753876598 (NM_000078.3:c.165del) (https://www.ncbi.nlm.nih.gov/gene/1071). The variant is annotated in the ClinVar database (ID1675625) and is registered in a control sample of gnomAD: four mutant alleles on 282,774 chromosomes (no homozygotes have been found). It is known that variants causing loss of function of this gene affect the HDL-C level (Millwood et al., 2018; Li et al., 2021). According to the totality of criteria for pathogenicity evaluation (PM2 and PVS1), we designated this variant as likely pathogenic. The *CETP* gene codes for a plasma protein that catalyzes the exchange of triglycerides and cholesterol esters between lipoprotein particles (Oliveira, Raposo, 2020).

Discussion

High-throughput sequencing is employed not only for molecular genetic diagnosis of FH but also as a tool for identifying i) variants that may be involved in lipid metabolism and ii) their effects on the phenotype of patients with FH (Miroshnikova et al., 2021). In the current study, 16 variants were identified in an FH population (15 single-nucleotide substitutions and one deletion) that have previously been classified as pathogenic or likely pathogenic in the ClinVar or LOVD database as well as two new missense variants in the *LDLR* gene that we classified as pathogenic. In our genome-wide analysis, in lipid metabolism-associated genes, we detected four additional variants that met our search criteria. Two of these four variants have been described before, and two are new.

One of the genes in which rare variants were found in patients with FH is *PLD1*, encoding phospholipase D1. This enzyme hydrolyzes membrane lipid phosphatidylcholine thereby generating phosphatidic acid (Bowling et al., 2021). Phosphatidic acid is an intermediate metabolite in the synthesis of all membrane glycerophospholipids and plays an important structural role in live cells by promoting membrane

biogenesis (Tanguy et al., 2018); furthermore, its involvement and phospholipase D1's participation in exocytosis have been demonstrated (Tanguy et al., 2022).

Alternative splicing of *PLD1* mRNA results in many different transcripts having both catalytic and regulatory functions (Nelson, Frohman, 2015). It has been shown that recessive variants in the *PLD1* gene are associated with severe right-sided congenital heart malformations in two families (Ta-Shma et al., 2017). In *Pld1* knockout mice, moderate dysfunction of pulmonary and tricuspid valves is observed (Ta-Shma et al., 2017). Recessive *PLD1* variants also correlate with isolated neonatal cardiomyopathies (Lahrouchi et al., 2021). In humans, missense variants of *PLD1* are reported to be concentrated in regions of the protein critical for catalytic activity, thus resulting in low enzymatic activity in most of such mutant proteins (Lahrouchi et al., 2021). It has also been demonstrated in cell lines that *PLD1* overexpression promotes the formation of lipid droplets, whereas an siRNA *PLD1* knockdown inhibits this process (Andersson et al., 2006).

The variant (NM_002662.5:c.2430, –1G>A) that we detected in a proband with FH without signs of congenital heart disease is in a heterozygous state. Considering the low prevalence of this variant and its possible role in subcellular transport and in the formation of lipid droplets, this substitution is of interest for further investigation in individuals with lipid metabolism disorders.

Another rare variant was found by us in the *LRP1B* gene (LDL receptor-related protein 1B). The LRP1B protein is a member of the LDL receptor family (Strickland et al., 2002). LRP1B takes part in lipoprotein catabolism; accordingly, research on rare variants of the *LRP1B* gene in individuals with FH is promising. Most of the recently identified ligands of LRP1B are well-known factors of blood coagulation and of lipoprotein metabolism, suggesting that LRP1B is implicated in atherosclerosis (Lee, 2019).

SIDT1 is a multispan transmembrane protein belonging to the SID1 transmembrane family and shares some sequence homology with *Caenorhabditis elegans* ChUP-1, which is a cholesterol-binding protein located in intracellular vesicles (Valdes et al., 2012). SIDT1 expression in endolysosomes has been documented (Nguyen et al., 2019). *SIDT1* has been shown to participate in cholesterol transport (Méndez-Acevedo et al., 2017) but has not been investigated in the context of the FH phenotype. Most likely, the variant that we found in this gene does not take part in the formation of the clinical phenotype of FH because our assessment using the American College of Medical Genetics and Genomics criteria classifies it as a variant of uncertain clinical significance; however, for unambiguous evaluation of its association with the FH phenotype, additional data are needed.

The *CETP* gene codes for the CETP protein, which carries cholesterol esters. This protein regulates the concentration and particle size of HDL-C in the blood and plays an important role in reverse cholesterol transport (Barter, Kastelein, 2006). It has been shown that elevated activity of *CETP* reduces HDL-C concentration and correlates with a higher risk of cardiovascular disease (Barter, 2011; Iwanicka et al., 2018). Variants in the *CETP* gene can alter the blood lipid profile (Wuni et al., 2022). In our previous study on one of the *CETP* variants, we reported its association with changes in the blood lipid profile and with the risk of myocardial infarction in a population of Western Siberia (Semaev et al., 2019). When a map of functional and physical associations was constructed in the present study, the APOB protein turned out to be a predicted functional partner of the CETP protein, and mutations in APOB represent some of FH etiologies.

Additional segregational and functional analyses are necessary to evaluate pathogenic effects of the identified variants on the formation of the clinical FH phenotype. Identification of new pathogenic variants will facilitate risk assessment of FH and of its complications among patients and members of their families.

Conclusion

A combination of machine learning and whole-genome sequencing in probands with a clinical diagnosis of FH revealed rare variants in genes *SIDT1*, *LRP1B*, *PLD1*, and *CETP*; these variants may influence the disease phenotype.

References

Andersson L., Boström P., Ericson J., Rutberg M., Magnusson B., Marchesan D., Ruiz M., Asp L., Huang P., Frohman M.A., Borén J., Olofsson S.O. PLD1 and ERK2 regulate cytosolic lipid droplet formation. *J. Cell Sci*. 2006;119(Pt. 11):2246-2257. DOI 10.1242/jcs.02941.

Assouline L., Leitersdorf E., Lambert M., Reshef A., Feoli-Fonseca J.C., Levy E. Identification of two novel LDL receptor gene defects in French-Canadian pediatric population: mutational analysis and biochemical studies. *Hum. Mutat*. 1997;9(6):555-562. DOI 10.1002/(SICI)1098-1004(1997)9:6<555::AID-HUMU9>3.0.CO;2-0.

Banda J.M., Sarraju A., Abbasi F., Parizo J., Pariani M., Ison H., Briskin E., Wand H., Dubois S., Jung K., Myers S.A., Rader D.J., Leader J.B., Murray M.F., Myers K.D., Wilemon K., Shah N.H., Knowles J.W. Finding missed cases of familial hypercholesterolemia in health systems using machine learning. *NPJ Digit. Med*. 2019;2:23. DOI 10.1038/s41746-019-0101-5.

Barter P.J. The causes and consequences of low levels of high density lipoproteins in patients with diabetes. *Diabetes Metab. J*. 2011;35(2):101-106. DOI 10.4093/dmj.2011.35.2.101.

Barter P.J., Kastelein J.J.P. Targeting cholesteryl ester transfer protein for the prevention and management of cardiovascular disease. *J. Am. Coll. Cardiol*. 2006;47(3):492-499. DOI 10.1016/j.jacc.2005.09.042.

Borén J., Chapman M.J., Krauss R.M., Packard C.J., Bentzon J.F., Binder C.J., Daemen M.J., Demer L.L., Hegele R.A., Nicholls S.J., Nordestgaard B.G., Watts G.F., Bruckert E., Fazio S., Ference B.A., Graham I., Horton J.D., Landmesser U., Laufs U., Masana L., Pasterkamp G., Raal F.J., Ray K.K., Schunkert H., Taskinen M.R., van de Sluis B., Wiklund O., Tokgozoglu L., Catapano A.L., Ginsberg H.N. Low-density lipoproteins cause atherosclerotic cardiovascular disease: pathophysiological, genetic, and therapeutic insights: a consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur. Heart J*. 2020;41(24):2313-2330. DOI 10.1093/eurheartj/ehz962.

Bowling F.Z., Frohman M.A., Airola M.V. Structure and regulation of human phospholipase D. *Adv. Biol. Regul.* 2021;79:100783. DOI 10.1016/j.jbior.2020.100783.

Damgaard D., Larsen M.L., Nissen P.H., Jensen J.M., Jensen H.K., Soerensen V.R., Jensen L.G., Faergeman O. The relationship of molecular genetic to clinical diagnosis of familial hypercholesterolemia in a Danish population. *Atherosclerosis*. 2005;180(1):155-160. DOI 10.1016/j.atherosclerosis.2004.12.001.

Day I.N., Whittall R.A., O'Dell S.D., Haddad L., Bolla M.K., Gudnason V., Humphries S.E. Spectrum of LDL receptor gene mutations in heterozygous familial hypercholesterolemia. *Hum. Mutat*.

1997;10(2):116-127. DOI 10.1002/(SICI)1098-1004(1997)10:2<116::AID-HUMU4>3.0.CO;2-I.

Ezhov M.V., Bazhan S.S., Ershova A.I., Meshkov A.N., Sokolov A.A., Kukharchuk V.V., Gurevich V.S., Voevoda M.I., Sergienko I.V., Shakhtshneider E.V., Pokrovsky S.N., Konovalov G.A., Leontyeva I.V., Konstantinov V.O., Shcherbakova M.Yu., Zakharova I.N., Balakhonova T.V., Filippov A.E., Akhmedzhanov N.M., Aleksandrova O.Yu., Lipovetsky B.M. Clinical guidelines for familial hypercholesterolemia. *Ateroscleroz i Dislipidemii = The Journal of Atherosclerosis and Dyslipidemias.* 2019;15(1):58-98. (in Russian)

Feussner G., Dobmeyer J., Nissen H., Hansen T.S. Unusual xanthomas in a young patient with heterozygous familial hypercholesterolemia and type III hyperlipoproteinemia. *Am. J. Med. Genet*. 1996; 65:149-154. DOI 10.1002/(SICI)1096-8628(19961016)65:2<149::AID-AJMG14>3.0.CO;2-Q.

Fouchier S.W., Kastelein J.J., Defesche J.C. Update of the molecular basis of familial hypercholesterolemia in the Netherlands. *Hum. Mutat*. 2005;26(6):550-556. DOI 10.1002/humu.20256.

García-García A.B., Real J.T., Puig O., Cebolla E., Marín-García P., Martínez Ferrandis J.I., García-Sogo M., Civera M., Ascaso J.F., Carmena R., Armengod M.E., Chaves F.J. Molecular genetics of familial hypercholesterolemia in Spain: Ten novel *LDLR* mutations and population analysis. *Hum. Mutat*. 2001;18(5):458-459. DOI 10.1002/humu.1218.

Graham C.A., McClean E., Ward A.J., Beattie E.D., Martin S., O'Kane M., Young I.S., Nicholls D.P. Mutation screening and genotype:phenotype correlation in familial hypercholesterolaemia. *Atherosclerosis*. 1999;147(2):309-316. DOI 10.1016/s0021-9150(99)00201-4.

Hegele R.A., Berberich A.J., Ban M.R., Wang J., Digenio A., Alexander V.J., D'Erasmo L., Arca M., Jones A., Bruckert E., Stroes E.S., Bergeron J., Civeira F., Witztum J.L., Gaudet D. Clinical and biochemical features of different molecular etiologies of familial chylomicronemia. *J. Clin. Lipidol*. 2018;12(4):920-927.e4. DOI 10.1016/j.jacl.2018.03.093.

Iacocca M.A., Hegele R.A. Recent advances in genetic testing for familial hypercholesterolemia. *Expert. Rev. Mol. Diagn*. 2017;17(7): 641-651. DOI 10.1080/14737159.2017.1332997.

Iwanicka J., Iwanicki T., Niemiec P., Balcerzyk A., Krauze J., Górczyńska-Kosiorz S., Ochalska-Tyka A., Grzeszczak W., Żak I. Relationship between *CETP* gene polymorphisms with coronary artery disease in Polish population. *Mol. Biol. Rep*. 2018;45(6):1929-1935. DOI 10.1007/s11033-018-4342-1.

Khera A.V., Won H.H., Peloso G.M., Lawson K.S., Bartz T.M., Deng X., van Leeuwen E.M., … Correa A., Boerwinkle E., Merlini P.A., Ardissino D., Saleheen D., Gabriel S., Kathiresan S. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J. Am. Coll. Cardiol*. 2016;67(22):2578-2589. DOI 10.1016/j.jacc.2016.03.520.

Lahrouchi N., Postma A.V., Salazar C.M., De Laughter D.M., Tjong F., Piherová L., Bowling F.Z., … Mulder B., Airola M.V., Kmoch S., Barnett J.V., Clur S.A., Frohman M.A., Bezzina C.R. Biallelic loss-of-function variants in *PLD1* cause congenital right-sided cardiac valve defects and neonatal cardiomyopathy. *J. Clin. Invest*. 2021; 131(5):e142148. DOI 10.1172/JCI142148.

Lee S. The genetic and epigenetic association of LDL Receptor Related Protein 1B (*LRP1B*) gene with childhood obesity. *Sci. Rep*. 2019; 9(1):1815. DOI 10.1038/s41598-019-38538-2.

Li W., Liu X., Huang C., Liu L., Tan X., Wang X. The loss-of-function mutation of *CETP* affects HDLc levels but not ApoA1 in patients with acute myocardial infarction. *Nutr. Metab. Cardiovasc. Dis*. 2021;31(2):602-607. DOI 10.1016/j.numecd.2020.10.019.

Ma Y., Liu M.S., Chitayat D., Bruin T., Beisiegel U., Benlian P., Foubert L., De Gennes J.L., Funke H., Forsythe I., Blaichman S., Papanikolaou M., Erkelens D.W., Kastelein J., Brunzell J.D., Hayden M.R. Recurrent missense mutations at the first and second base of codon Arg^{243} in human lipoprotein lipase in patients of different ancestries. *Hum. Mutat*. 1994;3(1):52-58. DOI 10.1002/humu.1380030109.

Méndez-Acevedo K.M., Valdes V.J., Asanov A., Vaca L. A novel family of mammalian transmembrane proteins involved in cholesterol transport. *Sci. Rep*. 2017;7(1):7450. DOI 10.1038/s41598-017-07077-z.

Millwood I.Y., Bennett D.A., Holmes M.V., Boxall R., Guo Y., Bian Z., Yang L., Sansome S., Chen Y., Du H., Yu C., Hacker A., Reilly D.F., Tan Y., Hill M.R., Chen J., Peto R., Shen H., Collins R., Clarke R., Li L., Walters R.G., Chen Z., China Kadoorie Biobank Collaborative Group. Association of *CETP* gene variants with risk for vascular and nonvascular diseases among chinese adults. *JAMA Cardiol*. 2018;3(1):34-43. DOI 10.1001/jamacardio.2017.4177.

Miroshnikova V.V., Romanova O.V., Ivanova O.N., Fedyakov M.A., Panteleeva A.A., Barbitoff Y.A., Muzalevskaya M.V., Urazgildeeva S.A., Gurevich V.S., Urazov S.P., Scherbak S.G., Sarana A.M., Semenova N.A., Anisimova I.V., Guseva D.M., Pchelina S.N., Glotov A.S., Zakharova E.Y., Glotov O.S. Identification of novel variants in the *LDLR* gene in Russian patients with familial hypercholesterolemia using targeted sequencing. *Biomed. Rep*. 2021;14(1):15. DOI 10.3892/br.2020.1391.

Mollaki V., Progias P., Drogari E. Familial hypercholesterolemia in Greek children and their families: genotype-to-phenotype correlations and a reconsideration of *LDLR* mutation spectrum. *Atherosclerosis*. 2014;237(2):798-804. DOI 10.1016/j.atherosclerosis.2014.09.031.

Nelson R.K., Frohman M.A. Physiological and pathophysiological roles for phospholipase D. *J. Lipid. Res*. 2015;56(12):2229-2237. DOI 10.1194/jlr.R059220.

Nguyen T.A., Smith B.R.C., Elgass K.D., Creed S.J., Cheung S., Tate M.D., Belz G.T., Wicks I.P., Masters S.L., Pang K.C. SIDT1 localizes to endolysosomes and mediates double-stranded RNA transport into the cytoplasm. *J. Immunol*. 2019;202(12):3483-3492. DOI 10.4049/jimmunol.1801369.

Nordestgaard B.G., Chapman M.J., Humphries S.E., Ginsberg H.N., Masana L., Descamps O.S., Wiklund O., … Kuivenhoven J.A., Pajukanta P., Ray K., Stalenhoef A.F., Stroes E., Taskinen M.R., Tybjærg-Hansen A., European Atherosclerosis Society Consensus Panel. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. *Eur. Heart J*. 2013;34(45):3478-3490. DOI 10.1093/eurheartj/eht273.

Oliveira H.C.F., Raposo H.F. Cholesteryl ester transfer protein and lipid metabolism and cardiovascular diseases. In: Jiang X.-C. (Ed.). Lipid Transfer in Lipoprotein Metabolism and Cardiovascular Disease. Advances in Experimental Medicine and Biology. Vol. 1276. Singapore: Springer, 2020;15-25. DOI 10.1007/978-981-15-6082-8_2.

Pullinger C.R., Hennessy L.K., Chatterton J.E., Liu W., Love J.A., Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P. Familial ligand-defective apolipoprotein B. Identification of a new mutation that decreases LDL receptor binding affinity. *J. Clin. Invest*. 1995;95(3):1225-1234. DOI 10.1172/JCI117772.

Richards S., Aziz N., Bale S., Bick D., Das S., Gastier-Foster J., Grody W.W., Hegde M., Lyon E., Spector E., Voelkerding K., Rehm H.L., ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med*. 2015;17(5):405-424. DOI 10.1038/gim.2015.30.

Sambrook J., Russell D.W. Purification of nucleic acids by extraction with phenol:chloroform. *Cold Spring Harb. Protoc*. 2006;2006(1): pdb.prot4455. DOI 10.1101/pdb.prot4455.

Santos R.D., Gidding S.S., Hegele R.A., Cuchel M.A., Barter P.J., Watts G.F., Baum S.J., Catapano A.L., Chapman M.J., Defesche J.C., Folco E., Freiberger T., Genest J., Hovingh G.K., Harada-Shiba M., Humphries S.E., Jackson A.S., Mata P., Moriarty P.M., Raal F.J., Al-Rasadi K., Ray K.K., Reiner Z., Sijbrands E.J., Yamashita S., International Atherosclerosis Society Severe Familial Hypercholesterolemia Panel. Defining severe familial hypercholesterolaemia and the

implications for clinical management: a consensus statement from the International Atherosclerosis Society Severe Familial Hypercholesterolemia Panel. *Lancet Diabetes Endocrinol*. 2016;4(10):850-861. DOI 10.1016/S2213-8587(16)30041-9.

Semaev S., Shakhtshneider E., Orlov P., Ivanoshchuk D., Malyutina S., Gafarov V., Ragino Y., Voevoda M. Association of rs708272 (*CETP* gene variant) with lipid profile parameters and the risk of myocardial infarction in the white population of Western Siberia. *Biomolecules*. 2019;9(11):739. DOI 10.3390/biom9110739.

Speranza F., Mahankali M., Henkels K.M., Gomez-Cambronero J. The molecular basis of leukocyte adhesion involving phosphatidic acid and phospholipase D. *J. Biol. Chem*. 2014;289(42):28885-28897. DOI 10.1074/jbc.M114.597146.

Strickland D.K., Gonias S.L., Argraves W.S. Diverse roles for the LDL receptor family. *Trends Endocrinol. Metab*. 2002;13(2):66-74. DOI 10.1016/s1043-2760(01)00526-4.

Surendran R.P., Visser M.E., Heemelaar S., Wang J., Peter J., Defesche J.C., Kuivenhoven J.A., Hosseini M., Péterfy M., Kastelein J.J., Johansen C.T., Hegele R.A., Stroes E.S., Dallinga-Thie G.M. Mutations in *LPL*, *APOC2*, *APOA5*, *GPIHBP1* and *LMF1* in patients with severe hypertriglyceridaemia. *J. Intern. Med*. 2012;272(2):185-196. DOI 10.1111/j.1365-2796.2012.02516.x.

Szklarczyk D., Gable A.L., Lyon D., Junge A., Wyder S., Huerta-Cepas J., Simonovic M., Doncheva N.T., Morris J.H., Bork P., Jensen L.J., Mering C.V. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019; 47(D1):D607-D613. DOI 10.1093/nar/gky1131.

Tanguy E., Kassas N., Vitale N. Protein-phospholipid interaction motifs: A focus on phosphatidic acid. *Biomolecules*. 2018;8(2):20. DOI 10.3390/biom8020020.

Tanguy E., Wolf A., Wang Q., Chasserot-Golaz S., Ory S., Gasman S., Vitale N. Phospholipase D1-generated phosphatidic acid modulates secretory granule trafficking from biogenesis to compensatory endocytosis in neuroendocrine cells. *Adv. Biol. Regul*. 2022;83:100844. DOI 10.1016/j.jbior.2021.100844.

Ta-Shma A., Zhang K., Salimova E., Zernecke A., Sieiro-Mosti D., Stegner D., Furtado M., Shaag A., Perles Z., Nieswandt B., Rein A.J., Rosenthal N., Neiman A.M., Elpeleg O. Congenital valvular defects associated with deleterious mutations in the *PLD1* gene. *J. Med. Genet*. 2017;54(4):278-286. DOI 10.1136/jmedgenet-2016-104259.

Teramoto R., Tada H., Kawashiri M.A., Nohara A., Nakahashi T., Konno T., Inazu A., Mabuchi H., Yamagishi M., Hayashi K. Molecular and functional characterization of familial chylomicronemia syndrome. *Atherosclerosis*. 2018;269:272-278. DOI 10.1016/j.atherosclerosis.2017.11.006.

Torres M.M.T., Mora-Hernández S., Vázquez Cárdenas N.A., González Jaimes A. Homozygous familial hypercholesterolemia: the c.1055G>A mutation in the *LDLR* gene and clinical heterogeneity. *J. Clin. Lipidol*. 2014;8(5):525-527. DOI 10.1016/j.jacl.2014.05.002.

Vaezi Z., Amini A. Familial Hypercholesterolemia. In: StatPearls. StatPearls Publishing: Treasure Island, FL, USA, 2022. Available online: https://www.ncbi.nlm.nih.gov/books/NBK556009/ (accessed on 24 July 2022).

Valdes V.J., Athie A., Salinas L.S., Navarro R.E., Vaca L. CUP-1 is a novel protein involved in dietary cholesterol uptake in *Caenorhabditis elegans*. *PLoS One*. 2012;7(3):e33962. DOI 10.1371/journal.pone.0033962.

Vasilyev V., Zakharova F., Bogoslovskaya T., Mandelshtam M. Familial hypercholesterolemia in Russia: three decades of genetic studies. *Front. Genet*. 2020;11:550591. DOI 10.3389/fgene.2020.550591.

Wiegman A., Gidding S.S., Watts G.F., Chapman M.J., Ginsberg H.N., Cuchel M., Ose L., … Santos R.D., Stalenhoef A.F., Steinhagen-Thiessen E., Stroes E.S., Taskinen M.R., Tybjærg-Hansen A., Wiklund O., European Atherosclerosis Society Consensus Panel. Familial hypercholesterolaemia in children and adolescents: gaining decades of life by optimizing detection and treatment. *Eur. Heart J*. 2015;36(36):2425-2437. DOI 10.1093/eurheartj/ehv157.

World Health Organization-Human genetics DoNDP. Familial hypercholesterolaemia: report of a second World Health Organization consultation, Geneva: WHO; 1999. URL: https://apps.who.int/iris/handle/10665/66346 (Application: 14.11.2022).

Wuni R., Kuhnle G.G.C., Wynn-Jones A.A., Vimaleswaran K.S. A nutrigenetic update on *CETP* gene-diet interactions on lipid-related outcomes. *Curr. Atheroscler. Rep*. 2022;24(2):119-132. DOI 10.1007/s11883-022-00987-y.

Zakharova F.M., Damgaard D., Mandelshtam M.Y., Golubkov V.I., Nissen P.H., Nilsen G.G., Stenderup A., Lipovetsky B.M., Konstantinov V.O., Denisenko A.D., Vasilyev V.B., Faergeman O. Familial hypercholesterolemia in St.-Petersburg: the known and novel mutations found in the low density lipoprotein receptor gene in Russia. *BMC Med. Genet*. 2005;6:6. DOI 10.1186/1471-2350-6-6.

ORCID ID

D.E. Ivanoshchuk orcid.org/0000-0002-0403-545X
A.B. Kolker orcid.org/0000-0002-7048-6419
O.V. Timoshchenko orcid.org/0000-0002-6584-2060
S.E. Semaev orcid.org/0000-0003-3999-8501
E.V. Shakhtshneider orcid.org/0000-0001-6108-1025

Acknowledgements. The study was funded by the Russian Science Foundation, project No. 22-25-00743.

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

Received December 13, 2022. Revised May 2, 2023. Accepted May 3, 2023.

Вавиловский журнал генетики и селекции. 2023;27(5):530-538
DOI 10.18699/VJGB-23-64

Original Russian text https://vavilovj-icg.ru/

Metabolic profile of blood serum in experimental arterial hypertension

A.A. Seryapina[1] ✉, A.A. Malyavko[1], Yu.K. Polityko[1], L.V. Yanshole[2], Yu.P. Tsentalovich[2], A.L. Markel[1, 3]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] International Tomography Center of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
✉ seryapina@bionet.nsc.ru

Abstract. The etiology of essential hypertension is intricate, since it employs simultaneously various body systems related to the regulation of blood pressure in one way or another: the sympathetic nervous system, renin-angiotensin-aldosterone and hypothalamic-pituitary-adrenal systems, renal and endothelial mechanisms. The pathogenesis of hypertension is influenced by a variety of both genetic and environmental factors, which determines the heterogeneity of the disease in human population. Hence, there is a need to perform research on experimental models – inbred animal strains, one of them being ISIAH rat strain, which is designed to simulate inherited stress-induced arterial hypertension as close as possible to primary (or essential) hypertension in humans. To determine specific markers of diseases, various omics technologies are applied, including metabolomics, which makes it possible to evaluate the content of low-molecular compounds – amino acids, lipids, carbohydrates, nucleic acids fragments – in biological samples available for clinical analysis (blood and urine). We analyzed the metabolic profile of the blood serum of male ISIAH rats with a genetic stress-dependent form of arterial hypertension in comparison with the normotensive WAG rats. Using the method of nuclear magnetic resonance spectroscopy (NMR spectroscopy), 56 metabolites in blood serum samples were identified, 18 of which were shown to have significant interstrain differences in serum concentrations. Statistical analysis of the data obtained showed that the hypertensive status of ISIAH rats is characterized by increased concentrations of leucine, isoleucine, valine, myo-inositol, isobutyrate, glutamate, glutamine, ornithine and creatine phosphate, and reduced concentrations of 2-hydroxyisobutyrate, betaine, tyrosine and tryptophan. Such a ratio of the metabolite concentrations is associated with changes in the regulation of glucose metabolism (metabolic markers – leucine, isoleucine, valine, myo-inositol), of nitric oxide synthesis (ornithine) and catecholamine pathway (tyrosine), and with inflammatory processes (metabolic markers – betaine, tryptophan), all of these changes being typical for hypertensive status. Thus, metabolic profiling of the stress-dependent form of arterial hypertension seems to be an important result for a personalized approach to the prevention and treatment of hypertensive disease.
Key words: arterial hypertension; ISIAH rats; metabolic markers.

For citation: Seryapina A.A., Malyavko A.A., Polityko Yu.K., Yanshole L.V., Tsentalovich Yu.P., Markel A.L. Metabolic profile of blood serum in experimental arterial hypertension. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(5):530-538. DOI 10.18699/VJGB-23-64

Метаболомный профиль сыворотки крови при экспериментальной артериальной гипертензии

А.А. Серяпина[1] ✉, А.А. Малявко[1], Ю.К. Политыко[1], Л.В. Яньшоле[2], Ю.П. Центалович[2], А.Л. Маркель[1, 3]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Институт «Международный томографический центр» Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ seryapina@bionet.nsc.ru

Аннотация. Этиология гипертонической болезни неочевидна, поскольку одновременно оказываются задействованы различные системы организма, тем или иным образом связанные с регуляцией артериального давления: симпатическая нервная, ренин-ангиотензин-альдостероновая и гипоталамо-гипофизарно-надпочечниковая системы, почечные и эндотелиальные механизмы. На патогенез гипертонической болезни влияет множество как генетических, так и средовых факторов, что обусловливает популяционную гетерогенность заболевания у людей. В связи с этим возникает необходимость в проведении исследований на экспериментальных моделях – инбредных линиях животных. Таковой является линия крыс НИСАГ (ISIAH), воспроизводящая наследственную индуцированную стрессом артериальную гипертензию, максимально приближенную к артериальной гипертонии у людей. Для определения специфических маркеров заболеваний используются «омиксные» технологии, в том числе метаболомные, которые дают представление о профиле концентраций низкомолекулярных соединений – аминокислот, липидов, углеводов, фрагментов нуклеиновых кислот – в биологических образцах, доступных

для клинического анализа (кровь и моча). В настоящей работе проведен анализ метаболомного профиля сыворотки крови самцов крыс линии НИСАГ с генетической стресс-зависимой формой артериальной гипертензии по сравнению с нормотензивной линией крыс WAG. С применением метода спектроскопии ядерно-магнитного резонанса (ЯМР-спектроскопия) в образцах сыворотки крови было идентифицировано 56 метаболитов, при этом для 18 метаболитов выявлены достоверные различия по концентрации в сыворотке крови между линиями крыс. Статистический анализ полученных данных показал, что гипертензивный статус крыс НИСАГ характеризуется сочетанным повышением концентраций лейцина, изолейцина, валина, мио-инозитола, изобутирата, глутамата, глутамина, орнитина и креатинфосфата и понижением концентраций 2-гидроксиизобутирата, бетаина, тирозина и триптофана. Такие изменения концентраций метаболитов ассоциированы с характерными для гипертензивного статуса изменениями в регуляции метаболизма глюкозы (метаболомные маркеры – лейцин, изолейцин, валин и мио-инозитол), синтеза оксида азота (орнитин) и катехоламинов (тирозин) и с воспалительными процессами (метаболомные маркеры – бетаин, триптофан). Таким образом, идентификация метаболомного профиля стресс-зависимой формы артериальной гипертонии представляется важным результатом, полезным для разработки персонализированного подхода к профилактике и лечению гипертонической болезни.
Ключевые слова: артериальная гипертензия; крысы НИСАГ (ISIAH); метаболомные маркеры.

Introduction

Hypertension is a complex multifactorial disease determined by both genetic and environmental factors, as well as the effects of genotype-environment interactions. Currently, a wide selection of antihypertensive drugs and their combinations is available for clinical medicine (Laurent, 2017). However, only a few of them are actually used (vasodilators, diuretics, blockers of certain receptors and ion channels): they affect the final links in the pathogenesis of arterial hypertension and usually do not address the initial etiological mechanisms of the disease. This can partly explain the fact that only 30 % of hypertensive patients successfully achieve and control blood pressure (BP) targets (Thoenes et al., 2010).

To improve the effectiveness of assigned therapy, objective criteria that enable positive identification of the individual etiological and pathogenetic characteristics of the disease are needed. First of all, of interest are genetic markers. Genes associated with arterial hypertension have been identified in numerous studies, including genome-wide analysis of a huge number of polymorphisms. However, these polymorphic loci account for only a small percentage (2–3 %) of BP variability in the tested populations (Hoffmann et al., 2017). Obviously, the contribution of environmental factors, as well as the effects of genotype-environment interaction, dominates. Non-additive intergenic interactions and epigenetic influences may also be of great importance (Toland et al., 2008; Niu et al., 2009; Friso et al., 2015).

In recent decades, along with the analysis of the genome and transcriptome, metabolomic and proteomic studies have been developed. Metabolic profiles of biological tissues represent the influence on the metabolism of both genes and the environment, which makes it possible to obtain an integral assessment of multifactorial influences. Therefore, the search for metabolic markers, along with genetic ones, provides a more comprehensive picture of pathogenetic processes occurring in a particular person, and also allows clustering patients according to various forms of hypertensive conditions. Awareness of the metabolic pathways underlying a particular type of arterial hypertension would make the treatment protocols more efficient (Byrd, 2016).

Comprehensive metabolomic studies of arterial hypertension pathogenesis are still few in number. However, hypertensive patients were found to have specific changes in the lipid profile of blood serum (Brindle et al., 2003), changes in carbohydrate metabolism – an increase in glucose and galactose levels and a decrease in fructose concentration (Liu et al., 2011), an increase in the concentration of alpha-1-acid glycoprotein, a marker of inflammatory processes (De Meyer et al., 2008). Some data were also obtained on the metabolic profile in the strain of rats with spontaneous hypertension – SHR: an age-related decrease in the concentrations of certain amino acids (serine, methionine, ornithine, phenylalanine) and an increase in the content of free fatty acids in blood plasma (Aa et al., 2010), reduced in comparison with normotensive control rats urinary citrate and alpha-ketoglutarate levels at 8 weeks of age (Akira et al., 2008), increased urinary taurine and creatine at 12 and 26 weeks of age (Akira et al., 2005).

In the present study, for the first time, we analyzed the metabolic profile of blood serum in experimental animals with hereditary stress-sensitive arterial hypertension – ISIAH rats.

Materials and methods

Experimental animals. Male ISIAH rats with inherited stress-induced arterial hypertension (n = 10), control normotensive male WAG rats (Wistar Albino Glaxo) (n = 10), all aged 3–4 months. The experimental animals were kept under standard conditions in the conventional vivarium of the Institute of Cytology and Genetics (Siberian Branch of the Russian Academy of Sciences – SB RAS), receiving standard chow (Chara, Russia) and drinking water *ad libitum*. All procedures involving animals complied with the ethical standards approved by the legal acts of the Russian Federation, the principles of the Basel Declaration and the recommendations of the Inter-Institute Committee on Biological Ethics at the Institute of Cytology and Genetics (SB RAS) (protocol No. 127, September 8, 2022).

Blood pressure monitoring. Performed on a device for non-invasive blood pressure measurement (BIOPAC, USA) using the tail-cuff method, after preliminary adaptation of animals to this procedure for 3–4 days.

Blood serum sampling. Carried out during the euthanasia of experimental animals by decapitation. Collected peripheral blood was kept for an hour to form a primary clot, then cen-

trifuged (+4 °C, 3000 rpm, 20 min), the obtained blood serum was stored at –70 °C.

Extraction of metabolites from blood serum samples. Performed at the Research Equipment Sharing Center "Mass-spectrometric Studies" of the International Tomography Center (SB RAS), at the Laboratory of Proteomics and Metabolomics. Metabolites were extracted using a mixture of methanol-chloroform-water in the ratio of 1:1:1, according to a previously developed protocol (Zelentsova et al., 2020; Fomenko et al., 2022). The volume of serum for the study was 300 µl. The lyophilized extracts were diluted in 600 µl of deuterated phosphate buffer (50 mM, pH 7.2) with the addition of internal standard DSS (2×10^{-5} M 3-(trimethylsilyl) propane-1-sulfonate sodium).

NMR spectra. Obtained on the AVANCE III HD 700 MHz NMR spectrometer (Bruker BioSpin, Germany) equipped with an Ascend cryomagnet with a field strength of 16.44 Tesla. The survey parameters are described in earlier articles (Zelentsova et al., 2020; Fomenko et al., 2022). MestReNova v12.0 software was used to process the spectra and integrate the signals.

Identification of metabolites in the studied samples. Carried out using the Human Metabolome Database (https://hmdb.ca/) and our own data on the metabolic profiles of human and animal biological fluids (Tsentalovich et al., 2020; Fomenko et al., 2022).

Statistical processing of metabolomic data. Performed using the Statistica 8 software package (http://statsoft.ru/) and the MetaboAnalyst 5.0 web platform (https://www.metaboanalyst.ca/) (Pang et al., 2021), applying multivariate statistics (principal component analysis) and non-parametric method for assessing intergroup differences (Mann–Whitney U-test). Values at $p < 0.05$ were considered statistically significant.

Results

As a result of the NMR spectra analysis, the concentrations of 56 metabolites were determined in the blood serum of ISIAH (BP = 205.6 ± 7.3 mm Hg) and WAG (BP = 136.6 ± 3.1 mm Hg) rats. Significant interstrain differences in serum concentrations of 18 metabolites were observed (see the Table).

In ISIAH rats, the concentrations of leucine, isoleucine, valine, isobutyrate, glutamate, glutamine, asparagine, creatine phosphate, ornithine, myo-inositol, histidine, 1-methylhistidine, methionine sulfoxide in blood serum were significantly higher than in WAG rats, whereas the concentrations of 2-hydroxyisobutyrate, 2′-deoxyuridine, betaine, tryptophan, and tyrosine in ISIAH rats were decreased compared to normotensive controls.

In order to isolate metabolites that are associated with elevated blood pressure in ISIAH rats, a multivariate analysis was performed. Principal component analysis revealed two main factors (two axes) that together account for 47.2 % of the total variation in serum concentrations of the studied metabolites.

As can be seen from Fig. 1, the experimental animals were clustered in the space of two principal components on the basis of belonging to a hyper- or normotensive strain. The projections of these clusters on the axis of the first compo-

Serum metabolite concentrations in ISIAH and WAG rats

Metabolites (nmol/ml)	WAG					ISIAH				
	Mean	Median	Std. Dev.	Q1	Q3	Mean	Median	Std. Dev.	Q1	Q3
Leucine	106.29**	104.72	9.65	99.89	108.94	138.70	135.55	26.61	129.40	156.83
Isoleucine	58.66**	58.64	5.82	52.98	59.57	93.26	95.46	25.46	75.59	114.66
Valine	142.48**	139.02	13.54	134.73	145.72	207.97	210.07	44.06	177.35	238.86
Isobutirate	6.59**	6.34	1.02	5.67	7.13	8.55	8.89	1.61	7.75	9.76
2-hydroxyisobutirate	28.87*	28.61	3.18	25.85	31.46	16.29	16.28	3.30	13.85	17.63
Glutamate	159.25***	159.60	12.76	150.60	166.95	208.58	197.27	22.40	194.68	222.23
Glutamine	687.06**	694.13	54.04	664.59	711.75	894.53	968.68	147.72	773.79	1007.59
Asparagine	55.78*	51.00	12.07	47.29	67.57	72.87	76.02	13.92	59.06	78.80
2′-deoxyuridine	70.53***	73.91	7.65	65.59	76.82	49.77	50.29	8.99	43.98	54.92
Creatine phosphate	7.99**	8.24	2.70	6.92	10.30	21.11	22.68	9.75	16.72	27.83
Ornithine	47.34***	48.02	4.22	44.13	49.70	86.21	89.47	14.19	80.54	93.64
Betaine	139.64**	138.45	23.52	123.96	160.52	98.62	88.76	29.18	83.20	97.21
Myo-inositol	89.79***	88.96	8.78	83.53	98.59	132.74	127.83	25.71	114.09	143.01
Tryptophan	110.88***	110.77	11.31	101.11	119.81	85.64	86.93	6.27	80.82	91.54
Tyrosine	92.50**	92.87	8.39	84.27	98.39	78.42	77.15	10.17	71.00	85.05
Histidine	66.68**	66.43	5.90	61.44	71.38	76.00	77.21	6.72	70.80	80.52
1-methylhistidine	9.71**	10.67	1.88	8.11	11.04	19.05	16.59	10.09	12.00	21.19
Methionine sulfoxide	36.88***	37.56	3.63	33.76	39.72	45.16	45.56	2.18	43.40	46.45

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

nent practically do not overlap, while their projections on the axis of the second component coincide. Thus, the first principal component can be defined as the axis of presence/absence of hypertensive status. In order to establish a relationship between the concentrations of the detected metabolites and the hypertensive status, it is necessary to consider their distribution against the first principal component. This is determined by the "loadings" that metabolites make on the first component.

Positive loadings on the axis of the first component were made by 2-hydroxyisobutyrate, tryptophan, tyrosine, betaine, 2′-deoxyuridine; ornithine, valine, isoleucine, leucine, isobutyrate, glutamate, glutamine, asparagine, creatine phosphate, myo-inositol, histidine, 1-methylhistidine, methionine sulfoxide made negative loadings (Fig. 2). Thus, the listed metabolites are largely responsible for the clustering of groups of experimental animals according to the level of their blood pressure.

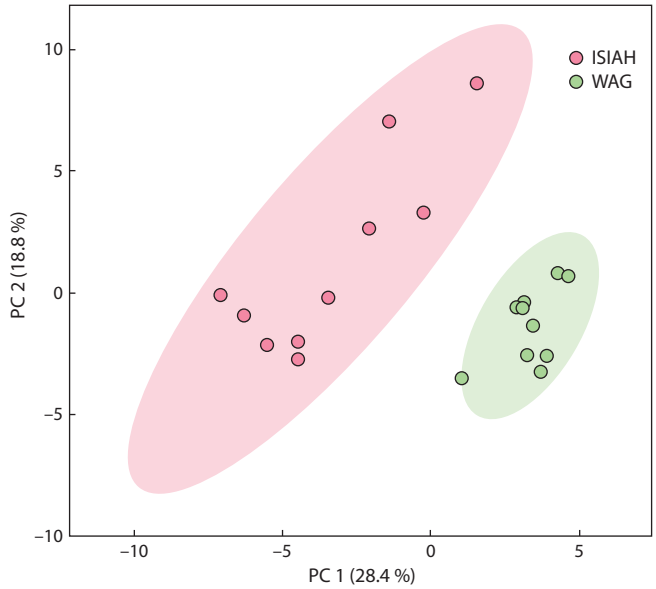


Fig. 1. Location of hypertensive ISIAH rats and normotensive WAG rats in principal component coordinates (PC 1 and PC 2) obtained by analyzing the metabolic profile of blood serum using the MetaboAnalyst 5.0 web platform.

Discussion

BCAA, branched-chain amino acids

Amino acids of the BCAA group – leucine, isoleucine and valine – are essential, and participate in the protein synthesis and degradation. They are also signal molecules in glucose metabolism, activating the mTORC1 complex, which phosphorylates the insulin receptor substrate IRS-1 (Yoshizawa, 2012; Yoon, 2016). Elevated plasma concentrations of BCAA amino acids have been associated with obesity, insulin resistance, decreased glucose tolerance, and type 2 diabetes, according to a number of studies (Newgard et al., 2009; Wang T.J. et al., 2011; Roberts et al., 2014). It has also been found that leucine, isoleucine, and valine are involved in the hypothalamic regulation of glucose metabolism in the liver (Arrieta-Cruz et al., 2016).

Prospective cohort studies involving a large number of patients (2243 (Hu et al., 2016) and 27,041 (Tobias et al., 2018)) show that elevated plasma concentrations of BCAA amino acids positively correlate with the risk of developing cardiovascular diseases (stroke, myocardial infarction, coronary disease). In ISIAH rats, a decrease in the level of immunoreactive insulin in the blood and glucose tolerance has been previously found, probably due to a genetically determined increased activity of the sympathoadrenal and thyroid systems (Shorin et al., 1990; Buzueva et al., 2006). Activation of the pancreas sympathetic innervation reduces insulin production by β-cells, acting through α_2-adrenergic receptors (van Duk et al., 1995), and thyroid hormones af-

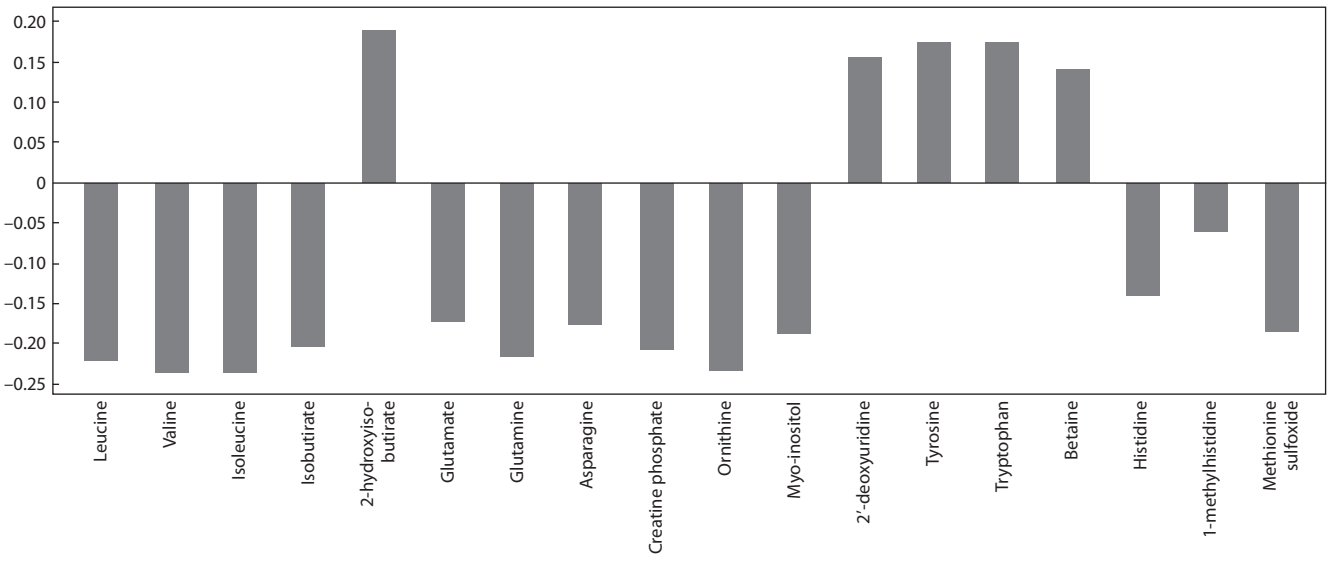


Fig. 2. Loadings on the first principal component made by those metabolites, the concentrations of which in the blood serum significantly differed in ISIAH and WAG rats.

fect insulin production through the regulation of insulin-like growth factor 1 secretion (Cavaliere et al., 1987). These data are consistent with the results of the present study: the concentrations of leucine, isoleucine, and valine in blood serum are significantly increased in ISIAH rats compared to controls (see the Table), which suggests that amino acids of the BCAA group can be considered as metabolic markers of hereditary stress-induced arterial hypertension.

Myo-inositol

Some inositol isomers (particularly myo-inositol) have insulin-like properties and may reduce insulin resistance in patients with metabolic syndrome (Giordano et al., 2011; Croze, Soulage, 2013). It has been shown that myo-inositol plasma level is associated with BP level in patients with hypertension (Yang M. et al., 2016), and the use of myo-inositol as part of a dietary supplement for six months reduced the concentration of cardiovascular diseases biomarkers in menopausal women and in women with a history of metabolic syndrome (D'Anna et al., 2014). It is assumed that inositol derivatives affect the IP_3 receptor, which regulates the contractility of the smooth muscle walls of blood vessels through L-type calcium channels (Abou-Saleh et al., 2013). An increased level of myo-inositol in the blood serum of ISIAH rats with hereditary stress-induced hypertension may indicate its involvement in the pathogenesis of the hypertensive status of rats of this strain.

SCFA, short-chain fatty acids

Short-chain fatty acids – formic, acetic, propionic, butyric, isobutyric, valeric, isovaleric and others – are produced in the large intestine during fiber fermentation, being an important source of energy for colonocytes, and having anti-inflammatory and antitumor properties (Andoh et al., 2003; Fernández et al., 2016). Short-chain fatty acids entering into acylation reactions can modify histones, thus regulating the expression of genes involved in the mechanisms of development of the metabolic syndrome, type 2 diabetes, and ischemic tissue damage (Sabari et al., 2017; Chen et al., 2020). Decreased production of short-chain fatty acids produced by gut bacteria leads to intestinal dysfunction, inflammation, kidney failure, and, as a result, to increased blood pressure (Kim et al., 2018; Felizardo et al., 2019). In SHR rats with spontaneous hypertension, elevated BP has been associated with a reduced content of acetate and butyrate-producing bacteria in the intestinal microbiota (Yang T. et al., 2015).

The relationship between BP levels and various acids of the SCFA group in salt-sensitive Dahl rats has also been studied: a high salt load resulted in an increase in the concentration of acetate, propionate and isobutyrate in fecal samples (Bier et al., 2018). Mechanisms of this relationship have not yet been studied in detail, but there is evidence that short-chain fatty acids can affect vessels and kidneys through endothelial receptors associated with G-proteins, which leads to a change in BP levels (Natarajan et al., 2016). In hypertensive ISIAH rats, a change in the ratio of SCFAs and their derivatives was also observed when compared with the normotensive control: isobutyrate blood concentration was significantly increased, while 2-hydroxyisobutyrate levels were decreased (see the Table).

Glutamate, glutamine

Associations of glutamate and glutamine concentrations, as well as hepatic aspartate aminotransferase activity, with insulin resistance and the development of the metabolic syndrome have been shown (Sookoian, Pirola, 2012). There is also evidence that plasma glutamate is positively correlated with blood pressure, body mass index, insulin and triglyceride levels. The glutamine/glutamate ratio is inversely related to these parameters (Liu X. et al., 2019). Considering that ISIAH rats in this study have increased serum levels of both glutamate and glutamine when compared to control WAG rats, but the glutamine concentration (894.53 nmol/ml) is several times higher than the glutamate content (208.58 nmol/ml), interpretation of observed interstrain differences in glutamate and glutamine concentrations requires more research.

Glutamate and glutamine also contribute to the metabolism of arginine and ornithine, which are involved in the urea and nitric oxide cycle (Wilson et al., 2001). Ornithine concentration in the serum of ISIAH rats is also increased compared to the control. It is established that α-difluoromethylornithine administration resulted in the restoration of endothelial function and prevented an increase in blood pressure in spontaneously hypertensive SHR rats (Demougeot et al., 2005). In an earlier SHR study, α-difluoromethylornithine reduced the rate of aortic and caudal artery contraction in response to electrical stimulation and norepinephrine administration, while a decrease in arterial wall thickness and a decrease in the content of polyamines in vessels was also observed (Soltis et al., 1994).

Metabolites associated with inflammation

In a study involving healthy volunteers (323 people) and ischemic stroke patients (323 people), choline, like its metabolite betaine, was found to reduce the risk of cardiovascular complications (Zhong et al., 2021). Long-term use of choline and betaine as a dietary supplement was also shown to lower blood pressure in hypertensive patients (Golzarand et al., 2021). Intragastric administration of betaine to rats modeling pulmonary hypertension resulted in a decrease in right ventricular and pulmonary artery blood pressure, in a decrease in the degree of ventricular hypertrophy and in remodeling of the arterial wall, presumably due to anti-inflammatory action – betaine also reduced the levels of MCP-1, ET-1, NF-κB, TNF-α, IL-1β (Yang J.M. et al., 2018).

Tryptophan is an essential aromatic amino acid. In mammals, tryptophan is metabolized in three partially overlapping pathways. The main pathway – kynurenine pathway – includes oxidation and destruction of the indole ring, producing derivatives: kynurenic and anthranilic acids. One of the 60 tryptophan molecules is converted into nicotinic acid (vitamin B3, niacin). The second pathway is the serotonin pathway, where tryptophan is converted to serotonin and melatonin. The third pathway is the indole pathway, the formation of indole derivatives, which are then excreted in the urine (Richard et al., 2009). It has been shown that disorders in the links of the kynurenine pathway facilitate development of cardiovascular diseases, including an increase in blood pressure (Song et al., 2017; Verheyen et al., 2017). It is possible that tryptophan and kynurenine promote vasodilation through participation in the adenylate cyclase and guanylate cyclase systems of secondary intracellular messengers, triggering a cascade of

reactions leading to the activation of nitric oxide receptors and to a decrease in the concentration of Ca^{2+} ions in the smooth muscle walls of blood vessels (Lincoln et al., 1990; Stasch et al., 2006; Wang Y. et al., 2010).

Betaine and tryptophan concentrations were significantly reduced in the blood serum of hypertensive ISIAH rats compared with normotensive control, which may indicate that inflammatory processes play a role in establishing and maintaining the hypertensive status of ISIAH rats. Recently, there has been even more evidence of the important role of vascular wall inflammation in the pathogenesis of hypertensive conditions, including those involving interleukins IL-1β and IL-18 (Patrick et al., 2021).

Metabolites associated with energy processes

Creatine phosphate is a source of rapidly mobilized energy in tissues where energy metabolism is most intense – skeletal muscles, myocardium, brain. Due to the fact that direct transport of ATP across the mitochondrial membrane is difficult, creatine phosphate serves as a "shuttle", participating in the transport of chemical energy between mitochondria and energy-consuming areas. ATP with mitochondrial creatine kinase phosphorylates creatine to creatine phosphate, which goes, for example, to myofibrils. Myofibrillic creatine kinase forces creatine phosphate to phosphorylate ADP to ATP, producing creatine, which again returns to the mitochondria, and the cycle repeats (Bessman, Carpeneter, 1985).

Changes in the content and ratio of creatine and phosphocreatine in tissues can be a signal of various pathologies (Strumia et al., 2012). It has been shown that a decrease in the ratio of creatine phosphate/ATP correlates with the severity of heart failure (Neubauer et al., 1992) and with the severity of myocardial hypertrophy (Ye et al., 2001). It is also known that exogenous creatine phosphate administration has a cardioprotective effect on the ischemic myocardium (Scattolin et al., 1993; Azova et al., 2015; Zhang et al., 2015). In our study, in ISIAH rats, serum creatine phosphate concentration was increased nearly three-fold compared with the normotensive control. To explain this difference in peripheral concentrations of creatine phosphate, additional studies are required, including an assessment of creatine phosphate concentration and the ratio of creatine phosphate/ATP in the myocardium of hypertensive ISIAH rats.

Metabolites associated with the synthesis of catecholamines

Tyrosine is an aromatic amino acid from which, via enzyme tyrosine hydroxylase, catecholamines are synthesized: dopamine, adrenaline, norepinephrine. Catecholamines are the main effectors of the sympathoadrenal system, affecting cardiac output and vascular resistance (Lee et al., 2016). The main indicators of the sympathoadrenal system functions are catecholamine concentrations and tyrosine hydroxylase activity (Yamabe et al., 1973; Moura et al., 2005), but tyrosine concentration may also be considered as a marker of catecholamine synthesis disorders: for example, in a metabolomic study of urine samples from patients with hypertensive nephrosclerosis, a decrease in tyrosine and dopamine levels has been found (Ovrehus et al., 2019).

It has previously been shown that the production of epinephrine by the adrenal glands and norepinephrine in the brain is increased in ISIAH rats compared with WAG (Markel et al., 2007; Redina et al., 2021), which allows to suggest that the reduced serum tyrosine level in ISIAH rats is a marker of changes in catecholamine synthesis.

Conclusion

Thus, we conclude that the metabolic profile of blood serum, which indicates the presence of a stress-dependent form of arterial hypertension, can be described as follows: an increase in the concentrations of leucine, isoleucine, valine, myo-inositol, isobutyrate, glutamate, glutamine, ornithine, creatine phosphate, and a decrease in the concentrations of 2-hydroxyisobutyrate, betaine, tryptophan, tyrosine. Elevated concentrations of leucine, isoleucine, valine, and myo-inositol are associated with glucose metabolism and insulin resistance observed in ISIAH rats (Shorin et al., 1990; Pivovarova et al., 2020). Ornithine plays an important role in the urea synthesis, and is also associated with the metabolism of arginine and the production of vasoactive factor – nitric oxide; therefore, its consideration as a metabolic marker of hypertension pathogenesis seems to be quite reasonable. Betaine is described as having an anti-inflammatory effect in various pathologies (Zhao et al., 2018), therefore, a decrease in its concentration in the serum of ISIAH rats may indicate the involvement of the inflammatory process in the pathogenesis of arterial hypertension. Serum tryptophan may play the same role as a negative marker of the inflammatory process (Sorgdrager et al., 2019); its decrease in ISIAH rats may have a pro-inflammatory effect.

The results obtained are the starting point for a more detailed study on the association of these metabolic markers with the development of hypertensive status at certain stages of the stress-dependent arterial hypertension pathogenesis.

References

Aa J., Wang G., Hao H., Huang Q., Lu Y., Yan B., Zha W., Liu L., Kang A. Differential regulations of blood pressure and perturbed metabolism by total ginsenosides and conventional antihypertensive agents in spontaneously hypertensive rats. *Acta Pharmacol. Sin*. 2010;31(8):930-937. DOI 10.1038/aps.2010.86.

Abou-Saleh H., Pathan A.R., Daalis A., Hubrack S., Abou-Jassoum H., Al-Naeimi H., Rusch N.J., Machaca K. Inositol 1, 4, 5-trisphosphate (IP_3) receptor up-regulation in hypertension is associated with sensitization of Ca^{2+} release and vascular smooth muscle contractility. *J. Biol. Chem*. 2013;288(46):32941-32951. DOI 10.1074/jbc.M113.496802.

Akira K., Imachi M., Hashimoto T. Investigations into biochemical changes of genetic hypertensive rats using ^{1}H nuclear magnetic resonance-based metabonomics. *Hypertens. Res*. 2005;28(5):425-430. DOI 10.1291/hypres.28.425.

Akira K., Masu S., Imachi M., Mitome H., Hashimoto M., Hashimoto T. ^{1}H nuclear magnetic resonance-based metabonomic analysis of urine from young spontaneously hypertensive rats. *J. Pharm. Biomed. Anal*. 2008;46(3):550-556. DOI 10.1016/j.jpba.2007.11.017.

Andoh A., Tsujikawa T., Fujiyama Y. Role of dietary fiber and short-chain fatty acids in the colon. *Curr. Pharm. Des*. 2003;9(4):347-358. DOI 10.2174/1381612033391973.

Arrieta-Cruz I., Su Y., Gutiérrez-Juárez R. Suppression of endogenous glucose production by isoleucine and valine and impact of diet composition. *Nutrients*. 2016;8(2):79. DOI 10.3390/nu8020079.

Azova M.M., Blagonravov M.L., Frolov V.A. Effect of phosphocreatine and ethylmethylhydroxypyridine succinate on the expression of Bax and Bcl-2 proteins in left-ventricular cardiomyocytes of spon-

taneously hypertensive rats. *Bull. Exp. Biol. Med*. 2015;158(3):313-314. DOI 10.1007/s10517-015-2749-4.

Bessman S.P., Carpenter C.L. The creatine-creatine phosphate energy shuttle. *Annu. Rev. Biochem*. 1985;54(1):831-862. DOI 10.1146/annurev.bi.54.070185.004151.

Bier A., Braun T., Khasbab R., Di Segni A., Grossman E., Haberman Y., Leibowitz A. A high salt diet modulates the gut microbiota and short chain fatty acids production in a salt-sensitive hypertension rat model. *Nutrients*. 2018;10(9):1154. DOI 10.3390/nu10091154.

Brindle J.T., Nicholson J.K., Schofield P.M., Grainger D.J., Holmes E. Application of chemometrics to ^{1}H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. *Analyst*. 2003;128(1):32-36. DOI 10.1039/B209155K.

Buzueva I.I., Filyushina E.E., Shmerling M.D., Markel A.L., Yakobson G.S. Age-related structural characteristics of the adrenal medulla in hypertensive NISAG rats. *Bull. Exp. Biol. Med*. 2006;142(6):651-653. DOI 10.1007/s10517-006-0441-4.

Byrd J.B. Personalized medicine and treatment approaches in hypertension: current perspectives. *Integr. Blood Press. Control*. 2016;9:59-67. DOI 10.2147/IBPC.S74320.

Cavaliere H., Meyer K., Geraldo M.-N. Effect of thyroid hormone therapy on plasma insulin-like growth factor I levels in normal subjects, hypothyroid patients and endemic cretins. *Horm. Res. Paediatr*. 1987;25(3):132-139. DOI 10.1159/000180644.

Chen X.F., Chen X., Tang X. Short-chain fatty acid, acylation and cardiovascular diseases. *Clin. Sci*. 2020;134(6):657-676. DOI 10.1042/CS20200128.

Croze M.L., Soulage C.O. Potential role and therapeutic interests of myo-inositol in metabolic diseases. *Biochimie*. 2013;95(10):1811-1827. DOI 10.1016/j.biochi.2013.05.011.

D'Anna R., Santamaria A., Cannata M.L., Interdonato M.L., Giorgianni G.M., Granese R., Corrado F., Bitto A. Effects of a new flavonoid and Myo-inositol supplement on some biomarkers of cardiovascular risk in postmenopausal women: a randomized trial. *Int. J. Endocrinol*. 2014;2014:653561. DOI 10.1155/2014/653561.

De Meyer T., Sinnaeve D., Van Gasse B., Tsiporkova E., Rietzschel E.R., De Buyzere M.L., Gillebert T.C., Bekaert S., Martins J.C., Van Criekinge W. NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Anal. Chem*. 2008;80(10):3783-3790. DOI 10.1021/ac7025964.

Demougeot C., Prigent-Tessier A., Marie C., Berthelot A. Arginase inhibition reduces endothelial dysfunction and blood pressure rising in spontaneously hypertensive rats. *J. Hypertens*. 2005;23(5):971-978. DOI 10.1097/01.hjh.0000166837.78559.93.

Felizardo R.J.F., Watanabe I.M., Dardi P., Rossoni L.V., Câmara N.O.S. The interplay among gut microbiota, hypertension and kidney diseases: The role of short-chain fatty acids. *Pharmacol. Res*. 2019;141:366-377. DOI 10.1016/j.phrs.2019.01.019.

Fernández J., Redondo-Blanco S., Gutiérrez-del-Río I., Miguélez E.M., Villar C.J., Lombo F. Colon microbiota fermentation of dietary prebiotics towards short-chain fatty acids and their roles as anti-inflammatory and antitumour agents: A review. *J. Funct. Foods*. 2016;25:511-522. DOI 10.1016/j.jff.2016.06.032.

Fomenko M.V., Yanshole L.V., Tsentalovich Y.P. Stability of metabolomic content during sample preparation: blood and brain tissues. *Metabolites*. 2022;12(9):811. DOI 10.3390/metabo12090811.

Friso S., Carvajal C.A., Fardella C.E., Olivieri O. Epige netics and arterial hypertension: the challenge of emerging evidence. *Transl. Res*. 2015;165(1):154-165. DOI 10.1016/j.trsl.2014.06.007.

Giordano D., Corrado F., Santamaria A., Quattrone S., Pintaudi B., Di Benedetto A., D'Anna R. Effects of myo-inositol supplementation in postmenopausal women with metabolic syndrome: a perspective, randomized, placebo-controlled study. *Menopause*. 2011;18(1):102-104. DOI 10.1097/gme.0b013e3181e8e1b.

Golzarand M., Bahadoran Z., Mirmiran P., Azizi F. Dietary choline and betaine intake and risk of hypertension development: a 7.4-year follow-up. *Food Funct*. 2021;12(9):4072-4078. DOI 10.1039/D0FO03208E.

Hoffmann T.J., Ehret G.B., Nandakumar P., Ranatunga D., Schaefer C., Kwok P.Y., Iribarren C., Chakravarti A., Risch N. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet*. 2017;49(1):54-64. DOI 10.1038/ng.3715.

Hu W., Sun L., Gong Y., Zhou Y., Yang P., Ye Z., Fu J., Huang A., Fu Z., Yu W., Zhao Y., Yang T., Zhou H. Relationship between branched-chain amino acids, metabolic syndrome, and cardiovascular risk profile in a Chinese population: a cross-sectional study. *Int. J. Endocrinol*. 2016;2016:8173905. DOI 10.1155/2016/8173905.

Kim S., Goel R., Kumar A., Qi Y., Lobaton G., Hosaka K., Mohammed M., Handberg E.M., Richards E.M., Pepine C.J., Raizada M.K. Imbalance of gut microbiome and intestinal epithelial barrier dysfunction in patients with high blood pressure. *Clin. Sci*. 2018;132(6):701-718. DOI 10.1042/CS20180087.

Laurent S. Antihypertensive drugs. *Pharmacol. Res*. 2017;124:116-125. DOI 10.1016/j.phrs.2017.07.026.

Lee Y.H., Kim Y.G., Moon J.Y., Kim J.S., Jeong K.H., Lee T.W., Ihm C.G., Lee S.H. Genetic variations of tyrosine hydroxylase in the pathogenesis of hypertension. *Electrolyte Blood Press*. 2016;14(2):21-26. DOI 10.5049/EBP.2016.14.2.21.

Lincoln T.M., Cornwell T.L., Taylor A.E. cGMP-dependent protein kinase mediates the reduction of Ca^{2+} by cAMP in vascular smooth muscle cells. *Am. J. Physiol. Cell Physiol*. 1990;258(3):C399-C407. DOI 10.1152/ajpcell.1990.258.3.C399.

Liu X., Zheng Y., Guasch-Ferré M., Ruiz-Canela M., Toledo E., Clish C., Liang L., Razquin C., Corella D., Estruch R., Fito M., Gómez-Gracia E., Arós F., Ros E., Lapetra J., Fiol M., Serra-Majem L., Papandreou C., Martínez-González M.A., Hu F.B., Salas-Salvadó J. High plasma glutamate and low glutamine-to-glutamate ratio are associated with type 2 diabetes: case-cohort study within the PREDIMED trial. *Nutr. Metab. Cardiovasc. Dis*. 2019;29(10):1040-1049. DOI 10.1016/j.numecd.2019.06.005.

Liu Y., Chen T., Qiu Y., Cheng Y., Cao Y., Zha A., Jia W. An ultrasonication-assisted extraction and derivatization protocol for GC/TOFMS-based metabolite profiling. *Anal. Bioanal. Chem*. 2011;400:1405-1417. DOI 10.1007/s00216-011-4880-z.

Markel A.L., Redina O.E., Gilinsky M.A., Dymshits G.M., Kalashnikova E.V., Khvorostova Yu.V., Fedoseeva L.A., Jacobson G.S. Neuroendocrine profiling in inherited stress-induced arterial hypertension rat strain with stress-sensitive arterial hypertension. *J. Endocrinol*. 2007;195(3):439-450. DOI 10.1677/JOE-07-0254.

Moura E., Costa P.M.P., Moura D., Guimarães S., Vieira-Coelho M.A. Decreased tyrosine hydroxylase activity in the adrenals of spontaneously hypertensive rats. *Life Sci*. 2005;76(25):2953-2964. DOI 10.1016/j.lfs.2004.11.017.

Natarajan N., Hori D., Flavahan S., Steppan J., Flavahan N.A., Berkowitz D.E., Pluznick J.L. Microbial short chain fatty acid metabolites lower blood pressure via endothelial G protein-coupled receptor 41. *Physiol. Genomics*. 2016;48(11):826-834. DOI 10.1152/physiolgenomics.00089.2016.

Neubauer S., Krahe T., Schindler R., Horn M., Hillenbrand H., Entzeroth C., Mader H., Kromer E.P., Riegger G.A., Lackner K. 31P magnetic resonance spectroscopy in dilated cardiomyopathy and coronary artery disease. Altered cardiac high-energy phosphate metabolism in heart failure. *Circulation*. 1992;86(6):1810-1818. DOI 10.1161/01.CIR.86.6.1810.

Newgard C.B., An J., Bain J.R., Muehlbauer M.J., Stevens R.D., Lien L.F., Haqq A.M., Shah S.H., Arlotto M., Slentz C.A., Rochon J., Gallup D., Ilkayeva O., Wenner B.R., Yancy W.S., Eisenson H., Musante G., Surwit R.S., Millington D.S., Butler M.D., Svetkey L.P. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab*. 2009;9(4):311-326. DOI 10.1016/j.cmet.2009.02.002.

Niu W.Q., Zhao H.Y., Zhou L., Dai X.X., Wang D.Y., Cao J., Wang B. Interacting effect of genetic variants of angiotensin II type 1 receptor on susceptibility to essential hypertension in Northern Han Chinese. *J. Hum. Hypertens*. 2009;23(1):68-71. DOI 10.1038/jhh.2008.77.

Ovrehus M.A., Bruheim P., Ju W., Zelnick L.R., Langlo K.A., Sharma K., de Boer I.H., Hallan S.I. Gene expression studies and targeted metabolomics reveal disturbed serine, methionine, and tyrosine metabolism in early hypertensive nephrosclerosis. *Kidney Int. Rep*. 2019;4(2):321-333. DOI 10.1016/j.ekir.2018.10.007.

Pang Z., Chong J., Zhou G., de Lima Morais D.A., Chang L., Barrette M., Gauthier C., Jacques P.É., Li S., Xia J. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res*. 2021;49(W1):W388-W396. DOI 10.1093/nar/gkab382.

Patrick D.M., van Beusecum J.P., Kirabo A. The role of inflammation in hypertension: novel concepts. *Curr. Opin. Physiol.* 2021;19:92-98. DOI 10.1016/j.cophys.2020.09.016.

Pivovarova E.N., Borisova M.A., Markel A.L. Experimental model of metabolic syndrome accompanied by non-alcoholic fatty liver disease in hypertensive ISIAH rats using fructose load. *Pisma v Vavilovskii Zhurnal Genetiki i Selektsii = Letters to Vavilov Journal of Genetics and Breeding*. 2020;6(1):10-14. DOI 10.18699/Letters2020-6-02. (in Russian)

Redina O.E., Smolenskaya S.E., Polityko Y.K., Ershov N.I., Gilinsky M.A., Markel A.L. Hypothalamic norepinephrine concentration and heart mass in hypertensive ISIAH rats are associated with a genetic locus on chromosome 18. *J. Pers. Med*. 2021;11(2):67. DOI 10.3390/jpm11020067.

Richard D.M., Dawes M.A., Mathias C.W., Acheson A., Hill-Kapturczak N., Dougherty D.M. L-tryptophan: basic metabolic functions, behavioral research and therapeutic indications. *Int. J. Tryptophan Res*. 2009;2:45-60. DOI 10.4137/ijtr.s2129.

Roberts L.D., Koulman A., Griffin J.L. Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome. *Lancet Diabetes Endocrinol*. 2014;2(1):65-75. DOI 10.1016/S2213-8587(13)70143-8.

Sabari B.R., Zhang D., Allis C.D., Zhao Y. Metabolic regulation of gene expression through histone acylations. *Nat. Rev. Mol. Cell Biol*. 2017;18(2):90-101. DOI 10.1038/nrm.2016.140.

Scattolin G., Gabellini A., Desideri A., Formichi M., Caneve F., Corbara F. Diastolic function and creatine phosphate: an echocardiographic study. *Curr. Ther. Res*. 1993;54(5):562-571. DOI 10.1016/S0011-393X(05)80677-0.

Shorin Yu.P., Markel' A.L., Selyatitskaia V.G., Pal'chikova N.A., Grinberg P.M., Amstislavskii S.Ya. Endocrine-metabolic relations in rats with genetic arterial hypertension. *Bull. Exp. Biol. Med*. 1990;109(6):575-576. DOI 10.1007/BF00841441.

Soltis E.E., Newman P.S., Olson J.W. Eflornithine treatment in SHR: potential role of vascular polyamines and ornithine decarboxylase in hypertension. *Clin. Exp. Hypertens*. 1994;16(5):595-610. DOI 10.3109/10641969409067964.

Song P., Ramprasath T., Wang H., Zou M.H. Abnormal kynurenine pathway of tryptophan catabolism in cardiovascular diseases. *Cell. Mol. Life Sci*. 2017;74(16):2899-2916. DOI 10.1007/s00018-017-2504-2.

Sookoian S., Pirola C.J. Alanine and aspartate aminotransferase and glutamine-cycling pathway: their roles in pathogenesis of metabolic syndrome. *World J. Gastroenterol*. 2012;18(29):3775-3781. DOI 10.3748/wjg.v18.i29.3775.

Sorgdrager F.J., Naudé P.J., Kema I.P., Nollen E.A., Deyn P.P.D. Tryptophan metabolism in inflammaging: from biomarker to therapeutic target. *Front. Immunol*. 2019;10:2565. DOI 10.3389/fimmu.2019.02565.

Stasch J.P., Schmidt P.M., Nedvetsky P.I., Nedvetskaya T.Y., Kumar A.H.S., Meurer S., Deile M., Taye A., Knorr A., Lapp H., Müller H., Turgay Y., Rothkegel C., Tersteegen A., Kemp-Harper B., Müller-Esterl W., Schmidt H.H.W. Targeting the heme-oxidized nitric oxide receptor for selective vasodilatation of diseased blood vessels. *J. Clin. Invest*. 2006;116(9):2552-2561. DOI 10.1172/JCI28371.

Strumia E., Pelliccia F., D'Ambrosio G. Creatine phosphate: pharmacological and clinical perspectives. *Adv. Ther*. 2012;29(2):99-123. DOI 10.1007/s12325-011-0091-4.

Thoenes M., Neuberger H.R., Volpe M., Khan B.V., Kirch W., Böhm M. Antihypertensive drug therapy and blood pressure control in men and women: an international perspective. *J. Hum. Hypertens*. 2010;24(5):336-344. DOI 10.1038/jhh.2009.76.

Tobias D.K., Lawler P.R., Harada P.H., Demler O.V., Ridker P.M., Manson J.E., Cheng S., Mora S. Circulating branched-chain amino acids and incident cardiovascular disease in a prospective cohort of US women. *Circ. Genom. Precis. Med*. 2018;11(4):e002157. DOI 10.1161/CIRCGEN.118.002157.

Toland E.J., Saad Y., Yerga-Woolwine S., Ummel S., Farms P., Ramdath R., Frank B.C., Lee N.H., Joe B. Closely linked non-additive blood pressure quantitative trait loci. *Mamm. Genome*. 2008;19(3):209-218. DOI 10.1007/s00335-008-9093-1.

Tsentalovich Y.P., Zelentsova E.A., Yanshole L.V., Yanshole V.V., Odud I.M. Most abundant metabolites in tissues of freshwater fish pike-perch (*Sander lucioperca*). *Sci. Rep*. 2020;10(1):17128. DOI 10.1038/s41598-020-73895-3.

van Duk G., Scheurink A., Ritter S., Steffens A. Glucose homeostasis and sympathoadrenal activity in mercaptoacetate-treated rats. *Physiol. Behav*. 1995;57(4):759-764. DOI 10.1016/0031-9384(94)00323-8.

Verheyen N., Meinitzer A., Grübler M.R., Ablasser K., Kolesnik E., Fahrleitner-Pammer A., Belyavskiy E., Trummer C., Schwetz V., Pieske-Kraigher E., Voelkl J., Alesutan I., Catena C., Sechi L.A., Brussee H., von Lewinski D., März W., Pieske B., Pilz S., Tomaschitz A. Low-grade inflammation and tryptophan-kynurenine pathway activation are associated with adverse cardiac remodeling in primary hyperparathyroidism: the EPATH trial. *Clin. Chem. Lab. Med*. 2017;55(7):1034-1042. DOI 10.1515/cclm-2016-1159.

Wang T.J., Larson M.G., Vasan R.S., Cheng S., Rhee E.P., McCabe E., Lewis G.D., Fox C.S., Jacques P.F., Fernandez C., O'Donnell C.J., Carr S.A., Mootha V.M., Florez J.C., Souza A., Melander A., Clish C.B., Gerszten R.E. Metabolite profiles and the risk of developing diabetes. *Nat. Med*. 2011;17(4):448-453. DOI 10.1038/nm.2307.

Wang Y., Liu H., McKenzie G., Witting P.K., Stasch J.-P., Hahn M., Changsirivathanathamrong D., Wu B.J., Ball H.J., Thomas S.R., Kapoor V., Celermajer D.S., Mellor A.L., Keaney J.F. Jr., Hunt N.H., Stocker R. Kynurenine is an endothelium-derived relaxing factor produced during inflammation. *Nat. Med*. 2010;16(3):279-285. DOI 10.1038/nm.2092.

Wilson C.J., Lee P.J., Leonard J.V. Plasma glutamine and ammonia concentrations in ornithine carbamoyltransferase deficiency and citrullinaemia. *J. Inherit. Metab. Dis*. 2001;24(7):691-695. DOI 10.1023/A:1012995701589.

Yamabe H., De Jong W., Lovenberg W. Further studies on catecholamine synthesis in the spontaneously hypertensive rat: catecholamine synthesis in the central nervous system. *Eur. J. Pharmacol*. 1973;22(1):91-98. DOI 10.1016/0014-2999(73)90188-X.

Yang J.M., Zhou R., Zhang M., Tan H.R., Yu J.Q. Betaine attenuates monocrotaline-induced pulmonary arterial hypertension in rats via inhibiting inflammatory response. *Molecules*. 2018;23(6):1274. DOI 10.3390/molecules23061274.

Yang M., Yu Z., Deng S., Chen X., Chen L., Guo Z., Zheng H., Chen L., Cai D., Wen B., Wu Q., Liang F. A targeted metabolomics MRM-MS study on identifying potential hypertension biomarkers in human plasma and evaluating acupuncture effects. *Sci. Rep*. 2016;6(1):25871. DOI 10.1038/srep25871.

Yang T., Santisteban M.M., Rodriguez V., Li E., Ahmari N., Carvajal J.M., Zadeh M., Gong M., Qi Y., Zubcevic J., Sahay B., Pepine C.J., Raizada M.K., Mohamadzadeh M. Gut dysbiosis is linked to hypertension. *Hypertension*. 2015;65(6):1331-1340. DOI 10.1161/HYPERTENSIONAHA.115.05315.

Ye Y., Gong G., Ochiai K., Liu J., Zhang J. High-energy phosphate metabolism and creatine kinase in failing hearts: a new porcine model. *Circulation*. 2001;103(11):1570-1576. DOI 10.1161/01.CIR.103.11.1570.

Yoon M.S. The emerging role of branched-chain amino acids in insulin resistance and metabolism. *Nutrients*. 2016;8(7):405. DOI 10.3390/nu8070405.

Yoshizawa F. New therapeutic strategy for amino acid medicine: notable functions of branched chain amino acids as biological regulators. *J. Pharmacol. Sci*. 2012;118(2):149-155. DOI 10.1254/jphs.11R05FM.

Zelentsova E.A., Yanshole L.V., Melnikov A.D., Kudryavtsev I.S., Novoselov V.P., Tsentalovich Y.P. Post-mortem changes in metabolomic profiles of human serum, aqueous humor and vitreous humor. *Metabolomics*. 2020;16(7):80. DOI 10.1007/s11306-020-01700-3.

Zhang W., Zhang H., Xing Y. Protective effects of phosphocreatine administered post-treatment combined with ischemic post-conditioning on rat hearts with myocardial ischemia/reperfusion injury. *J. Clin. Med. Res*. 2015;7(4):242-247. DOI 10.14740/jocmr2087w.

Zhao G., He F., Wu C., Li P., Li N., Deng J., Zhu G., Ren W., Peng Y. Betaine in inflammation: mechanistic aspects and applications. *Front. Immunol*. 2018;9:1070. DOI 10.3389/fimmu.2018.01070.

Zhong C., Miao M., Che B., Du J., Wang A., Peng H., Bu X., Zhang J., Ju Z., Xu T., He J., Zhang Y. Plasma choline and betaine and risks of cardiovascular events and recurrent stroke after ischemic stroke. *Am. J. Clin. Nutr*. 2021;114(4):1351-1359. DOI 10.1093/ajcn/nqab199.

ORCID ID

A.A. Seryapina orcid.org/0000-0002-8807-2580
Yu.K. Polityko orcid.org/0000-0003-2343-5085
L.V. Yanshole orcid.org/0000-0002-8265-6446
Yu.P. Tsentalovich orcid.org/0000-0002-1380-3000
A.L. Markel orcid.org/0000-0002-1550-1647

Acknowledgements. This work was supported by the Russian Science Foundation (project No. 22-25-20025) in cooperation with the Ministry of Science of the Novosibirskaya Oblast (agreement No. 000000540699598225120532 p-36, April 6, 2022).

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

Received November 29, 2022. Revised March 1, 2023. Accepted March 2, 2023.

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".

Сетевое издание «Вавиловский журнал генетики и селекции» – реестровая запись СМИ
Эл № ФС77-85772, зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 14 августа 2023 г.

Издание включено 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, Russian Science Citation Index на платформе Web of Science, Российский индекс научного цитирования, ВИНИТИ, Web of Science CC, Scopus, PubMed Central, DOAJ, ROAD, Ulrich's Periodicals Directory, Google Scholar.

Открытый доступ к полным текстам:
русскоязычная версия – на сайте ИЦиГ СО РАН, https://vavilovj-icg.ru/
и платформе Научной электронной библиотеки, elibrary.ru/title_about.asp?id=32440

англоязычная версия – на сайте vavilov.elpub.ru/index.php/jour
и платформе PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/journals/3805/

✉ email: 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.