Сетевое издание

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

СОДЕРЖАНИЕ • 2024 • 28 • 3

VAVILOV JOURNAL OF GENETICS AND BREEDING

CONTENTS • 2024 • 28 • 3

Assessment of the genetic diversity of the alleles of gliadin-coding loci in common wheat (*Triticum aestivum* L.) collections in Kazakhstan and Russia

M.U. Utebayev (iD)[1] ✉, S.M. Dashkevich (iD)[1], O.O. Kradetskaya (iD)[1], I.V. Chilimova (iD)[1], N.A. Bome (iD)[2]

[1] A.I. Barayev Research and Production Centre of Grain Farming, Shortandy-1, Akmola Region, Kazakhstan
[2] University of Tyumen, Tyumen, Russia
✉ phytochem@yandex.ru

Abstract. The study of genetic resources using prolamin polymorphism in wheat cultivars from countries with different climatic conditions makes it possible to identify and trace the preference for the selection of the alleles of gliadine-coding loci characteristic of specific conditions. The aim of the study was to determine the "gliadin profile" of the collection of common wheat (*Triticum aestivum* L.) from breeding centers in Russia and Kazakhstan by studying the genetic diversity of allelic variants of gliadin-coding loci. Intrapopulation ($\mu \pm S_\mu$) and genetic (H) diversity, the proportion of rare alleles ($h \pm S_h$), identity criterion (I) and genetic similarity (r) of common wheat from eight breeding centers in Russia and Kazakhstan have been calculated. It has been ascertained that the samples of common wheat bred in Kostanay region (Karabalyk Agricultural Experimental Station, Kazakhstan) and Chelyabinsk region (Chelyabinsk Research Institute of Agriculture, Russia) had the highest intrapopulation diversity of gliadin alleles. The proportion of rare alleles (h) at *Gli-B1* and *Gli-D1* loci was the highest in the wheat cultivars bred by the Federal Center of Agriculture Research of the South-East Region (Saratov region, Russia), which is explained by a high frequency of occurrence of *Gli-B1e* (86 %) and *Gli-D1a* (89.9 %) alleles. Based on identity criterion (I), the studied samples of common wheat from different regions of Kazakhstan and Russia have differences in gliadin-coding loci. The highest value of $I = 619.0$ was found when comparing wheat samples originated from Kostanay and Saratov regions, and the lowest $I = 114.4$, for wheat cultivars from Tyumen and Chelyabinsk regions. Some region-specific gliadin alleles in wheat samples have been identified. A combination of *Gli-A1f*, *Gli-B1e* and *Gli-Da* alleles has been identified in the majority of wheat samples from Kazakhstan and Russia. Alleles (*Gli-A1f*, *Gli-A1i*, *Gli-A1m*, *Gli-A1o*, *Gli-B1e*, *Gli-D1a*, *Gli-D1f*, *Gli-A2q*, *Gli-B2o*, and *Gli-D2a*) turned out to be characteristic and were found with varying frequency in wheat cultivars in eight regions of Russia and Kazakhstan. The highest intravarietal polymorphism (51.1 %) was observed in wheat cultivars bred in Omsk region (Russia) and the lowest (16.6 %), in Pavlodar region (Kazakhstan). On the basis of the allele frequencies, a "gliadin profile" of wheat from various regions and breeding institutions of Russia and Kazakhstan was compiled, which can be used for the selection of parent pairs in the breeding process, the control of cultivars during reproduction, as well as for assessing varietal purity.

Key words: gliadin-coding loci; genetic diversity; genetic similarity; common wheat; electrophoresis.

Оценка генетического разнообразия глиадинкодирующих локусов у образцов яровой пшеницы (*Triticum aestivum* L.), созданных в различных селекционных центрах Казахстана и России

М.У. Утебаев (iD)[1] ✉, С.М. Дашкевич (iD)[1], О.О. Крадецкая (iD)[1], И.В. Чилимова (iD)[1], Н.А. Боме (iD)[2]

[1] Научно-производственный центр зернового хозяйства им. А.И. Бараева, пос. Шортанды-1, Акмолинская область, Казахстан
[2] Тюменский государственный университет, Тюмень, Россия
✉ phytochem@yandex.ru

Аннотация. Изучение генетических ресурсов с использованием полиморфизма проламинов сортообразцов пшеницы из стран с различными климатическими условиями позволяет выявить и проследить предпочтительность отбора аллелей глиадинкодирующих локусов, характерных для конкретных условий. Цель исследования – определить

«глиадиновый профиль» коллекции яровой мягкой пшеницы (*Triticum aestivum* L.) из селекционных центров России и Казахстана на основе изучения генетического разнообразия аллельных вариантов глиадинкодирующих локусов. Проведен расчет внутрипопуляционного ($\mu \pm S_\mu$) и генетического (*H*) разнообразия, доли редких аллелей ($h \pm S_h$), критерия идентичности (*I*) и генетического сходства (*r*) яровой мягкой пшеницы из восьми селекционных центров России и Казахстана. Установлено, что наибольшим внутрипопуляционным разнообразием аллелей глиадина отличались образцы яровой мягкой пшеницы, созданные в Костанайской (Карабалыкская СХОС, Казахстан) и Челябинской (Челябинский НИИСХ, Россия) областях. Доля редких аллелей (*h*) по локусам *Gli-B1* и *Gli-D1* оказалась максимальной у сортов пшеницы селекции НИИСХ Юго-Востока (Саратовская область, Россия), что объясняется высокой частотой встречаемости аллелей *Gli-B1e* (86 %) и *Gli-D1a* (89.9 %). Статистически доказано, что изученные образцы яровой мягкой пшеницы из разных областей Казахстана и России отличаются друг от друга по глиадинкодирующим локусам на основе критерия идентичности (*I*). Наибольшее значение *I* = 619.0 установлено при сравнении образцов пшеницы, происходящих из Костанайской и Саратовской областей, а минимальное *I* = 114.4 отмечено для сортов пшеницы из Тюменской и Челябинской областей. Выявлены аллели глиадина, которые были идентифицированы только образцах, созданных в определенных регионах. Сочетание аллелей *Gli-A1f*, *Gli-B1e*, *Gli-Da* идентифицировано у большинства образцов пшеницы Казахстана и России. Аллели *Gli-A1f*, *Gli-A1i*, *Gli-A1m*, *Gli-A1o*, *Gli-B1e*, *Gli-D1a*, *Gli-D1f*, *Gli-A2q*, *Gli-B2o* и *Gli-D2a* оказались характерными и с различной частотой встречались в сортах пшеницы восьми областей России и Казахстана. Наибольший внутрисортовой полиморфизм (51.1 %) наблюдался у сортов пшеницы селекции СибНИИСХ (Омская область, Россия), а наименьший (16.6 %) – у образцов Павлодарской СХОС (Павлодарская область, Казахстан). На основе частот встречаемости аллелей составлен «глиадиновый профиль» пшеницы из разных областей и селекционных учреждений России и Казахстана, который может быть использован для подбора родительских пар в селекционном процессе, контроле сортов при репродукции, а также для установления сортовой чистоты.

Ключевые слова: глиадинкодирующие локусы; генетическое разнообразие; генетическое сходство; мягкая пшеница; электрофорез.

Introduction

Over the decades, scientists have found that the use of electrophoresis of the wheat storage protein, gliadin, is one of the methods that make it possible to distinguish cultivars from each other (Autran et al., 1979; Watry et al., 2020). Differences in gliadin spectra are associated with the presence of allelic diversity of genes localized at the main loci: *Gli-A1*, *Gli-B1*, *Gli-D1*, *Gli-A2*, *Gli-B2*, *Gli-D2*. Locus alleles control the synthesis of several gliadin components, which are inherited linked together and form a block. At the same time, gliadin blocks may differ from each other in the number, intensity, electrophoretic mobility and molecular weight of the components (Sozinov, Poperelya, 1980).

Based on the study of the world wheat collection, allelic blocks of gliadin were identified and cataloged for common wheat (Metakovsky et al., 2018) and durum wheat (Melnikova et al., 2012). It has been established that the cultivars created in different breeding centers can be similar to each other in some alleles of gliadin-coding loci (Novoselskaya-Dragovich et al., 2011; Melnikova et al., 2012), despite the fact that no special allele selection was performed. The reason for this is probably the linkage of these alleles with genes or groups of genes that affect the selection-relevant traits of wheat (Xynias et al., 2006); it may also be due to the involvement in the breeding process of the same genotype ("masterpiece cultivar"), valuable for many biological and economic traits, such as: Saratovskaya 29, Bezostaya 1, Mironovskaya 808, etc. Therefore, frequently occurring gliadin alleles in the samples created for specific climatic conditions can be used in the identification of cultivars and as markers of valuable traits in the breeding process, such as grain qualities and resistance to abiotic factors (Sozinov, 1985).

The data obtained on the basis of the polymorphism of storage proteins may not be inferior in informativeness to DNA markers. An additional advantage of using such markers for plant breeding is inexpensive equipment and ease of analysis. The analysis of the prolamins composition is still used in the identification of crop cultivars, i. e. alfalfa (Kakaei, Ahmadian, 2021), millet (Ma et al., 2022) and rice (Kaur et al., 2023), in the study of genetic control of the synthesis of storage proteins in oats (Lyubimova et al., 2020). The method of electrophoresis of storage proteins is recommended for use in the identification of varietal material in the UPOV rules for barley (Barley, UPOV Code(s)…, 2018) and wheat (Wheat, UPOV Code(s)…, 2022). For the identification and registration of wheat samples created in the Russian Federation, a methodological guide for electrophoresis of storage proteins has been published (Laboratory Analysis…, 2013). The use of the electrophoresis method for the identification of wheat cultivars is prescribed in the state standard of the Republic of Kazakhstan (ST RK, 2018) and the Republic of Mali (MN-01-03, 2001).

The results of the studies of wheat based on protein polymorphism can become the basis for a strategy for selecting genotypes with a certain combination of gliadin alleles. At the same time, the study of genetic resources based on the polymorphism of wheat prolamins from countries with different climatic conditions makes it possible to identify and trace the preference for selection and to establish the gliadin "profile" of the cultivar characteristic of specific conditions. In previous studies, the predominant or "leading" alleles of wheat prolamins characteristic of Northern Kazakhstan and Russia have been identified (Utebayev et al., 2016, 2019a, 2021, 2022). However, it is important to determine the gliadin spectrum of common wheat, which is characteristic of a specific breeding institution in Kazakhstan and Russia. Such information reflects the direction of breeding, the intensity of involvement of wheat genotypes from other breeding institutions, and the likelihood of "genetic erosion".

In this regard, the aim of the study is to determine the characteristic "gliadin profile" of common wheat (*Triticum aestivum* L.) samples created in various breeding centers in

М.У. Утебаев, С.М. Дашкевич
О.О. Крадецкая, И.В. Чилимова, Н.А. Боме

Генетическое разнообразие аллелей глиадина у образцов
пшеницы из селекционных центров Казахстана и России

2024
28 · 3

Russia and Kazakhstan, based on the study and statistical calculation of the genetic diversity of allelic variants of gliadin-coding loci.

Materials and methods

The object of study was 347 (177 Russian and 170 Kazakh) cultivars and breeding lines of common wheat (Supplementary Material 1)[1], the gliadin spectra of which were described and published earlier (Dobrotvorskaya et al., 2009; Novoselskaya-Dragovich et al., 2013; Utebayev et al., 2016, 2019a, 2022).

Unfortunately, it was not possible to carry out a temporal periodization by years of creation for all cultivars and breeding lines. Therefore, the calculations were based on the principle of belonging of a particular sample to a breeding institution (region). The genetic formulas of gliadin from common wheat samples created in ten breeding institutions in Russia and Kazakhstan were analyzed (see Supplementary Material 1). In addition, the electrophoresis of gliadin of a new cultivar Tselinnaya Niva (Akmola region) was carried out, the formula of which was included in the total number of analyzed wheat samples. Gliadin spectra of wheat were obtained according to the method proposed by E.V. Metakovsky (Metakovsky, Novoselskaya, 1991), gliadins were identified according to the catalog of alleles of gliadin-coding loci (Metakovsky, 1991).

Gliadin loci are designated according to the wheat gene catalog: *Gli-A1*, *Gli-B1*, *Gli-D1*, *Gli-A2*, *Gli-B2* and *Gli-D2* (McIntosh et al., 2003). Loci alleles were denoted by letters of the Latin alphabet in the following sequence, i. e. the genetic formula of gliadin of the cultivar Chinese Spring: *Gli-A1a*, *Gli-B1a*, *Gli-D1a*, *Gli-A2a*, *Gli-B2a*, *Gli-D2a* has an abbreviated notation: ***a, a, a, a, a, a***; while the genetic formula of the cultivar Mironovskaya 808: *Gli-A1f*, *Gli-B1b*, *Gli-D1g*, *Gli-A2n*, *Gli-B2m*, *Gli-D2e*, in abbreviated form looks like: ***f, b, g, n, m, e***.

Statistical analysis. Intrapopulation diversity (μ), which demonstrates the frequency of different genotypes, was calculated according to L.A. Zhivotovsky (1991):

$$\mu = (\sqrt{p_1} + \ldots + \sqrt{p_n})^2,$$

where p is the frequency of alleles calculated by the formula: $p = n/N$, in which N is the sample size, n is the number of alleles of one locus in the cultivar (breeding line). With equal frequencies of all alleles of the locus $\mu = n$, with an uneven distribution of frequencies $\mu < n$, and with monomorphism $\mu = 1$. The standard error of μ was calculated using the formula: $S_\mu = \sqrt{\mu(n - \mu)/N}$; where n is the number of alleles of one locus.

The calculation of the proportion of rare alleles (h) was determined by the formula:

$$h_\mu = 1 - (\mu/n).$$

To calculate the standard error of the proportion of rare alleles, the following formula was used:

$$S_h = \sqrt{h(1 - h)/N}.$$

In pairwise comparison of a group of wheat samples of different origins, the similarity index (r) was used (Zhivotovsky, 1979):

$$r = \sum \sqrt{pq},$$

where p is the frequency of the allele in the first population; q is the frequency of the allele in the second population. The statistical error of the r indicator was expressed by the formula:

$$S_r = 0.5\sqrt{\frac{q_0 - r^2}{N_1} + \frac{p_0 - r^2}{N_2}}.$$

In the case when all the identified alleles are common in the compared groups, the error was calculated using the formula:

$$S_r = 0.5\sqrt{\frac{N_1 + N_2}{N_1 N_2}(1 - r)}.$$

Based on the similarity index (r), the identity criterion (I) was calculated:

$$I = \frac{8N_1 N_2}{N_1 + N_2}(1 - r).$$

At I exceeding the table value of χ^2 with a 95 % significance level, cultivar populations were considered to have a significant difference.

The degree of genetic diversity (H) is calculated according to M. Nei (1973):

$$H = 1 - \sum p_i^2.$$

Results and discussion

The cultivars and breeding lines selected for study are presented in Supplementary Material 1. It is known that not all breeding lines reach the level of a cultivar, and not all cultivars reach the level of regionalization, nevertheless, this study presents the wheat samples (cultivars and breeding lines) that in one way or another used to be or are valuable for breeding, regardless of the year of creation or zoning. With this in mind, we made an attempt to show the allelic diversity of gliadin-coding loci that is found in one or another breeding center in Russia and Kazakhstan.

Gli-1 loci

Kazakhstan. The number of identified alleles of *A1* locus in wheat from Pavlodar and Karaganda regions was nine, from Akmola and Kostanay regions, 12 and 14 alleles, respectively (Fig. 1, Table 1). According to *B1* locus, two alleles were identified in the wheat of Karaganda origin, four were from Pavlodar, five and six alleles were of Kostanay and Akmola regions, respectively.

According to *D1* locus, the greatest diversity was recorded in the wheat from Kostanay region (nine alleles), the minimum from Karaganda (two alleles), for Akmola and Pavlodar regions, six and four alleles were recorded, respectively. The analysis of the gliadin formulas showed that alleles found in wheat in one region were absent in another. Thus, *Gli-A1d* and *Gli-A1y* were identified only in Akmola wheat, and *Gli-A1n* in the sample from Karaganda region.

At the same time, in the wheat samples from Akmola, Kostanay, Pavlodar and Karaganda regions, *Gli-A1f*, *Gli-A1i*
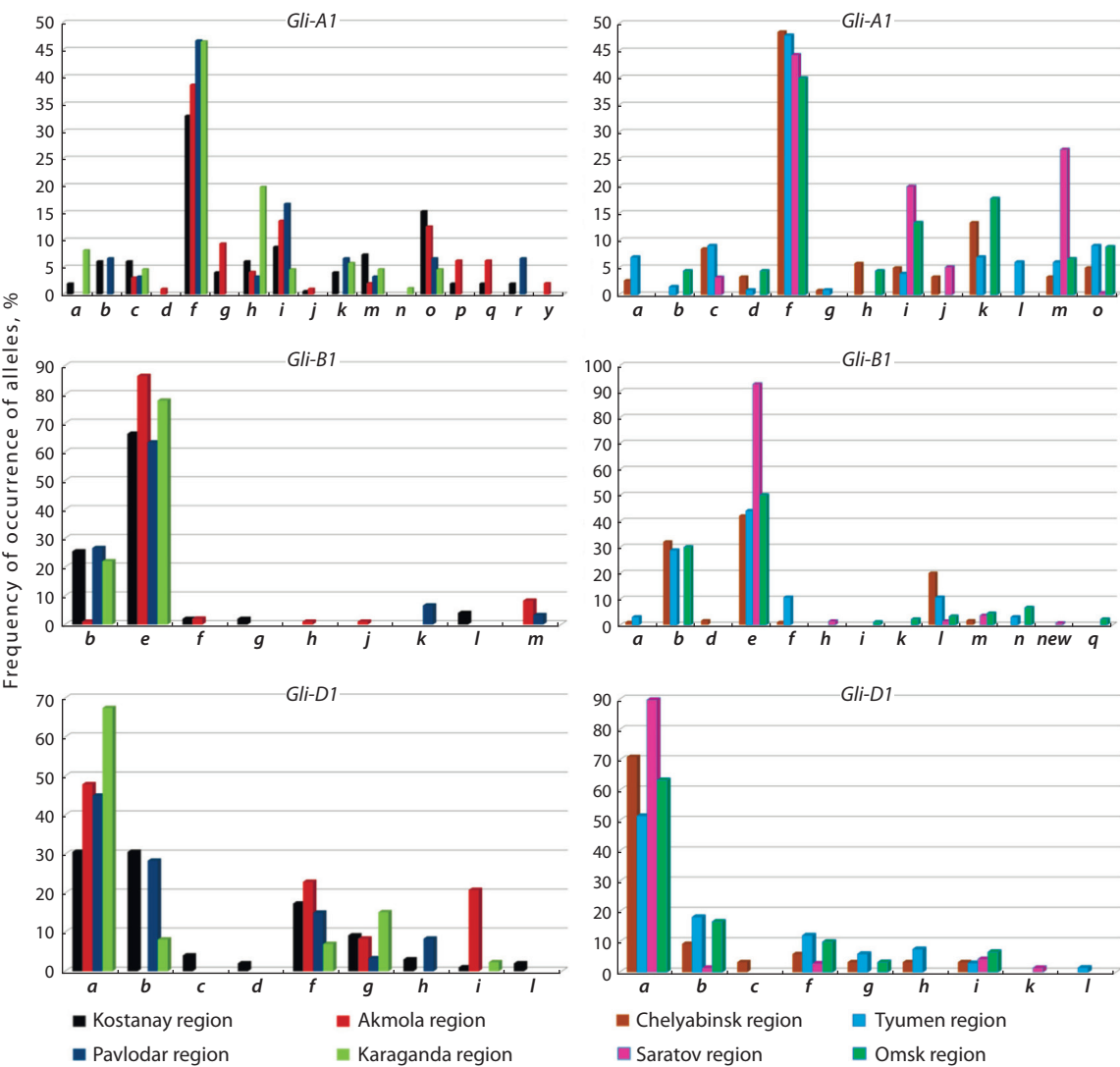
Fig. 1. Frequency of occurrence of alleles (%) of *Gli-1* loci of spring soft wheat by regions of Kazakhstan and Russia.

and *Gli-A1o* alleles were distributed at *Gli-A1* locus. *Gli-A1f* allele was common in all regions, its occurrence was 38.5 % in Akmola region, 32.9 % in Kostanay region, 46.7 % in Pavlodar region, and 46.5 % in Karaganda region (Fig. 1). The second most common allele in the wheat of Kostanay selection was *Gli-A1o* (15.3 %), and in Pavlodar, *Gli-A1i* with a frequency of 16.67 %. It should be noted that *Gli-A1o* and *Gli-A1i* alleles occur in the wheat from Akmola region with a frequency of 12.50 and 13.54 %, respectively.

It should be added that *Gli-A1h* allele identified in the wheat samples from Karaganda region with a frequency of 19.8 % was not widespread in other regions of Kazakhstan. On the other hand, the blocks of gliadin components controlled by *Gli-A1h* and *Gli-A1i* alleles are quite similar in the number and electrophoretic mobility of the components, differing only in the mobility of one component in the γ zone (Metakovsky, 1991).

Since each gliadin locus is characterized by multiple allelisms, it is not uncommon to have a polymorphism of a cultivar or line. That is, polymorphic samples are a mixture of caryopses that differ in alleles of one or more gliadin-

coding loci. *Gli-1* loci polymorphism was 27.9 % (12 out of 43 samples) for the samples from Karaganda region, 20.4 % (10 out of 49 samples) for those from Kostanay region, 18.7 % (9 out of 48 samples) and 13.3 % (4 out of 30 samples) for the wheat from Akmola and Pavlodar regions, respectively. It should be noted that the values given in Table 1 characterize the polymorphism of a single locus. In this regard, the largest polymorphism at *A1* and *D1* loci is from Karaganda region – 18.6 and 9.3 %, respectively. Note that the polymorphism at *B1* locus is more often represented by the combination *e+b*, while *A1* locus is more often represented by *f* allele in various combinations. The lowest allelic diversity of *Gli-B1* locus was observed in the wheat samples from Karaganda and Pavlodar regions – two and four alleles, respectively. In all regions, the highest percentage of occurrence was recorded for *Gli-B1e* allele (Fig. 1).

Gli-B1l found only in the samples of Lutescens 71 and Liniya 19CHS (Karabalyk Agricultural Experimental Station, Kostanay region) is of interest, since this locus is a marker of wheat-rye translocation. The genes included in this translocation control the plant's resistance to a number

М.У. Утебаев, С.М. Дашкевич
О.О. Крадецкая, И.В. Чилимова, Н.А. Боме

Генетическое разнообразие аллелей глиадина у образцов
пшеницы из селекционных центров Казахстана и России

2024
28 • 3

Table 1. Number of alleles and polymorphism of *Gli-* loci in varieties of spring soft wheat created in various breeding centers in Russia and Kazakhstan

Breeding center (region)	Gli-A1	Gli-B1	Gli-D1	Gli-A2	Gli-B2	Gli-D2	Total number, units	
	Number of alleles, units (polymorphism, %)						alleles	variety samples
Russian breeding centers (*N* = 177)								
Federal Center of Agriculture Research of the South-East Region (Saratov region)	6 (23.2)	5 (2.9)	5 (5.8)	10 (14.5)	13 (21.7)	9 (14.5)	48	69
Siberian Research Institute of Agriculture and Omsk State Agrarian University (Omsk region)	8 (22.2)	8 (13.3)	5 (6.7)	9 (13.3)	12 (15.5)	9 (8.9)	51	45
Chelyabinsk Research Institute of Agriculture (Chelyabinsk region)	11 (26.6)	7 (30.0)	7 (13.3)	12 (33.3)	15 (30.0)	17 (33.3)	69	30
Research Institute of Agriculture of the Northern Trans-Urals and State Agrarian University of the Northern Trans-Urals (Tyumen region)	11 (12.1)	6 (9.1)	7 (6.1)	11 (30.3)	14 (18.2)	12 (18.2)	61	33
Kazakhstan breeding centers (*N* = 170)								
A.I. Barayev Research and Production Centre of Grain Farming (Akmola region)	12 (10.4)	6 (6.2)	6 (8.3)	10 (8.3)	14 (2.1)	13 (4.2)	61	48
Karabalyk Agricultural Experimental Station (Kostanay region)	14 (6.1)	5 (10.2)	9 (8.2)	15 (6.1)	15 (4.1)	15 (4.1)	73	49
Pavlodar Agricultural Experimental Station (Pavlodar region)	9 (–)	4 (6.6)	4 (6.6)	9 (6.6)	13 (6.6)	9 (10.0)	48	30
Karaganda Agricultural Experimental Station named after A.F. Khristenko (Karaganda region)	9 (18.6)	2 (6.9)	2 (9.3)	10 (11.6)	12 (9.3)	8 (4.6)	43	43

of fungal diseases, such as various types of rust (brown, stem, yellow) and powdery mildew (Kozub et al., 2012). However, the presence of this translocation turned out to reduce the technological characteristics of the grain, which ultimately affects the baking quality of wheat (Sozinov, 1985). On the other hand, the negative effects of wheat-rye translocation can be neutralized by the presence of "good" glutenin subunits such as *1Dx5+1Dy10*, *1Bx7+1By9* and *1Bx7+1By8* (Sharma et al., 2018). It should be stated that the cultivar Lutescens 71 contains the components *1Dx5+1Dy10* and *1Bx7+1By9* in terms of the composition of high-molecular glutenin subunits (Utebayev et al., 2019b).

Gli-B1b allele is widely distributed among the studied samples, with the exception of the wheat from Akmola region. The low frequency of occurrence of this allele is probably due to the fact that most of the cultivars of A.I. Barayev Research and Production Centre of Grain Farming were created on the basis of the cultivars of Federal Center of Agriculture Research of the South-East Region (Saratov region), which are characterized by *Gli-B1e* allele (Novoselskaya-Dragovich et al., 2003).

The largest polymorphism at *Gli-D1* locus was observed for the wheat from Karaganda region, 9.3 %, and was expressed by the combination of *Gli-D1g+a*.

Gli-D1a, *Gli-D1f* and *Gli-D1g* alleles are common in the wheat from all four regions of Kazakhstan (Fig. 1). At the same time, *Gli-D1a* had the maximum frequency of occurrence. It should be noted that *Gli-D1a* and *Gli-D1f* alleles control gliadin blocks that are very similar in the number and electrophoretic mobility of components, with the exception

of the most mobile component located in the γ zone. There is an opinion that the less gliadin blocks differ in component composition, the closer they are to each other in terms of nucleotide composition (Chebotar et al., 2012). In this case, it can be assumed that the influence of such blocks on qualitative characteristics may be similar.

Russia. The number of identified alleles of *A1* locus in the wheat from Chelyabinsk and Tyumen regions was 11, from Saratov and Omsk regions, six and eight alleles, respectively. The largest number of identified alleles for *B1* locus was observed in the wheat from Omsk region – eight, the smallest one was in Saratov wheat – five (Fig. 1). Seven alleles were identified for *D1* locus in the wheat from Chelyabinsk and Tyumen regions, while five alleles were identified in the wheat from Saratov and Omsk regions.

The analysis of gliadin formulas showed that for each locus there were alleles characteristic only for the samples from one region, i. e. *Gli-B1h*, *Gli-B1new* and *Gli-D1k* alleles were found only in the wheat of Saratov selection (Dobrotvorskaya et al., 2009), *Gli-B1i*, *Gli-B1k* and *Gli-B1q* – Omsk selection (Novoselskaya-Dragovich et al., 2013), *Gli-A1l* and *Gli-D1l* – Tyumen selection, and *Gli-B1d* – Chelyabinsk region (Fig. 1).

The largest polymorphism of *Gli-1* loci was observed in the wheat of Chelyabinsk origin – 33.3 % (10 out of 30 samples), then in Omsk wheat – 31.1 % (14 out of 45 samples), in Saratov wheat – 26.1 % (18 out of 69 samples) and the smallest one, in Tyumen wheat – 18.2 % (6 out of 33 samples). It should be noted that such samples as Kukushka 12-6, Milturum 12013, Rossiyanka, Chelyabinskaya 17, Selivanovskiy Rusak and Omskaya 9 turned out to be polymorphic for all three *Gli-1*

loci, with the largest number of alleles per locus found in the cultivar Chelyabinskaya 17 (see Supplementary Material 1).

According to *A1* locus, the high frequency of occurrence of *Gli-A1f* allele was recorded in the wheat from Tyumen region – 47.8 %, from Chelyabinsk region – 48.5 %, from Saratov region – 44.3 %, and from Omsk region – 40.0 % (Fig. 1). It should be stated that the allele is common among Australian (Metakovsky et al., 1990), Iranian (Salavati et al., 2008), Ukrainian (Kozub et al., 2009) selection, as well as in the cultivars from Western and Eastern Siberia (Nikolaev et al., 2009) and may be associated with some economically valuable traits of wheat.

Gli-A1i, *Gli-A1m* and *Gli-A1o* alleles were also "common" (Fig. 1). As it turned out, *Gli-A1m* and *Gli-A1o* alleles make up the "gliadin profile" of the wheat from Canada, Mexico, Scandinavian countries, Spain, and China (Metakovsky et al., 2018).

According to *B1* locus, *Gli-B1e* allele "is in the lead" in the wheat of four regions, with different occurrence (Fig. 1). It should be added that *Gli-B1e* has a wide distribution area among the wheat cultivars of Kazakh and Russian selection (Novoselskaya-Dragovich et al., 2003; Nikolaev et al., 2009; Utebayev et al., 2019a). Also, according to *B1* locus, the largest number of alleles occurring in a certain region was identified as following: *Gli-B1d* in Chelyabinsk region, *Gli-B1h* and *Gli-B1new* in Saratov region, *Gli-B1i*, *Gli-B1k* and *Gli-B1q* in Omsk region. During the analysis of genealogies, it was found that the wheat from Federal Center of Agriculture Research of the South-East Region (Saratov region), for which *Gli-B1e* allele is characteristic, was actively involved in breeding when creating the wheat cultivars of Tyumen and Chelyabinsk selection (GRIS, 2017). In their turn, most of the cultivars of Federal Center of Agriculture Research of the South-East Region (Saratov region), in one way or another, originate from two cultivar-populations: the genetic formula of Poltavka is *Gli-A1o+f+c+j*, *Gli-B1e+m*, *Gli-D1a*, *Gli-A2q*, *Gli-B2o+s*, *Gli-D2e+a* and for Selivanovskiy Rusak it is *Gli-A1f+i+j**, *Gli-B1e+new*, *Gli-D1a+i*, *Gli-A2j+q+s*, *Gli-B2o+q*, *Gli-D2e+s* (Novoselskaya-Dragovich et al., 2003). Historically, most Kazakh cultivars are based on the cultivars from Saratov and Omsk regions, so it is quite expected that the gliadin profile of the wheat of the two countries is similar. Nevertheless, DNA diagnostics methods have proven the phylogenetic difference between Kazakh and Russian cultivars (Shavrukov et al., 2014). *Gli-B1b* allele was often found with the frequency of 32.0 % in the wheat from Chelyabinsk region, 28.8 %, from Tyumen region, and 30.0 %, from Omsk region. Since *Gli-B1b* is distributed from Scandinavian countries to Australia (Metakovsky et al., 2018), it is likely valuable for breeding.

At *Gli-D1* locus, the highest occurrence was found for *Gli-D1a* allele (Fig. 1). In addition, alleles such as *Gli-D1b*, *Gli-D1f* and *Gli-D1i* are common to four regions of Russia (Saratov, Omsk, Chelyabinsk, Tyumen). It is worth paying attention to *Gli-D1b* allele, which is characteristic of the wheat from France, Mexico, Portugal, Bulgaria, Serbia (Metakovsky et al., 2018), Iran (Salavati et al., 2008) and England (Chernakov, Metakovsky, 1994). Based on the study of proteolysis of wheat prolamins, it is proposed to use *Gli-D1b* together with *Gli-D1a* as markers of adaptability in spring bread wheat (Upelniek et al., 2003).

Gli-2 loci

Kazakhstan. When analyzing the genetic formulas of gliadin at *Gli-A2* locus, 10 alleles were identified in the wheat from Akmola and Karaganda regions. 15 and 9 alleles were identified in the wheat of Kostanay and Pavlodar selection, respectively. *B2* locus is represented by 12 alleles in Karaganda wheat, 13 were found in Pavlodar wheat, 14, in Akmola wheat, and 15, in the wheat of Kostanay selection. According to *D2* locus, eight alleles were identified in the wheat from Karaganda, nine, from Pavlodar, 13, from Akmola, and 15, from Kostanay regions (Table 1, Fig. 2).

Table 1 shows that the wheat of Karaganda selection is again "in the leading position" in terms of polymorphism of a single locus, since the values of *A2* and *B2* are the highest – 11.6 and 9.3 %, respectively. The "common" alleles were observed with varying frequencies in four regions of Kazakhstan: *Gli-A2b*, *Gli-A2l*, *Gli-A2q*, *Gli-B2a*, *Gli-B2f*, *Gli-B2l*, *Gli-B2m*, *Gli-B2t*, *Gli-D2a* and *Gli-D2q*.

The polymorphism of wheat by *Gli-2* loci was 11.6 % (5 out of 43 samples) for Karaganda region, 10.4 % (5 out of 48 samples) for Akmola region, 10.0 % (3 out of 30 samples) and 8.2 % (4 out of 49 samples) for Pavlodar and Kostanay regions, respectively. Such specimens as Karabalykskaya 9 (Kostanay region), Lutescens 65, Lutescens 261 (Pavlodar region), Lutescens 1220, Lutescens 1242 (Karaganda region) turned out to be polymorphic at three *Gli-2* loci.

The analysis of gliadin genetic formulas showed that the alleles *Gli-A2v* (2.3 %), *Gli-B2k* (2.3 %), *Gli-B2new* (2.3 %), *Gli-B2p* (4.6 %), *Gli-D2o* (2.3 %) and *Gli-D2r* (10.5 %) were found only in Karaganda wheat cultivars, and *Gli-A2h* (5.0 %), only in the samples of Pavlodar selection. Six alleles were identified in the samples from Kostanay region (*Gli-A2a*, *Gli-A2w*, *Gli-B2i*, *Gli-B2j*, *Gli-D2f* and *Gli-B2j*) and from Akmola region (*Gli-B2h*, *Gli-B2q*, *Gli-B2s*, *Gli-D2d* and *Gli-B2g*). At the same time, *Gli-B2s* allele with a frequency of 16.67 % is the second most common allele after *Gli-B2d* among the wheat of Akmola selection.

It should be stated that *Gli-A2l* allele, which occurs among Kazakh wheat samples, especially those from Karaganda region, turned out to be common among English (Chernakov, Metakovsky, 1994) and Iranian (Salavati et al., 2008) wheat samples. Also, *Gli-A2f* allele, which is the second most common wheat of Kostanay origin (15.31 %), was often found in Mexico and Portugal (Metakovsky et al., 2018). *Gli-A2q* allele, which has a high percentage of occurrence in Akmola and Pavlodar regions – 18.7 and 30.0 %, respectively, is of interest. It turned out that it is associated with the qualitative characteristics of grain, which are characteristic of strong cultivars of wheat (Dobrotvorskaya et al., 2009). On the other hand, it has been established that wheat genotypes with *Gli-A2q* allele have a long stem and low productivity (Khrunov et al., 2011).

Gli-B2s allele with a frequency of 16.7 %, identified only among the cultivars of Akmola region, constitutes the "profile" of the wheat of Saratov selection (Novoselskaya-Dragovich et al., 2003).

Gli-D2a allele, identified in the wheat samples from four regions of Kazakhstan, is widely distributed in common wheat cultivars from England (Chernakov, Metakovsky, 1994), Italy (Metakovsky et al., 1994), France (Metakovsky, Branlard, 1998), and Spain (Metakovsky et al., 2000). This is probably

М.У. Утебаев, С.М. Дашкевич
О.О. Крадецкая, И.В. Чилимова, Н.А. Боме

Генетическое разнообразие аллелей глиадина у образцов
пшеницы из селекционных центров Казахстана и России

2024
28•3
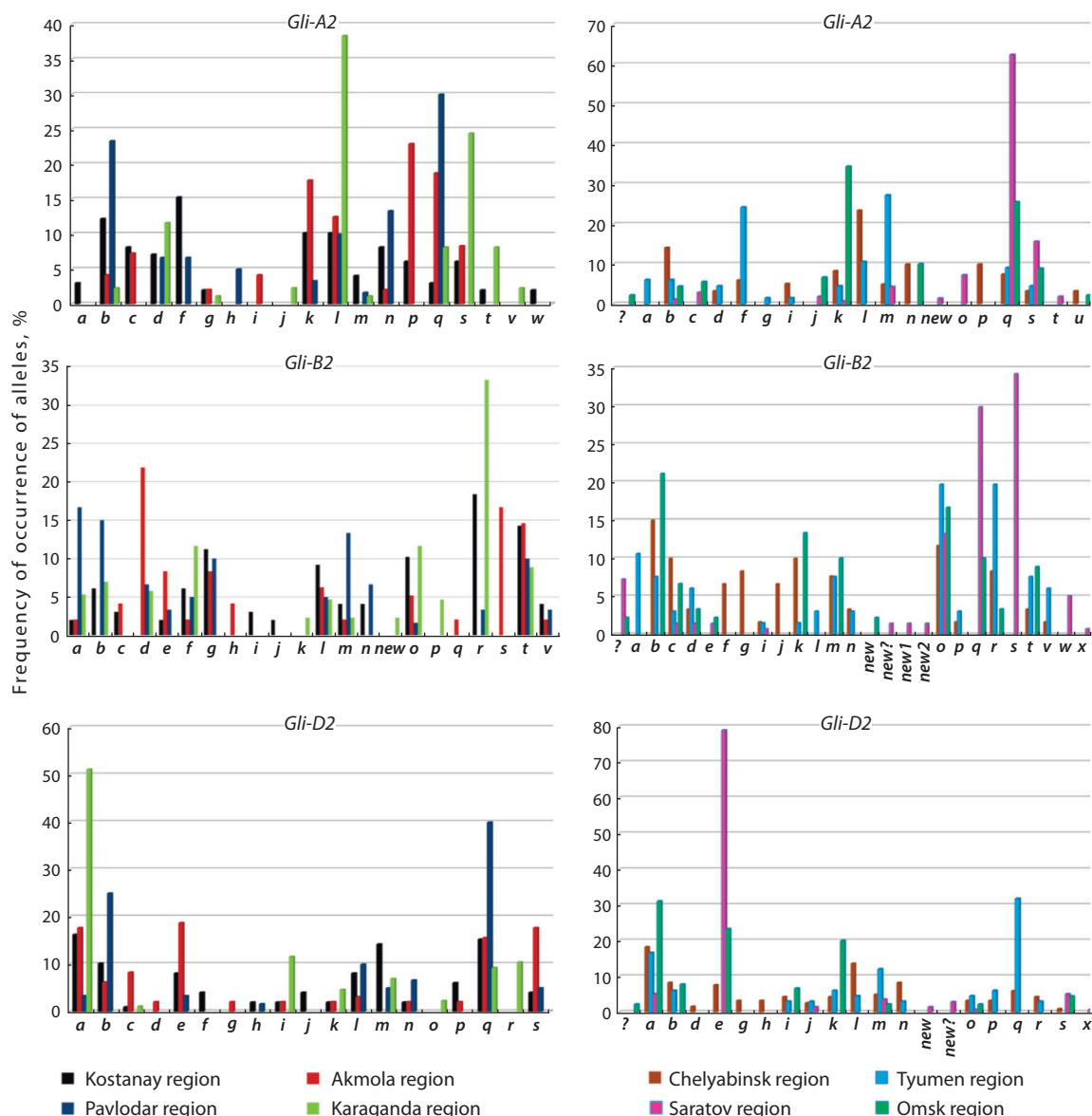
Fig. 2. Frequency of occurrence of alleles (%) of *Gli-2* loci of spring soft wheat by regions of Kazakhstan and Russia.

due to its association with adaptive traits, since the climate of European countries compared to Kazakhstan differs both in terms of precipitation, solar activity, and soil cover (Kunan-bayev et al., 2022). *Gli-D2q* allele, which is found in the wheat from Pavlodar region, is widespread in Australia (Metakovsky et al., 1990), which may also be associated with economically valuable traits.

Russia. According to *Gli-A2* locus, 12 alleles were identified in the wheat from Chelyabinsk region, 11, from Tyumen region, 10 and 9 alleles, from Saratov and Omsk regions, respectively. According to *B2* locus, genetic diversity is represented by 12 alleles in the wheat from Omsk region, 13 alleles, from Saratov region, 14 alleles, from Tyumen region, and 15 alleles, from Chelyabinsk region. According to *D2* locus, 17 and 12 alleles were identified in the wheat from Chelyabinsk and Tyumen regions, respectively, whereas 9 alleles were found in the wheat of Saratov and Omsk selection (Table 1, Fig. 2).

Gli-A2b, *Gli-A2k*, *Gli-A2q*, *Gli-A2s*, *Gli-B2c*, *Gli-B2d*, *Gli-B2o*, *Gli-D2a*, *Gli-D2m*, and *Gli-D2o* alleles with different frequencies turned out to be "common" for the wheat samples from the analyzed areas. Polymorphism for all *Gli-2* loci was stated for the wheat of Chelyabinsk origin at the level of 36.6 % (11 out of 30 samples), of Saratov origin – 34.8 % (24 out of 69 samples), of Omsk origin – 31.1 % (14 out of 45 samples) and of Tyumen origin – 30.3 % (10 out of 33 samples).

A high polymorphism of individual loci was observed for the wheat of Chelyabinsk origin: *Gli-A2* (33.3 %), *Gli-B2* (30.0 %) and *Gli-D2* (33.3 %); the lowest one was found in the wheat of Omsk selection: *Gli-A2* (13.3 %), *Gli-B2* (15.5 %) and *Gli-D2* (8.9 %) (Table 1).

Such samples as Kukushka 12-6, Milturum 12013, Rossiyanka, Uralskaya Kukushka, Chelyaba 2, Chelyabin-skaya 17, Erythrospermum 24841 (Chelyabinsk region), Tyumenskaya 30, Surenta 4, Surenta 6, Rechka, Lutescens 70,

Tyumenskaya Yubileynaya (Tyumen region), Lutescens 55-11, Saratovskaya 50, Selivanovskiy Rusak (Saratov region), Pamyati Azieva (Omsk region) are polymorphic for three *Gli-2* loci.

Based on the gliadin formulas, alleles that do not occur in other areas have been identified, i.e. eight alleles *Gli-A2p*, *Gli-A2u*, *Gli-B2f*, *Gli-B2g*, *Gli-B2j*, *Gli-D2d*, *Gli-D2g* and *Gli-D2h* have been identified only in the wheat of Chelyabinsk Research Institute of Agriculture; four alleles *Gli-A2a*, *Gli-A2g*, *Gli-B2a* and *Gli-B2l*, only in the cultivars from Tyumen region. 11 region-specific alleles *Gli-A2o*, *Gli-A2t*, *Gli-B2s*, *Gli-B2w* and *Gli-B2x* and several new alleles for each locus have been identified in the wheat from Saratov region (Dobrotvorskaya et al., 2009). Four alleles *Gli-A2u* and one new allele of *A2*, *B2* and *D2* loci were identified in the wheat of Omsk selection (Novoselskaya-Dragovich et al., 2013) (see Supplementary Material 1).

A high percentage of occurrence of *Gli-A2q* allele (62.5 %) was stated for the wheat of Federal Center of Agriculture Research of the South-East Region (Saratov region). At the same time, the highest allele frequency was identified for *Gli-A2k* (34.4 %) in the wheat of Omsk region and *Gli-A2l* (23.5 %) and *Gli-A2m* (27.3 %) for Chelyabinsk and Tyumen regions, respectively. It should be noted that *Gli-A2l* allele is common among the wheat from England (Chernakov, Metakovsky, 1994) and Iran (Salavati et al., 2008), and *Gli-A2m* allele, among the wheat from Canada and France (Metakovsky et al., 2018).

Gli-B2o allele turned out to be "common" for four regions of Russia. It has been stated that this allele is found in the wheat of Iranian (Salavati et al., 2008) and Italian (Metakovsky et al., 1994) origin, and in some cultivars of Saratov selection (Dobrotvorskaya et al., 2009), as well as in winter forms of wheat (Novoselskaya-Dragovich et al., 2015). In general, it should be added that according to *B2* locus, the wheat of Saratov selection has the largest number of unknown alleles (Dobrotvorskaya et al., 2009).

In the cultivars of Tyumen origin, a high frequency of occurrence was stated for the following alleles: *Gli-D2q* (31.8 %) and *Gli-D2a* (16.6 %), respectively; whereas for the wheat of Chelyabinsk Research Institute of Agriculture, *Gli-D2a* (18.3 %) and *Gli-D2l* (13.7 %) alleles are predominant (Fig. 2). It should be noted that *Gli-D2a* allele is probably associated with valuable traits, since it is widely distributed among Italian wheat cultivars (Metakovsky et al., 2018), and among Omsk cultivars its occurrence reaches 31.1 % (Novoselskaya-Dragovich et al., 2013).

Thus, in eight regions of Russia and Kazakhstan, the following alleles have become widespread: *Gli-A1f*, *Gli-A1i*, *Gli-A1m*, *Gli-A1o*, *Gli-B1e*, *Gli-D1a*, *Gli-D1f*, *Gli-A2q*, *Gli-B2o* and *Gli-D2a* (Supplementary Material 2). In the analysis of the general polymorphism, heterogeneity in all six gliadin-coding loci was identified for four wheat samples of Chelyabinsk Research Institute of Agriculture (Kukushka 12-6, Milturum 12013, Rossiyanka, Chelyabinskaya 17) and one cultivar from Federal Center of Agriculture Research of the South-East Region (Selivanovskiy Rusak). Polymorphisms at five gliadin loci (*A1*, *B1*, *A2*, *B2* and *D2*) were observed for the following samples: Karabalykskaya 9 (Karabalyk Agricultural Ex-
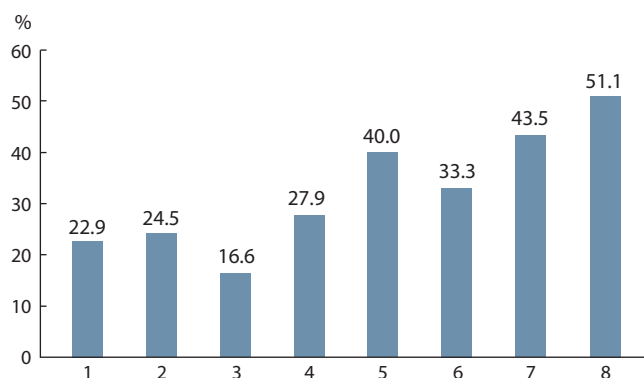
Fig. 3. Polymorphism of spring soft wheat samples from various regions of Kazakhstan and Russia.

1 – Akmola region (A.I. Barayev Research and Production Centre of Grain Farming); 2 – Kostanay region (Karabalyk Agricultural Experimental Station); 3 – Pavlodar region (Pavlodar Agricultural Experimental Station); 4 – Karaganda region (Karaganda Agricultural Experimental Station named after A.F. Khristenko); 5 – Chelyabinsk region (Chelyabinsk Research Institute of Agriculture); 6 – Tyumen region (Research Institute of Agriculture of the Northern Trans-Urals and State Agrarian University of the Northern Trans-Urals); 7 – Saratov region (Federal Center of Agriculture Research of the South-East Region); 8 – Omsk region (Siberian Research Institute of Agriculture and Omsk State Agrarian University).

perimental Station), Lutescens 1242 (Karaganda Agricultural Experimental Station named after A.F. Khristenko), Surenta 4 (Research Institute of Agriculture of the Northern Trans-Urals and State Agrarian University of the Northern Trans-Urals), Chelyaba 2 (Chelyabinsk Research Institute of Agriculture); heterogeneity of *B1*, *D1*, *A2*, *B2* and *D2* loci was stated for the cultivars Tyumenskaya 30 and Tyumenskaya Yubileynaya (Research Institute of Agriculture of the Northern Trans-Urals and State Agrarian University of the Northern Trans-Urals); the Lutescens 55-11 sample (Federal Center of Agriculture Research of the South-East Region) is polymorphic at *A1*, *D1*, *A2*, *B2* and *D2* loci, and the Omskaya 9 sample (Siberian Research Institute of Agriculture and Omsk State Agrarian University) is polymorphic at *A1*, *B1*, *D1*, *A2*, and *B2* loci (see Supplementary Material 1). The general polymorphism of the wheat samples depending on the origin is shown in Figure 3.

As can be seen, the greatest polymorphism was observed in the wheat of Omsk origin. It is believed that the presence of biotypes within the cultivar is an additional means to obtain a stable yield and increase its resistance to various environmental stress factors (Metakovsky et al., 2020).

Summarizing the results obtained by the frequencies of gliadin alleles, a "gliadin profile" of the wheat of Russian and Kazakh selection was compiled (Table 2). As can be seen, the combination of *Gli-1* loci alleles (*Gli-A1f*, *Gli-B1e* and *Gli-D1a*) is the same for eight regions, while *Gli-2* loci are different. The highest occurrence of the combination of *Gli-A1f*, *Gli-B1e* and *Gli-D1a* was in the samples of Saratov origin (33 out of 69 samples) and Karaganda origin (16 out of 43 samples) – 47.8 and 37.2 %, respectively; the lowest one was observed in the samples of Kostanay origin – 6.1 % (3 out of 49 samples). The maximum combination of *Gli-A1f* and *Gli-B1e* alleles found in the wheat from Akmola region (9 out of 48 samples) and Kostanay region (9 out of 49 samples) is

М.У. Утебаев, С.М. Дашкевич
О.О. Крадецкая, И.В. Чилимова, Н.А. Боме

Генетическое разнообразие аллелей глиадина у образцов
пшеницы из селекционных центров Казахстана и России

2024
28 • 3

Table 2. General "gliadin profile" of spring soft wheat created in various breeding centers from Russia and Kazakhstan

Region (Breeding center)	Gli- loci					
	A1	B1	D1	A2	B2	D2
Akmola (A.I. Barayev Research and Production Centre of Grain Farming)	f	e	a	p	d	a+e
Kostanay (Karabalyk Agricultural Experimental Station)	f	e	a+b	f	r	a+q
Pavlodar (Pavlodar Agricultural Experimental Station)	f	e	a	q	a+b	q
Karaganda (Karaganda Agricultural Experimental Station named after A.F. Khristenko)	f	e	a	l	r	a
Tyumen (Research Institute of Agriculture of the Northern Trans-Urals and State Agrarian University of the Northern Trans-Urals)	f	e	a	m+f	o+r	q
Chelyabinsk (Chelyabinsk Research Institute of Agriculture)	f	e	a	l	b	a
Saratov (Federal Center of Agriculture Research of the South-East Region)	f	e	a	q	s	e
Omsk (Siberian Research Institute of Agriculture and Omsk State Agrarian University)	f	e	a	k	b	a

18.8 and 18.4 %, respectively, the minimum one is 1.4 % in the samples from Saratov region (1 out of 69 samples), and there are none from Tyumen region.

The association of *Gli-B1e* and *Gli-D1a* was most often found in the wheat from Saratov region (27 out of 69 samples) and Akmola region (13 out of 48 samples) – 39.1 and 27.1 %, respectively; only 9.1 % (3 out of 33 samples) are from Tyumen region. The information on the relationship between gliadin alleles and grain quality indicators is contradictory, i. e. the presence of *Gli-A1m* allele has been shown to cause a decrease in flour sedimentation. Later, it turned out that in most cases it is associated with *Glu-A3e*, the "worst" allele. On the other hand, *Gli-A1m* is present in many high-quality cultivars of Canadian selection (Metakovsky et al., 2019).

It was found that *Gli-A2b* and *Gli-B2c* alleles were statistically related to W-energy of the pastry deformation determined on the alveograph (Metakovsky et al., 1997). Although it has been suggested that alleles encoded by *Gli-2* loci have a negative effect on grain quality (Masci et al., 2002), nonetheless, the use of *Gli-A2s* and *Gli-B2o* has been proposed as markers of increased protein, gluten content and grain nature (Khrunov et al., 2011). Later, on the basis of molecular genetic methods, the results were obtained indicating the presence of genes localized at *Gli-2* loci that have a positive effect on the rheological properties of pastry (Noma et al., 2019).

Gli-B1e allele constitutes the "gliadin profile" of many high-quality wheat cultivars of Russian and Kazakh selection (Novoselskaya-Dragovich et al., 2013; Utebayev et al., 2019a, 2022), which is probably due to the fact that it encodes the synthesis of the so-called ω-gliadin d4, which is associated with increased grain quality (Branlard et al., 2003).

It is worth noting that not all gliadin alleles, which are "positioned" as markers of quality grain, increase qualitative characteristics. Weather and climatic conditions play a significant corrective role in the grain formation. Therefore, till the present moment, there is no information about "universal" alleles, the presence of which would contribute to the production of high-quality wheat grain. Such controversies concerning the relationship between gliadin alleles and grain characteristics

contribute to an in-depth study of this phenomenon. On the other hand, the use of gliadin polymorphism for identification and determination of varietal purity does not lose its relevance due to the simplicity of execution and the constancy of the gliadin spectrum.

Statistical analysis

Based on the statistical calculations, the intrapopulation (μ) and genetic diversity (H) at *A1*, *D1* and *A2* loci turned out to be maximum for the wheat samples from Kostanay region; at *B2* and *D2* loci, for the samples from Chelyabinsk Research Institute of Agriculture, and at *B1*, for the cultivars from Tyumen region. The minimum values of μ and H were observed for the wheat from Akmola region at $Gli\text{-}B1 - 2.78 \pm 0.43$ and 0.24, respectively (Supplementary Material 3).

It turned out that the applied H indicator cannot always describe the genetic diversity of the population satisfactorily, since it "underestimates" rare alleles (alleles with a low frequency of occurrence in the population or cultivar). Therefore, the additional application of the μ parameter allows for a more accurate assessment of the degree of diversity by taking into account the number of rare alleles and their frequency, i. e. 11 alleles were identified at *Gli-A1* and *Gli-A2* loci in the set of the cultivars of Tyumen origin. At the same time, the intrapopulation diversity μ for *Gli-A1* locus was 8.00 ± 0.85, while for *Gli-A2* it was 9.12 ± 0.72. This difference is explained by the fact that one allele with a frequency of 0.47 was "in the lead" for *Gli-A1*, and two alleles with frequencies of 0.24 and 0.27 prevailed for *Gli-A2* locus. The applied indicator shows how variable the population is depending on the frequency of alleles.

The rare allele ratio criterion (h) characterizes the distribution of frequencies, which is always $h > 0$ in case of unevenness, compared to μ, which evaluates the degree of diversity of the population. Based on this, the genetic and intrapopulation diversity at *Gli-B1* and *Gli-D1* loci turned out to be the lowest for the wheat samples from Saratov region (see Supplementary Material 3). Such a low value is explained by the predominance of *Gli-B1e* allele over other alleles

Table 3. Average values of the proportion of rare alleles ($h \pm S_h$), genetic (H) and intrapopulation ($\mu \pm S_\mu$) diversity at the *Gli-1* and *Gli-2* loci

Region (Breeding center)	H	$\mu \pm S_\mu$	$h \pm S_h$
Akmola (A.I. Barayev Research and Production Centre of Grain Farming)	0.72	7.78 ± 0.55	0.22 ± 0.05
Kostanay (Karabalyk Agricultural Experimental Station)	0.80	10.15 ± 0.62	0.19 ± 0.05
Pavlodar (Pavlodar Agricultural Experimental Station)	0.73	6.82 ± 0.54	0.18 ± 0.06
Karaganda (Karaganda Agricultural Experimental Station named after A.F. Khristenko)	0.68	5.97 ± 0.48	0.21 ± 0.06
Tyumen (Research Institute of Agriculture of the Northern Trans-Urals and State Agrarian University of the Northern Trans-Urals)	0.78	8.34 ± 0.70	0.20 ± 0.06
Chelyabinsk (Chelyabinsk Research Institute of Agriculture)	0.77	9.40 ± 0.76	0.22 ± 0.07
Saratov (Federal Center of Agriculture Research of the South-East Region)	0.46	4.61 ± 0.47	0.44 ± 0.06
Omsk (Siberian Research Institute of Agriculture and Omsk State Agrarian University)	0.74	6.83 ± 0.49	0.20 ± 0.06

(92.8 % of occurrence). Accordingly, the *h* indicator will be the maximum – 0.56 ± 0.06. The same situation is observed in the analysis of allele frequencies at *Gli-D1* locus. With a high percentage of occurrence of *Gli-D1a* allele (89.9 %), the value of parameters μ (2.46 ± 0.30) and *H* (0.19) decreases and the value of *h* increases accordingly. On average, the samples created in the Kostanay (10.15 ± 0.62) and Chelyabinsk (9.40 ± 0.76) regions had the highest intrapopulation diversity of alleles (Table 3). It should be noted that the intrapopulation diversity and the proportion of rare alleles in the wheat samples from Chelyabinsk region have increased markedly compared to the results that were published earlier: μ = 6.15 ± 0.33 and *h* = 0.12 ± 0.05 (Chernakov, Metakovsky, 1994). It should be noted that *H* values of the wheat bred in Tyumen are higher (0.78) than in Chelyabinsk wheat (0.77), but at the same time, the index of intrapopulation diversity (μ) in the wheat in Chelyabinsk region is higher.

If we take into account μ errors of both areas, then the difference in their values lies within the statistical error, and the intrapopulation diversity is approximately at the same level. However, the mean values were derived from calculations of the diversity of each locus, in which case the allelic diversity of cultivars (populations) within the locus must be taken into account. It turned out that with the same number (seven) of identified alleles of *D1* locus, *Gli-D1a* allele prevailed in the wheat of Chelyabinsk Research Institute of Agriculture with a frequency of 71 %, and the rest had frequencies of no more than 10 %. At the same time, *Gli-D1a* allele was also "leading" in Tyumen cultivars, but with a lower frequency of 51.5 %, and *Gli-D1b* and *Gli-D1f* alleles with frequencies of 18.2 and 12.1 %, respectively, were found together with it. In other words, the diversity of Tyumen wheat cultivars at *D1* locus is higher than that of Chelyabinsk wheat, which ultimately affected the average values of genetic and intrapopulation diversity.

Comparative analysis of genetic diversity of gliadin coding loci of common wheat in breeding centers of Kazakhstan and Russia

To determine the similarities and differences between the wheat samples from various breeding centers (regions) of Russia and Kazakhstan for gliadin alleles, a cluster analysis was carried out, as a result of which three groups A, B, and
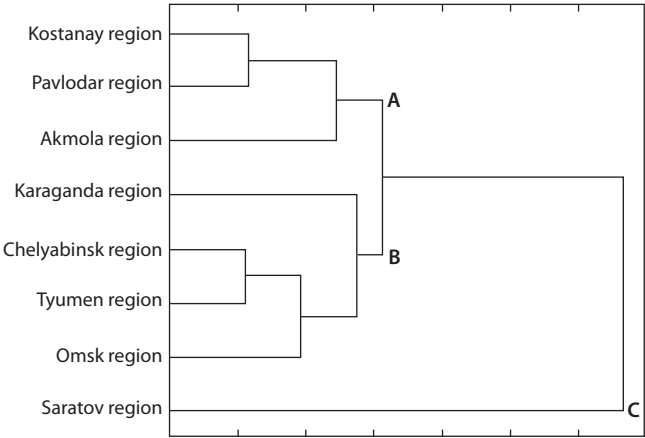
Fig. 4. Clustering by frequency of occurrence (%) of alleles of *Gli-1* and *Gli-2* loci of spring bread wheat depending on the region of origin.

C were formed (Fig. 4). Group A consisted of the samples from North Kazakhstan regions, while the wheat of Kostanay and Pavlodar selection turned out to be quite close. This is explained by the fact that there were "common" alleles with different frequencies and for each locus, for example, out of 14 alleles identified by *A1* locus, nine alleles turned out to be common, for a total of six loci out of 77 alleles, 45 were common, i.e. 58.4 %. Whereas in the wheat from Akmola region only 35.6 % of alleles were common to Kostanay and Pavlodar regions, which was reflected in the dendrogram. A similar situation was observed with the samples of cluster B. The "isolation" of Saratov wheat samples is due to the fact that only 10.1 % of gliadin alleles were common to the wheat from other regions of Kazakhstan and Russia.

For further establishment of the significant degree of differences between the groups of common wheat in terms of the frequency of occurrence of gliadin-coding loci alleles, the identity criterion (*I*) was used. In its essence, if the obtained value exceeds the table value of χ^2 at a given level of significance, then there is a significant difference between the groups (Zhivotovsky, 1979).

Supplementary Material 4 shows the values of genetic similarity (*r*), the criterion of pairwise similarity of the studied

М.У. Утебаев, С.М. Дашкевич
О.О. Крадецкая, И.В. Чилимова, Н.А. Боме

Генетическое разнообразие аллелей глиадина у образцов
пшеницы из селекционных центров Казахстана и России

2024
28•3

Table 4. Average values of genetic similarity (r) and identity criterion (I)
of spring bread wheat samples for *Gli-1* and *Gli-2* loci by region of origin

Regions	Akmola	Kostanay	Pavlodar	Karaganda	Tyumen	Chelyabinsk	Saratov	Omsk
Akmola	0	0.62 ± 0.05 287.1 (106.4)	0.72 ± 0.06 248.2 (92.8)	0.67 ± 0.05 355.7 (95.1)	0.65 ± 0.06 328.5 (104.1)	0.72 ± 0.06 243.6 (104.1)	0.69 ± 0.06 414.4 (103.0)	0.71 ± 0.06 325.1 (99.6)
Kostanay		0	0.85 ± 0.04 135.5 (97.4)	0.75 ± 0.04 271.2 (104.1)	0.86 ± 0.04 135.7 (104.1)	0.86 ± 0.05 128.0 (107.5)	0.55 ± 0.05 619.0 (116.5)	0.74 ± 0.05 296.4 (110.9)
Pavlodar			0	0.69 ± 0.06 260.1 (84.8)	0.80 ± 0.06 152.2 (92.8)	0.79 ± 0.06 147.6 (96.2)	0.55 ± 0.06 455.0 (96.2)	0.72 ± 0.05 244.9 (88.3)
Karaganda				0	0.79 ± 0.05 191.2 (88.3)	0.77 ± 0.06 195.1 (98.5)	0.55 ± 0.05 568.0 (95.1)	0.71 ± 0.05 304.3 (87.1)
Tyumen					0	0.85 ± 0.05 114.4 (97.4)	0.53 ± 0.05 507.2 (104.1)	0.71 ± 0.05 265.0 (98.5)
Chelyabinsk						0	0.57 ± 0.06 430.0 (110.9)	0.78 ± 0.05 187.2 (103.0)
Saratov							0	0.69 ± 0.05 407.5 (90.5)
Omsk								0

N o t e. The upper number is the indicator of genetic similarity (r), the lower number is the criterion of identity (I). In parentheses χ^2 for a 5 % significance level, if $I > \chi^2$, then the differences are significant.

groups and the criterion of identity (I) for each locus separately. Genetic similarity (r) does not exceed 1, but can be equal to 1 only if the groups being compared are identical in number and frequency of alleles. When averaged, the obtained values of the identity criterion (I) exceeded the tabular value of χ^2 for all pairwise comparisons. Accordingly, the studied groups of common wheat samples from different regions and breeding centers of Kazakhstan and Russia significantly differ from each other in gliadin-coding loci (Table 4).

However, when analyzing the I values for individual loci, it turned out that even in the presence of alleles characteristic of a certain area, a significant difference between the groups was not always achieved (see Supplementary Material 4), i. e. when comparing the wheat of Tyumen and Omsk origin, the difference at *D1* locus was insignificant I = 7.6 (12.6), since five out of seven identified alleles occurred in both groups, and with a fairly high frequency. Note that in most cases, there was a slight difference at *Gli-1* loci, while at *Gli-2* loci, the differences were statistically significant.

This is probably due to the fact that wheat breeding has traditionally been aimed at increasing yield, grain quality and resistance to various stressors, and alleles of *Gli-A1*, *Gli-B1* and *Gli-D1* loci are associated with baking quality (Nieto-Taladriz et al., 1994; Li et al., 2009; Demichelis et al., 2019), and resistance to leaf, stem rust (Czarnecki, Lukow, 1992; Cox et al., 1994) and powdery mildew (Hsam et al., 2015).

Conclusion
Based on the study, description and statistical calculation of the genetic diversity of allelic variants of gliadin-coding loci of common wheat, a significant difference in genotypes from different regions of Kazakhstan and Russia has been established. The revealed genetic differentiation on the basis

of protein polymorphism is likely adaptive. The gliadin alleles that are characteristic of a certain region have been identified. The "gliadin profile" of the wheat of Kazakhstan and Russian origin has been established, which shows the preference of wheat genotypes for gliadin alleles as a result of selection. This information can be used for the selection of parent pairs in the breeding process, the control of cultivars during reproduction, as well as for the establishment of varietal purity.

References
Autran J.C., Bushuk W., Wrigley C.W., Zillman R.R. Wheat cultivar identification by gliadin electrophoregrams. IV. Comparison of international methods. *Cereal Foods World*. 1979;24(9):471-475

Barley. UPOV Code(s): HORDE_VUL, *Hordeum vulgare* L. Guidelines for the conduct of tests for distinctness, uniformity and stability. Geneva: International Union for the Protection of New Varieties of Plants, 2018. Available at: https://www.upov.int/edocs/tgdocs/en/tg019.pdf

Branlard G., Dardevet M., Amiour N., Igrejas G. Allelic diversity of HMW and LMW glutenin subunits and omega-gliadins in French bread wheat (*Triticum aestivum* L.). *Genet. Resour. Crop Evol.* 2003; 50:669-679. DOI 10.1023/A:1025077005401

Chebotar S.V., Blagodarova E.M., Kurakina E.A., Semenyuk I.V., Polishchuk A.M., Kozub N.A., Sozinov I.A., Khokhlov A.N., Ribalka A.I., Sivolap Yu.M. Genetic polymorphism of loci determining bread making quality in Ukrainian wheat varieties. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2012;16(1):87-98 (in Russian)

Chernakov V.M., Metakovsky E.V. Diversity of gliadin-coding locus allelic variants and evaluation of genetic similarity of common wheat varieties from different breeding centers. *Genetika = Genetics (Moscow)*. 1994;30(4):509-517 (in Russian)

Cox T.S., Raupp W.J., Gill B.S. Leaf rust-resistance genes *Lr41*, *Lr42*, and *Lr43* transferred from *Triticum tauschii* to common wheat. *Crop Sci.* 1994;34(2):339-343. DOI 10.2135/cropsci1994.0011183 X003400020005x

Czarnecki E.M., Lukow O.M. Linkage of stem rust resistance gene *Sr33* and the gliadin (*Gli-D1*) locus on chromosome 1DS. *Genome*. 1992;35(4):565-568. DOI 10.1139/g92-084

Demichelis M., Vanzetti L.S., Crescente J.M., Nisi M.M., Pflüger L., Bainotti C.T., Helguera M. Significant effects in bread-making quality associated with the gene cluster *Glu-D3/Gli-D1* from the bread wheat cultivar Prointa Guazú. *Cereal Res. Commun*. 2019; 47(1):111-122. DOI 10.1556/0806.46.2018.055

Dobrotvorskaya T.V., Dragovich A.Yu., Martynov S.P., Pukhal'skii V.A. Genealogical and statistical analyses of the inheritance of gliadin-coding alleles in a model set of common wheat *Triticum aestivum* L. cultivars. *Russ. J. Genet*. 2009;45(6):685-695. DOI 10.1134/S1022795409060088

GRIS – Genetic Resources Information System for Wheat and Triticale. 2017. Avialiable at: http://www.wheatpedigree.net/

Hsam N.O., Kowalczyk K., Zeller F.J., Hsam S.L. Characterization of powdery mildew resistance and linkage studies involving the *Pm3* locus on chromosome 1A of common wheat (*Triticum aestivum* L.). *J. Appl. Genet*. 2015;56(1):37-44. DOI 10.1007/s13353-014-0236-7

Kakaei M., Ahmadian S. Genetic diversity study of some Iranian alfalfa genotypes based on seed storage proteins patterns. *Iran. J. Sci. Technol. Trans. A Sci*. 2021;45(4):1223-1228. DOI 10.1007/s40995-021-01142-z

Kaur R., Kaur R., Sharma N., Kumari N., Khanna R., Singh G. Protein profiling in a set of wild rice species and rice cultivars: a stepping stone to protein quality improvement. *Cereal Res. Commun*. 2023; 51:163-177. DOI 10.1007/s42976-022-00273-2

Khrunov A.A., Fisenko A.V., Beletsky S.L., Dragovich A.Yu. Study of the relationship between the composition of gliadins and economically valuable traits of common wheat. *Izvestiya Timiryazevskoj Sel'skohozyajstvennoj Akademii = Izvestiya of Timiryazev Agricultural Academy*. 2011;2:11-19 (in Russian)

Kozub N.A., Sozinov I.A., Sobko T.A., Kolyuchii V.T., Kuptsov S.V., Sozinov A.A. Variation at storage protein loci in winter common wheat cultivars of the Central Forest-Steppe of Ukraine. *Cytol. Genet*. 2009;43(1):55-62. DOI 10.3103/S0095452709010101

Kozub N.A., Sozinov I.A., Sobko T.A., Dedkova O.S., Badaeva E.D., Netsvetaev V.P. Rye translocations in the varieties of winter common wheat. *Sel'skokhozyaistvennaya Biologiya = Agricultural Biology*. 2012;47(3):68-74. DOI 10.15389/agrobiology.2012.3.68eng

Kunanbayev K., Churkina G., Filonov V., Utebayev M., Rukavitsina I. Influence of cultivation technology on the productivity of spring wheat and the humus state of Southern carbonate soils of Northern Kazakhstan. *J. Ecol. Eng*. 2022;23(3):49-58. DOI 10.12911/22998993/145459

Laboratory Analysis of Wheat Seed Proteins. Technological instruction. Moscow, 2013 (in Russian)

Li Y., Song Y., Zhou R., Branlard G., Jia J. Detection of QTLs for breadmaking quality in wheat using a recombinant inbred line population. *Plant Breed*. 2009;128(3):235-243. DOI 10.1111/j.1439-0523.2008.01578.x

Lyubimova A.V., Tobolova G.V., Eremin D.I., Loskutov I.G. Dynamics of genetic diversity of oat varieties in the Tyumen region at avenin-coding loci. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(2):123-130. DOI 10.18699/VJ20.607

Ma G., Li Q., Li S., Liu Z., Cui Y., Zhang J., Liu D. Genetic diversity and classification of chinese elite foxtail millet [*Setaria italica* (L.) P. Beauv.] revealed by acid-PAGE prolamin. *Agric. Sci*. 2022;13(3): 404-428. DOI 10.4236/as.2022.133028

Masci S., Rovelli L., Kasarda D.D., Vensel W.H., Lafiandra D. Characterisation and chromosomal localisation of C-type low-molecular-weight glutenin subunits in the bread wheat cultivar Chinese Spring. *Theor. Appl. Genet*. 2002;104(2-3):422-428. DOI 10.1007/s001220100761

McIntosh R.A., Devos K.M., Dubkovsky J., Morris C.F., Rogers W.J. Catalogue of Gene Symbols for Wheat. Supplement. 2003. Available at: https://wheat.pw.usda.gov/ggpages/wgc/2003upd.html

Melnikova N.V., Kudryavtseva A.V., Kudryavtsev A.M. Catalogue of alleles of gliadin-coding loci in durum wheat (*Triticum durum* Desf.). *Biochimie*. 2012;94(2):551-557. DOI 10.1016/j.biochi.2011.09.004

Metakovsky E.V. Gliadin allele identification in common wheat. II. Cataloque of gliadin alleles in common wheat. *J. Genet. Breed*. 1991; 45(4):325-344

Metakovsky E.V., Branlard G. Genetic diversity of French common wheat germplasm based on gliadin alleles. *Theor. Appl. Genet*. 1998; 96:209-218. DOI 10.1007/s001220050729

Metakovsky E.V., Novoselskaya A.Yu. Gliadin allele identification in common wheat. I. Methodological aspects of the analysis of gliadin pattern by one-dimensional polyacrylamide gel electrophoresis. *J. Genet Breed*. 1991;45(4):317-324

Metakovsky E.V., Wrigley C.W., Bekes F., Gupta R.B. Gluten polypeptides as useful genetic markers of dough quality in Australian wheats. *Aust. J. Agric. Res*. 1990;41(2):289-306. DOI 10.1071/AR9900289

Metakovsky E.V., Pogva N.E., Blancardi A.M., Redaelli R. Gliadin allele composition of common wheat cultivars grown in Italy. *J. Gen. Breed*. 1994;48(1):55-66

Metakovsky E.V., Annicchiarico P., Boggini G.E., Pogna N.E. Relationship between gliadin alleles and dough strength in Italian bread wheat cultivars. *J. Cereal Sci*. 1997;25(3):229-236. DOI 10.1006/jcrs.1996.0088

Metakovsky E.V., Gómez M., Vázquez J.F., Carrillo J.M. High genetic diversity of Spanish common wheats as judged from gliadin alleles. *Plant Breed*. 2000;119(1):37-42. DOI 10.1046/j.1439-0523.2000.00450.x

Metakovsky E., Melnik V., Rodriguez-Quijano M., Upelniek V., Carrillo J.M. A catalog of gliadin alleles: polymorphism of 20th-century common wheat germplasm. *Crop J*. 2018;6(6):628-641. DOI 10.1016/j.cj.2018.02.003

Metakovsky E., Melnik V.A., Pascual L., Wrigley C.W. Gliadin genotypes worldwide for spring wheats (*Triticum aestivum* L.) 1. Genetic diversity and grain-quality gliadin alleles during the 20th century. *J. Cereal Sci*. 2019;87:172-177. DOI 10.1016/j.jcs.2019.03.008

Metakovsky E., Melnik V., Pascual L., Wrigley C.W. Over 40 % of 450 registered wheat cultivars (*Triticum aestivum*) worldwide are composed of multiple biotypes. *J. Cereal Sci*. 2020;96:103088. DOI 10.1016/j.jcs.2020.103088

MN-01-03/001:2000 – Blé – Identification des varieties par électrophorèse. In: Les projets de normes, adoptés par le Conseil National de Normalisation et de Contrôle de Qualité lors de la session du 20 décembre 2000, sont homologués comme normes maliennes. Bamako, 2001

Nei M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA*. 1973;70(12):3321-3323. DOI 10.1073/pnas.70.12.3321

Nieto-Taladriz M.T., Perretant M.R., Rousset M. Effect of gliadins and HMW and LMW subunits of glutenin on dough properties in the F_6 recombinant inbred lines from a bread wheat cross. *Theor. Appl. Genet*. 1994;88(1):81-88. DOI 10.1007/BF00222398

Nikolaev A.A., Pukhal'sky V.A., Upelniek V.P. Genetic diversity of local spring bread wheats (*Triticum aestivum* L.) of West and East Siberia in gliadin genes. *Russ. J. Genet*. 2009;45(2):189-197. DOI 10.1134/S1022795409020094

Noma S., Hayakawa K., Abe C., Suzuki S., Kawaura K. Contribution of α-gliadin alleles to the extensibility of flour dough in Japanese wheat cultivars. *J. Cereal Sci*. 2019;86:15-21. DOI 10.1016/j.jcs.2018.12.017

Novoselskaya-Dragovich A.Y., Krupnov V.A., Saifulin R.A., Pukhalskiy V.A. Dynamics of genetic variation at gliadin-coding loci in Saratov cultivars of common wheat *Triticum aestivum* L. over eight decades of scientific breeding. *Russ. J. Genet*. 2003;39(10):1130-1137. DOI 10.1023/A:1026170709964

Novoselskaya-Dragovich A.Y., Fisenko A.V., Yankovsky N.K., Kudryavtsev A.M., Yang Q., Lu Z., Wang D. Genetic diversity of storage protein genes in common wheat (*Triticum aestivum* L.) cultivars

М.У. Утебаев, С.М. Дашкевич
О.О. Крадецкая, И.В. Чилимова, Н.А. Боме

Генетическое разнообразие аллелей глиадина у образцов
пшеницы из селекционных центров Казахстана и России

2024
28·3

from China and its comparison with genetic diversity of cultivars from other countries. *Genet. Resour. Crop Evol*. 2011;58(4):533-543. DOI 10.1007/s10722-010-9596-y

Novoselskaya-Dragovich A.Y., Fisenko A.V., Puhal'skii V.A. Genetic differentiation of common wheat cultivars using multiple alleles of gliadin coding loci. *Russ. J. Genet*. 2013;49(5):487-496. DOI 10.1134/S1022795413020087

Novoselskaya-Dragovich A.Yu., Bespalova L.A., Shishkina A.A., Melnik V.A., Upelniek V.P., Fisenko A.V., Dedova L.V., Kudryavtsev A.M. Genetic diversity of common wheat varieties at the gliadin-coding loci. *Russ. J. Genet.* 2015;51(3):323-333. DOI 10.1134/S1022795415030102

Salavati A., Sameri H., Boushehri A.A.S., Yazdi-Samadi B. Evaluation of genetic diversity in Iranian landrace wheat *Triticum aestivum* L. by using gliadin alleles. *Asian J. Plant Sci*. 2008;7(5):440-446. DOI 10.3923/ajps.2008.440.446

Sharma A., Sheikh I., Kumar R., Kumar K., Vyas P., Dhaliwal H.S. Evaluation of end use quality and root traits in wheat cultivars associated with 1RS.1BL translocation. *Euphytica*. 2018;214(4):62. DOI 10.1007/s10681-018-2144-0

Shavrukov Y., Suchecki R., Eliby S., Abugalieva A., Kenebayev S., Langridge P. Application of next-generation sequencing technology to study genetic diversity and identify unique SNP markers in bread wheat from Kazakhstan. *BMC Plant Biol*. 2014;14:258. DOI 10.1186/s12870-014-0258-7

Sozinov A.A. Protein Polymorphism and its Significance in Genetics and Breeding. Moscow, 1985 (in Russian)

Sozinov A.A., Poperelya F.A. Genetic classification of prolamins and its use for plant breeding. *Ann. Technol. Agric*. 1980;29(2):229-245

ST RK 3323-2018. Seeds of Wheat. Identification of varieties by electrophoresis. Astana, 2018 (in Russian)

Upelniek V.P., Brezhneva T.A., Dadashev S.Y., Novozhilova O.A., Molkanova O.I., Semikhov V.F. On the use of alleles of gliadin-coding loci as possible adaptability markers in the spring wheat (*Triticum aestivum* L.) cultivars during seed germination. *Russ. J. Genet*. 2003;39(12):1680-1686. DOI 10.1023/B:RUGE.0000009158.41760.67

Utebayev M., Dashkevich S., Babkenov A., Shtefan G., Fahrudenova I., Bayahmetova S., Sharipova B., Kaskarbayev Zh., Shavru-kov Y. Application of gliadin polymorphism for pedigree analysis in common wheat (*Triticum aestivum* L.) from Northern Kazakhstan. *Acta Physiol. Plant*. 2016;38:204. DOI 10.1007/s11738-016-2209-4

Utebayev M., Dashkevich S., Bome N., Bulatova K., Shavrukov Y. Genetic diversity of gliadin-coding alleles in bread wheat (*Triticum aestivum* L.) from Northern Kazakhstan. *PeerJ*. 2019a;7:e7082. DOI 10.7717/peerj.7082

Utebayev M., Dashkevich S., Kunanbayev K., Bome N., Sharipova B., Shavrukov Y. Genetic polymorphism of glutenin subunits with high molecular weight and their role in grain and dough qualities of spring bread wheat (*Triticum aestivum* L.) from Northern Kazakhstan. *Acta Physiol. Plant*. 2019b;41(5):71. DOI 10.1007/s11738-019-2862-5

Utebayev M.U., Bome N.A., Zemtsova E.S., Kradetskaya O.O., Chilimova I.V. Diversity of high-molecular-weight glutenin subunits and evaluation of genetic similarities in spring bread wheats from different breeding centers. *Trudy po Prikladnoy Botanike, Genetike i Selektsii = Proceedings on Applied Botany, Genetics and Breeding*. 2021;182(1):99-109. DOI 10.30901/2227-8834-2021-1-99-109 (in Russian)

Utebayev M.U., Dolinny Y.Y., Dashkevich S.M., Bome N.A. Allelic composition of gliadin-coding loci as a 'portrait' in spring soft wheat selections of Russian and Kazakh origins. *SABRAO J. Breed. Genet*. 2022;54(4):755-766. DOI 10.54910/sabrao2022.54.4.7

Watry H., Zerkle A., Laudencia-Chingcuanco D. Modified acid-PAGE method for rapid screening and phenotyping of wheat gliadin mutant lines. *MethodsX*. 2020;7:100858. DOI 10.1016/j.mex.2020.100858

Wheat. UPOV Code(s): TRITI_AES, *Triticum aestivum* L. emend. Fiori et Paol. Guidelines for the conduct of tests for distinctness, uniformity and stability. Geneva: International Union for the Protection of New Varieties of Plants, 2022. Available at: https://www.upov.int/edocs/tgdocs/en/tg003.pdf

Xynias I.N., Kozub N.O., Sozinov I.A. Seed storage protein composition of Hellenic bread wheat cultivars. *Plant Breed*. 2006;125(4):408-410. DOI 10.1111/j.1439-0523.2006.01242.x

Zhivotovsky L.A. Population similarity measure for polymorphic characters. *Zhurnal Obshchey Biologii = Journal of General Biology*. 1979;40(4):587-602 (in Russian)

Zhivotovsky L.A. Population Biometry. Moscow, 1991 (in Russian)

Search for biocontrol agents among endophytic lipopeptide-synthesizing bacteria *Bacillus* spp. to protect wheat plants against Greenbug aphid (*Schizaphis graminum*)

S.D. Rumyantsev (iD), V.Y. Alekseev (iD), A.V. Sorokan (iD), G.F. Burkhanova (iD), E.A. Cherepanova (iD),
I.V. Maksimov (iD), S.V. Veselova (iD) (✉)

Institute of Biochemistry and Genetics of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
(✉) veselova75@rambler.ru

Abstract. Beneficial endophytic bacteria can suppress the development of insect pests through direct antagonism, with the help of metabolites, or indirectly by the induction of systemic resistance through the regulation of hormonal signaling pathways. Lipopeptides are bacterial metabolites that exhibit direct antagonistic activity against many organisms, including insects. Also, lipopeptides are able to trigger induced systemic resistance (ISR) in plants against harmful organisms, but the physiological mechanisms of their action are just beginning to be studied. In this work, we studied ten strains of bacteria isolated from the tissues of wheat and potatoes. Sequencing of the 16S rRNA gene showed that all isolates belong to the genus *Bacillus* and to two species, *B. subtilis* and *B. velezensis*. The genes for lipopeptide synthetase – surfactin synthetase (*Bs_srf*), iturin synthetase (*Bs_ituA*, *Bs_ituB*) and fengycin synthetase (*Bs_fenD*) – were identified in all bacterial isolates using PCR. All strains had high aphicidal activity against the Greenbug aphid (*Schizaphis graminum* Rond.) due to the synthesis of lipopeptides, which was proven using lipopeptide-rich fractions (LRFs) isolated from the strains. Endophytic lipopeptide-synthesizing strains of *Bacillus* spp. indirectly affected the viability of aphids, the endurance of plants against aphids and triggered ISR in plants, which manifested itself in the regulation of oxidative metabolism and the accumulation of transcripts of the *Pr1*, *Pr2*, *Pr3*, *Pr6* and *Pr9* genes due to the synthesis of lipopeptides, which was proven using LRF isolated from three strains: *B. subtilis* 26D, *B. subtilis* 11VM, and *B. thuringiensis* B-6066. We have for the first time demonstrated the aphicidal effect of fengycin and the ability of the fengycin-synthesizing strains and isolates, *B. subtilis* Ttl2, *Bacillus* sp. Stl7 and *B. thuringiensis* B-6066, to regulate components of the pro-/antioxidant system of aphid-infested plants. In addition, this work is the first to demonstrate an elicitor role of fengycin in triggering a systemic resistance to *S. graminum* in wheat plants. We have discovered new promising strains and isolates of endophytes of the genus *Bacillus*, which may be included in the composition of new biocontrol agents against aphids. One of the criteria for searching for new bacteria active against phloem-feeding insects can be the presence of lipopeptide synthetase genes in the bacterial genome.

Key words: *Bacillus* spp.; *Schizaphis graminum*; endophytes; PCR; RT-PCR; plant-microbial interactions; lipopeptides; biological control agents.

С.Д. Румянцев, В.Ю. Алексеев, А.В. Сорокан, Г.Ф. Бурханова
Е.А. Черепанова, И.В. Максимов, С.В. Веселова

Эндофитные липопептид-синтезирующие *Bacillus* ssp.
для защиты пшеницы от *Schizaphis graminum*

2024
28•3

Аннотация. Полезные эндофитные бактерии могут подавлять развитие вредителей за счет прямого антагонизма, с помощью метаболитов или опосредованно индуцировать системную устойчивость через регуляцию гормональных сигнальных путей. Липопептиды – бактериальные метаболиты, проявляющие прямую антагонистическую активность ко многим организмам, в том числе к насекомым. Также липопептиды способны запускать системную индуцированную устойчивость у растений против вредных организмов. В настоящее время механизм действия бактериальных метаболитов липопептидов на защитную систему растений только начинают исследовать. В данной работе изучено десять штаммов и изолятов бактерий, выделенных из внутренних тканей культурной и дикой пшеницы и картофеля. Секвенирование гена 16S рРНК показало принадлежность всех изолятов к роду *Bacillus* и двум видам – *B. subtilis* и *B. velezensis*. У всех бактериальных изолятов методом ПЦР были идентифицированы гены липопептид синтаз – сурфактин синтазы (*Bs_srf*), итурин синтаз (*Bs_ituA*, *Bs_ituB*) и фенгицин синтазы (*Bs_fenD*). Все штаммы обладали афицидной активностью в отношении обыкновенной злаковой тли (*Schizaphis graminum* Rond.) за счет синтеза липопептидов, что было доказано с помощью липопептид-богатых фракций (ЛБФ), выделенных из штаммов. Эндофитные липопептид-синтезирующие штаммы *Bacillus* spp. опосредованно влияли на жизнеспособность тли, выносливость растений по отношению к тле и запускали системную индуцированную устойчивость у растений, что проявлялось в регуляции окислительного метаболизма и накоплении транскриптов генов *Pr1*, *Pr2*, *Pr3*, *Pr6* и *Pr9*, за счет синтеза липопептидов, что подтверждено с помощью ЛБФ, выделенных из трех штаммов – *B. subtilis* 26D, *B. subtilis* 11VM и *B. thuringiensis* B-6066. В нашей работе впервые показано афицидное действие фенгицина и способность штаммов и изолятов *B. subtilis* Ttl2, *Bacillus* sp. Stl7 и *B. thuringiensis* B-6066, синтезирующих фенгицин, регулировать компоненты про-/антиоксидантной системы растений, зараженных тлей. Кроме того, впервые продемонстрирована элиситорная роль фенгицина в запуске системной устойчивости растений пшеницы к *S. graminum*. Обнаружены новые перспективные штаммы и изоляты эндофитных бактерий рода *Bacillus*, которые могут стать основой будущих биопрепаратов против тлей. Одним из критериев поиска новых бактерий, активных против насекомых, питающихся флоэмным соком, может быть наличие в бактериальном геноме генов липопептид синтаз.

Ключевые слова: *Bacillus* spp.; *Schizaphis graminum*; эндофитные бактерии; ПЦР; ПЦР в реальном времени; растительно-микробные взаимодействия; липопептиды; биопрепараты.

Introduction

Insects of the order Hemiptera, aphids, whiteflies, planthoppers, including the Greenbug aphid *Schizaphis graminum*, which are sap-sucking insects, can cause severe yield losses of up to 60–80 % due to their influence on photosynthesis processes and biomass growth rate (Koch et al., 2016; Radchenko et al., 2022). Currently, chemical insecticides remain the main agents of controlling phloem-feeding pests, leading to the emergence of new pesticide-resistant forms of pests. Therefore, it is necessary to find environmentally friendly biological control agents to defend plants from pests. Such effective biological control agents can be endophytic growth-promoting bacteria that can live inside plants without causing diseases in them (Rani et al., 2022).

Currently, many researchers suppose that endophytes protect plants from stress through the mechanisms of direct or indirect protective effects on harmful organisms due to the synthesis and secretion of diverse metabolites (Oukala et al., 2021; Xia et al., 2022). The direct action of endophytes is carried out due to the biocidal activity of some metabolites (bacteriocins, biosurfactants, lipopeptides). Indirect action is expressed in the ability of endophytes to stimulate growth processes in plants, improve the immune system of plants, and build a durable defense of the host against harmful organisms, which is known as priming (Rashid, Chung, 2017; Xia et al., 2022). Bacteria-induced priming provides faster and longer-lasting plant protection throughout the growing season with low physiological costs, making endophyte-based biocontrol agents very promising (Oukala et al., 2021; Rani et al., 2022; Xia et al., 2022). Activation of the plant immune system and priming by endophytes is realized by triggering induced systemic resistance (ISR) against harmful organisms, which has been shown by many researchers and summarized in recent reviews (Oukala et al., 2021; Rani et al., 2022; Xia et al., 2022). Endophyte-activated ISR is regulated by bacterial-produced hormone-like substances with growth-regulating activity such as abscisic (ABA), salicylic (SA), jasmonic acids (JA), and ethylene (ET) (Pieterse et al., 2014; Rashid, Chung, 2017). The characteristic features of ISR are jumps in the generation of reactive oxygen species (ROS) and changes in the gene expression with a focus on defense-related genes of pathogenesis-related proteins (PR proteins) (Oukala et al., 2021; Xia et al., 2022).

Bacteria of *Bacillus* spp. are famous for their ability to synthesize a wide range of diverse metabolites (Miljaković et al., 2020). Bacterial metabolites are the active ingredient of any biocontrol agent. Lipopeptides are one of the major classes of bacterial metabolites intensively researched in recent years. Lipopeptides are small peptides that have biocidal properties against mycoplasmas, bacteria, yeasts, fungi, oomycetes, nematodes, and pests due to their capability to connect to the lipid bilayer of the plasmalemma and change its permeability (Andrić et al., 2021). Bacteria of the genus *Bacillus* produce lipopeptides of three families: surfactins, fengycins and iturins (Andrić et al., 2021). Recently, the insecticidal activity of lipopeptides against the orders Diptera, Coleoptera, Hemiptera, and Lepidoptera have been shown in some studies (Rodríguez et al., 2018; Denoirjean et al., 2022). Currently, the eliciting role of lipopeptides in triggering systemic resistance in plants is being actively studied (Rashid et al., 2018; Tunsagool et al., 2019; Miljaković et al., 2020). The elicitor role of lipopeptides against a wide range of pathogens of plants has been shown in many studies (Tunsagool et al., 2019; Jiang et al., 2021). However, information on the elicitor role of lipopeptides in triggering ISR in plants against sucking insects is limited (Rashid et al., 2018; Rumyantsev et al., 2023).

Thus, the search for highly effective endophytic strains for plant protection against sap-sucking insects using the priming mechanism, the study of the metabolic composition and mechanisms of action of endophytes is an urgent task. In this regard, the aim of our work was to study the elicitor role of lipopeptides and the ability of endophytic bacteria that synthesize lipopeptides to protect plants through the priming mechanism. To do this, in our work we searched for strains and isolates of the genus *Bacillus* capable of synthesizing lipopeptides, studied the insecticidal activity of bacteria in relation to Greenbug aphid, and also studied the indirect effect of endophytes and lipopeptide-rich fractions (LRFs) of three strains – *B. subtilis* 26D, *B. subtilis* 11VM and *B. thuringiensis* B-6066 – on the redox status, indicators of resistance (antibiosis and endurance) to the pest, and changes in the expression of defense-related genes of PR proteins of wheat plants populated by *S. graminum*.

Materials and methods

Bacteria, plants and insects. In this work, gram-positive aerobic endophytic bacteria from the collection of the Laboratory of Biochemistry of Plant Immunity of the Institute of Biochemistry and Genetics of the Ufa Federal Research Centre of the Russian Academy of Sciences (UFRC RAS) were used. Three strains of *Bacillus subtilis*, 26D (Russian Collection of Agricultural Microorganisms (RCAM), No. 128), 11VM (RCAM No. 519), Ttl2 (isolated from the leaves of *Triticum timopheevii* Zhuk., Republic of Bashkortostan), one strain of *B. thuringiensis*, B-6066 (All-Russian collection of industrial microorganisms (ARCIM), No. 6066), and six isolates of *Bacillus* spp. isolated from leaves of wheat and potatoes growing on the territory of the Republic of Bashkortostan were used. Bacteria were grown on liquid lysogenic broth (LB) medium (1 % tryptone, 0.5 % yeast extract and 0.5 % NaCl) in 50 ml flasks at 28 °C using laboratory shakers (120 rpm) within 72 h until complete sporulation.

In this work, we studied the population of Greenbug aphid (*Schizaphis graminum* Rond.), 2020, which was maintained under laboratory-controlled conditions on plants of common spring wheat (*Triticum aestivum* L.) cv. Salavat Yulaev (SY) as described previously (Rumyantsev et al., 2023). Seeds of cv. SY were obtained from the Bashkir Research Institute of Agriculture – Subdivision of the UFRC RAS.

Isolation of DNA from bacteria. Genomic DNA from bacteria was isolated with a lysis buffer containing 1 % Che-

lex 100 resin (BioRad Laboratories, USA), as described earlier (Veselova et al., 2022).

16S rRNA gene sequencing. The gene of 16S rRNA was amplified using the universal primers 27F (5′-CAGAGTTT GATCCTGGCT-3′) and 1492R (5′-AGGAGGTGATCCAG CCGCA-3′). Amplified fragments of the 16S RNA gene of *Bacillus* spp. isolates were visualized on a 1 % agarose gel stained with ethidium bromide. Then, PCR fragments of the 16S RNA gene were excised from the agarose gel and purified using a diaGene agarose gel DNA elution kit (DiaM, Russia). Sanger sequencing of PCR fragments was performed on a 3500xL genetic analyzer from Applied Biosystems (Evrogen, Russia). BLAST software was used for alignment and comparison of the obtained sequences of *Bacillus* spp. isolates with sequences deposited in GenBank. These results were used for identifying what matched the searched sequence and what species the isolates under consideration belonged to. Data on sequences and species of bacteria were submitted in GenBank (see Table 3).

Detection of genes of lipopeptide synthetase in the *Bacillus* spp. strains and isolates by PCR. The genes of lipopeptide synthetase – surfactin synthetase (*srf*), iturin synthetases (*ituA*, *ituB*) and fengycin synthetase (*fenD*) – were identified in bacterial strains and isolates using PCR with gene-specific primers. Primers to the *bac* gene encoding 16S RNA of *Bacillus* spp. were used as an inner control. The sequences of all primers are presented in Table 1.

Isolation of the lipopeptide-rich fraction (LRF) from the *Bacillus* spp. strains. LRFs from the acidified liquid bacterial culture medium of three *Bacillus* spp. strains *B. subtilis* 11VM, *B. subtilis* 26D and *B. thuringiensis* B-6066 and two isolates *Bacillus* sp. Tas2.1 and *Bacillus* sp. Tas8.2 were obtained by ethanol extraction followed by purification on an Amicon Ultracel-3K filter (Merck KGaA, Darmstadt, Germany) as described previously (Maksimov et al., 2020). The purified lipopeptide fraction was weighed and dissolved in 80 % ethanol, the growth-promoting concentrations selected earlier were used (Maksimov et al., 2020).

Aphicidal activity of the *Bacillus* spp. strains and isolates. Aphicidal activity of bacterial strains and LRF was studied on first cut leaves of wheat seedlings cv. SY, placed in test tubes with 5 ml of the bacterial suspension at the concentration of 10^7 cells/ml (control tubes contained 5 ml of sterile water) or those with 5 ml of LRF at various concentrations from 2.5 to 200 μg/ml according to a method modified for

Table 1. Nucleotide sequences of primers, bacterial genes encoding lipopeptide synthases

Genes	GenBank Accession number	Sequence (5′–3′)	Amplicon size, bp
Bs_srf	EU882341	F – ATGAAGATTTACGGAATTTATATG R – TTATAAAAGCTCTTCGTACGAG	675
Bs_ituA	D21876.1	F – ATGAAAATTTACGGAGTATATATG R – TTATAACAGCTCTTCATACGTT	674
Bs_ituB	KR149331	F – AAGAAGGCGTTTTTCAAGCA R – CGACATACAGTTCTCCCGGT	508
Bs_fenD	AJ011849	F – TTTGGCAGCAGGAGAAGTTT R – GCTGTCCGTTCTGCTTTTTC	964
Bs_Bac	NR102783	F – ACCAGAAAGCCACGGCTAACTAC R – GGCGGAAACCCCCTAACACT	356

С.Д. Румянцев, В.Ю. Алексеев, А.В. Сорокан, Г.Ф. Бурханова
Е.А. Черепанова, И.В. Максимов, С.В. Веселова

Эндофитные липопептид-синтезирующие *Bacillus* ssp.
для защиты пшеницы от *Schizaphis graminum*

2024
28 • 3

wheat and described earlier (Veselova et al., 2019). Aphicidal activity was expressed as mortality rate (%) among the total number of aphids (Veselova et al., 2019).

Experimental conditions. Before planting, some wheat seeds were treated with a liquid culture of bacteria in a semi-dry manner at growth-stimulating concentrations selected earlier (Alekseev et al., 2021; Rumyantsev et al., 2023). The cell titer in the suspension was counted at 600 nm using a SmartSpectm Plus spectrophotometer (Bio-Rad, USA) certified for this task. The cell titer of the studied cultures was $(1.8–2)\cdot10^9$ cells/ml; by adding distilled water, the suspensions were diluted to a final titer of $(2–4)\cdot10^6$ cells/ml and the resulting suspensions were used for seed treatment. The final titer of *B. subtilis* 26D, *Bacillus* sp. Tas2.1 and *B. thuringiensis* B-6066 was $4\cdot10^6$ cells/ml. The final titer of *Bacillus* sp. Tas8.2, *B. subtilis* 11VM, *B. subtilis* Ttl2 and *Bacillus* sp. Stl7 was $2\cdot10^6$ cells/ml.

Wheat seedlings were grown in 1-liter vessels on an aquatic culture (10 % Hoagland–Arnon solution) under aphid breeding conditions. Solutions of LRFs at growth-stimulating concentrations selected earlier (Alekseev et al., 2021) were added to the plant nutrient medium 24 h before aphid colonization. After 24 h, the medium was replaced with Hoagland–Arnon solution without LRFs. Growth-promoting concentrations of LRF of *Bacillus* spp. strains *B. subtilis* 11VM, *B. subtilis* 26D and *B. thuringiensis* B-6066 and isolates *Bacillus* sp. Tas2.1 and *Bacillus* sp. Tas8.2 were 1.5, 2.5, 1.5, 2.5 and 2.0 µg/ml, respectively (Alekseev et al., 2021). Plant treatment with LRFs was carried out to establish the elicitor role of lipopeptides in the induction of defensive signaling pathways in plants and did not pursue the goal of studying these metabolites as independent biocontrol agents. The colonization of 4-day-old wheat seedlings by aphids was carried out as described earlier (Rumyantsev et al., 2023).

Antibiosis test and endurance test. The antibiosis test was carried out as described earlier (Veselova et al., 2019). Mortality and fecundity of aphids were expressed as % of the total number of aphids. The propagation coefficient was calculated as described earlier (Veselova et al., 2019). Plant endurance was assessed by measuring the length of the first and second leaves of seedlings as described previously and expressed as % leaf growth compared to uninfected control (Veselova et al., 2019).

The content of hydrogen peroxide (H_2O_2) and the activity of enzymes – peroxidase (POD) and catalase (CAT) were analyzed according to standard methods (Rumyantsev et al., 2023). To measure the content of hydrogen peroxide (H_2O_2) and enzyme activity, plant material was homogenized 24 and 72 hours after colonization by *S. graminum* in 0.05 M Na-phosphate buffer (PB), pH 6.2, in a ratio of 1:5 (wt/vol) and incubated at 4 °C for 30 min. The supernatant was separated by centrifugation at 15,000 g for 15 min (5415K Eppendorf, Germany). The concentration of H_2O_2 in the supernatant was determined according to the method of (Bindschedler et al., 2001; Maksimov et al., 2011), using orange xylenol in the presence of Fe^{2+} ions. After coloring, the mixture was centrifuged for 5 min at 10,000 g and the optical density was measured at a wavelength of 560 nm on an LS 55 Luminescence Spectrometer (Perkin Elmer, USA). H_2O_2 content was calculated using a calibration curve and expressed in µmol H_2O_2/g fresh

weight (FW). POD activity was determined by a micromethod in 96-well plates (Corning-Costar, USA) by the oxidation of (o-)phenylenediamine in the presence of H_2O_2 at 490 nm on a Benchmark Microplate Reader spectrophotometer (Bio-Rad Laboratories, USA) (Veselova et al., 2014). The enzyme activity was expressed in optical density/mg protein per minute, which corresponded to the amount of oxidized substrate causing an increase in optical density in 1 min. CAT activity was determined by a micromethod based on the ability of H_2O_2 to form a stable colored complex with molybdate salts (Veselova et al., 2014). Optical density was measured at 405 nm on a Benchmark Microplate Reader spectrophotometer. CAT activity was calculated using a calibration curve and expressed in µmol H_2O_2/(mg protein per min). Protein content was determined by the Bradford method.

Performing qPCR. Isolation of RNA from wheat leaves (five plants per repeat) fixed in liquid nitrogen 1, 3, and 6 days after aphid infestation was performed using Lira® (Biolabmix, Russia) according to the manufacturer's instructions. cDNA synthesis was performed as described previously (Veselova et al., 2022). Expression of genes encoding PR proteins was analyzed by quantitative real-time PCR using a CFX Connect real-time PCR Detection System device (BioRad Laboratories, USA) and a set of predefined reagents EvaGreen I (Sintol, Russia). In the work, primers for the genes encoding PR1 protein, PR2 protein – glucanase, PR3 protein – chitinase, PR6 protein – proteinase inhibitors and PR9 protein peroxidase were used. To standardize the data, the wheat gene *TaRli* (RNaseLinhibitor-like) was used as an inner reference for the real-time qPCR analysis. Primers for qRT-PCR were designed using a web-based primer designing tool from IDT (http://eu.idtdna.com/Scitools/Applications/Primerquest) (USA). Primer sequences were validated by the presence of only a single peak on the thermal dissociation (Tm) curve generated by the thermal denaturing protocol. The sequences of all primers are presented in Table 2.

To quantify relative gene expression, the delta-delta Ct method was applied using the CFX Connect real-time PCR Detection System (BioRad Laboratories, USA) as described

Table 2. Nucleotide sequences of primers for wheat genes encoding PR-proteins

Genes	GenBank Accession number	Sequence (5′–3′)
TaPr1	AF384143	F – ATAACCTCGGCGTCTTCATC R – GCTTATTACGGCATTCCTTTT
TaPr2	DQ090946	F – CTGACCTACACATCCCTGTTC R – CTCGGAAATCACCACCTTCA
TaPr3	AB029936	F – CCATCCAGATCTCACACAACTAC R – ACCACAACGCCGTCTTAAA
TaPr6	EU293132.1	F – GGGCCCTGCAAGAAGTACTG R – ACACGCATAGGCACGATGAC
TaPr9	AK333699	F – CAACTGCAGGGTTCCCAATA R – CCTAGCTACCCGTTCATCTTTC
TaRli	AY059462	F – GCTGTGTATTGGTTGTGGTATTT R – GCGATGGGTAGTATCTTTCTCC

Table 3. Characterization of bacteria isolated from the inner tissues of plants

Isolate number	Source of origin	Species designation	Accession number in GenBank
Ttl1	*Triticum timopheevii*	*Bacillus velezensis*	OR775749
Tas2	*Triticum aestivum*	*Bacillus subtilis*	OR775745
Tas2.1	*Triticum aestivum*	*Bacillus subtilis*	OR775746
Tas8.2	*Triticum aestivum*	*Bacillus subtilis*	OR775748
TV2	*Triticum aestivum*	*Bacillus velezensis*	OR775756
Ttl2	*Triticum timopheevii*	*Bacillus subtilis*	OK427265
Stl7	*Solanum tuberosum*	*Bacillus* sp.	MT613864

previously (Veselova et al., 2022). Three independent biological and three technical replications were performed for each experiment.

Statistical analysis. The experiments were carried out in triplicate with a different number of biological repetitions, from 3 to 10, depending on the type of analysis. The exact number of replicates for each analysis is indicated in the table note or figure legend. Experimental data were expressed as means ± SE, which were calculated in all treatments using MS Excel. The significance of differences was assessed by ANOVA followed by Duncan's test ($p \leq 0.05$) with STATISTICA 10.0 software.

Results

Characterization of bacteria isolated from the inner tissues of plants

Two strains, *B. subtilis* 26D and *B. subtilis* 11VM, were used in the work as reference endophytic strains with known properties and protective action against Greenbug aphid (Rumyantsev et al., 2023). Previously, it was shown that the *B. subtilis* 26D strain synthesizes surfactin, and the *B. subtilis* 11VM strain synthesizes iturin (Rumyantsev et al., 2023). *B. thuringiensis* B-6066 also induced resistance against aphids, but was not tested for the ability to synthesize lipopeptides (Veselova et al., 2019).

Two isolates from the UFRC RAS collection of microorganisms were previously sequenced using 16S RNA gene fragments: *Bacillus* sp. Stl7 (GenBank: MT613864) (isolated from the inner tissues of leaves of *Solanum tuberosum* L., Republic of Bashkortostan) and *B. subtilis* Ttl2 (GenBank: OK427265) (Sorokan et al., 2020; Veselova et al., 2022) (Table 3). For the remaining five isolates, presented in Table 3, fragments of the 16S RNA gene were sequenced in this work. Isolate of *Bacillus* sp. Ttl1 was isolated from the inner tissues of the leaves of *T. timopheevii*, the remaining isolates of *Bacillus* sp. Tas2, Tas8.2, TV2 and Tas2.1 were isolated from the inner tissues of common spring wheat leaves (*T. aestivum*) (Table 3). Isolates of *Bacillus* sp. Ttl1 and TV2 were designated as *Bacillus velezensis*. Isolates of *Bacillus* sp. Tas2, Tas2.1, Tas8.2 were designated as *B. subtilis* (Table 3).

Detection of genes of lipopeptide synthetases in the *Bacillus* spp. strains and isolates

Ten strains and isolates of the *Bacillus* spp. were tested for the presence of lipopeptide synthetases genes (Fig. 1).
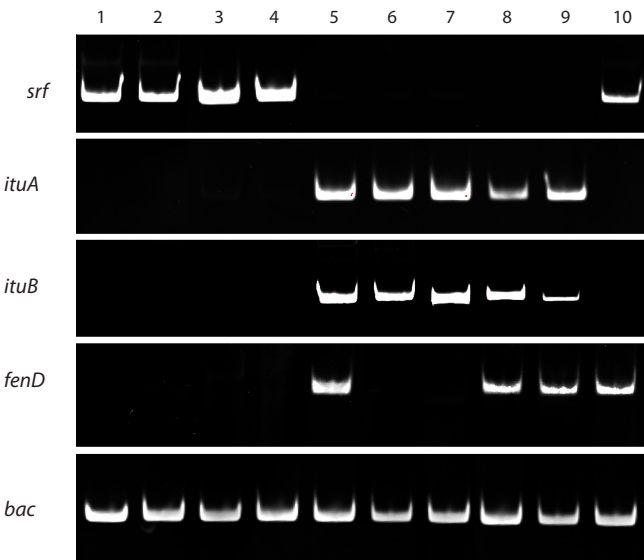
Fig. 1. PCR analysis of bacteria *Bacillus* spp. for the presence of lipopeptide synthetase genes: *srf* – surfactin synthetase, *ituA* and *ituB* – iturin synthetase, and *fenD* – fengycin synthetase, *bac* – reference gene. The samples are indicated as follows: 1 – *B. subtilis* 26D; 2 – *B. subtilis* Tas2; 3 – *B. subtilis* Tas8.2; 4 – *B. subtilis* Tas2.1; 5 – *B. subtilis*11VM; 6 – *B. velezensis* TV2; 7 – *B. velezensis* Ttl1; 8 – *B. subtilis* Ttl2; 9 – *Bacillus* sp. Stl7; 10 – *B. thuringiensis* B-6066.

As in *B. subtilis* 26D, in the strains of *B. subtilis* Tas2, Tas8.2 and Tas2.1, gene encoding surfactin synthetase *srf* was found (Fig. 1). As in the *B. subtilis* 11VM strain, in the *B. subtilis* Ttl2 strain and the *Bacillus* sp. Stl7 isolate, genes encoding iturin synthetase *ituA* and *ituB* and fengycin synthetase *fenD* were found, and in the strains of *B. velezensis*, TV2 and Ttl1, only genes encoding inturin synthetase were detected. The genes encoding surfactin and fengycin synthetase were identified in the *B. thuringiensis* B-6066 strain (Fig. 1).

Direct aphicidal effect of endophytic strains and isolates of bacteria *Bacillus* spp. and LRF on the *S. graminum*

Analysis of the aphicidal activity of ten strains and isolates of the genus *Bacillus* showed that all bacteria had high insecticidal activity against Greenbug aphid (Table 4).

Aphid mortality increased from 8 to 50–77 % during feeding on bacterial suspension (Table 4). Accordingly, the fecundity of aphids decreased. In addition, bacteria reduced the propagation coefficient of aphids by 2–5 times (Table 4).

С.Д. Румянцев, В.Ю. Алексеев, А.В. Сорокан, Г.Ф. Бурханова
Е.А. Черепанова, И.В. Максимов, С.В. Веселова

Эндофитные липопептид-синтезирующие *Bacillus* ssp.
для защиты пшеницы от *Schizaphis graminum*

2024
28 • 3

Table 4. Aphicidal (insecticidal) effect of endophytic strains and isolates of the genus *Bacillus* against *S. graminum*

Isolate/Strain	Mortality, %	Fecundity, %	Propagation coefficient
Control	8.0 ± 1.1[a]	89.1 ± 4.5[a]	2.47 ± 0.15[a]
B. subtilis 26D	66.7 ± 5.3[b]	33.3 ± 1.8[b]	0.78 ± 0.07[b]
B. subtilis Tas2	61.2 ± 5.3[c]	38.8 ± 2.9[c]	1.07 ± 0.12[c]
B. subtilis Tas8.2	76.7 ± 6.7[d]	23.3 ± 1.3[d]	0.71 ± 0.04[b]
B. subtilis Tas2.1	73.3 ± 5.7[d]	26.7 ± 1.5[d]	0.80 ± 0.10[b]
B. subtilis 11VM	72.3 ± 8.1[d]	27.7 ± 1.6[d]	0.50 ± 0.02[d]
B. velezensis TV2	49.8 ± 2.3[e]	50.2 ± 4.2[e]	0.92 ± 0.05[c]
B. velezensis Ttl1	58.3 ± 4.1[c]	41.7 ± 2.2[f]	1.03 ± 0.08[c]
B. subtilis Ttl2	69.5 ± 5.5[b]	30.5 ± 1.7[b]	1.10 ± 0.05[c]
Bacillus sp. Stl7	68.2 ± 6.4[b]	31.8 ± 1.9[b]	0.80 ± 0.03[b]
B. thuringiensis B-6066	76.8 ± 8.7[d]	23.2 ± 1.4[d]	0.10 ± 0.001[e]

N o t e. The same Latin letters in one column indicate that the values aren't statistically different according to Duncan's test ($n = 30$, $p \leq 0.05$).

The greatest aphicidal activity was shown by the *B. subtilis* 26D, *B. subtilis* 11VM, *B. subtilis* Ttl2, *B. subtilis* Tas2.1, *B. subtilis* Tas8.2 and *B. thuringiensis* B-6066 strains and the *Bacillus* sp. Stl7 isolate (Table 4).

All studied strains synthesized lipopeptides (Fig. 1). To confirm the hypothesis about the role of lipopeptides in the aphicidal activity of *Bacillus* spp. LRFs were isolated from five strains. First of all, the aphicidal activity of LRFs was tested. The aphicidal activity of LRFs of the strains *B. subtilis* 26D (LRFBs26D) and *B. subtilis* 11VM (LRFBs11VM) was studied previously (Rumyantsev et al., 2023). And it was shown that the concentration of 25 µg/ml of LRFBs26D or LRFBs11VM caused the death of 50 % of aphids, and 100 % death of aphids was caused by 150 µg/ml already on the 5th day of feeding (Rumyantsev et al., 2023). LRFs of the strains *B. subtilis* Tas8.2, *B. subtilis* Tas2.1 and *B. thuringiensis* B-6066 (LRFBsTas8.2, LRFBsTas2.1 and LRFBt B-6066) as well as the strains themselves had a negative effect on the viability of *S. graminum* at direct influence (Fig. 2). The concentration of 25 µg/ml of LRFBsTas8.2 and LRFBt B-6066 caused death in more than 50 % of aphids, but not LRFBsTas2.1. However, 100 % of aphids died on the 5th day of feeding with solutions of LRFBsTas8.2, LRFBsTas2.1 and LRFBt B-6066 at a concentration of 50 µg/ml (Fig. 2).

The plant-mediated effect of endophytes of *Bacillus* spp. and LRF on various types of resistance (antibiosis, endurance) of wheat plants against *S. graminum*

In further study of the indirect effect of bacteria on the pest, seven strains and isolates were taken that showed the highest aphicidal activity, which are presented in Table 5. All seven bacterial strains and isolates had an indirect effect on aphid mortality and propagation coefficient. Aphid mortality increased from 10.9 to 36.3 % during aphid feeding on bacteria-treated plants (Table 5). Some bacteria reduced the propagation coefficient of aphids by 1.5–2 times (Table 5). The *B. subtilis* 26D and *B. thuringiensis* B-6066 strains and the *Bacillus* sp. Stl7 isolate had the greatest effect on aphid mortality, and the propagation coefficient was most strongly influenced by the *B. subtilis* 26D, *B. subtilis* Ttl2, *B. subtilis*
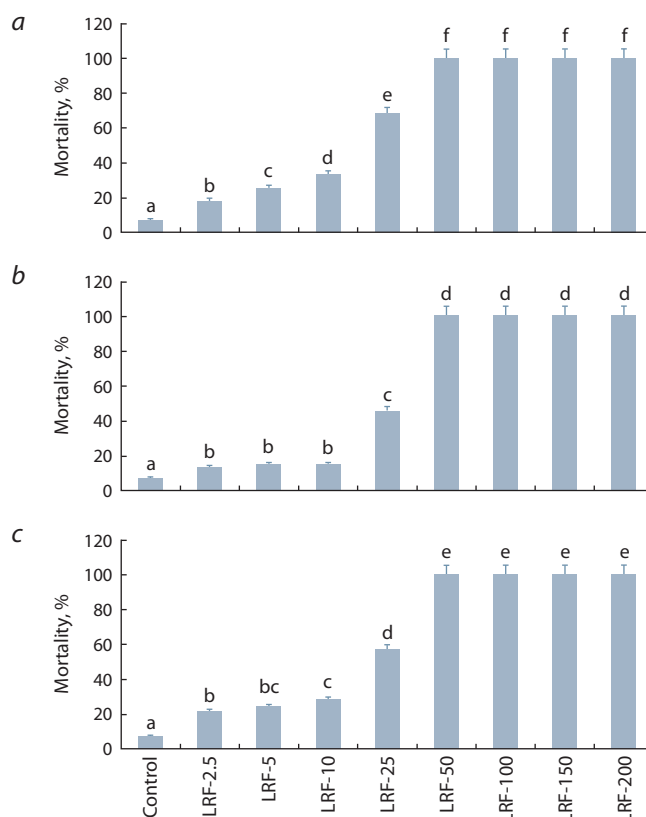
Fig. 2. Aphicidal activity of LRFs of the strains *B. subtilis* Tas8.2 (*a*), *B. subtilis* Tas2.1 (*b*) and *B. thuringiensis* B-6066 (*c*) against Greenbug aphid *S. graminum*.

Concentrations used for LRFs were 2.5, 5, 10, 25, 50, 100, 150 and 200 µg/ml. Figures present means ± SE ($n = 15$). Columns of each histogram marked with the same Latin letters indicate that the values aren't statistically different according to Duncan's test ($p \leq 0.05$).

Tas2.1 and *B. subtilis* Tas8.2 strains (Table 5). Moderate susceptible cv. SY showed low tolerance (endurance) to *S. graminum*, which manifested itself in inhibition of the growth of the first and second leaves in seedlings by 20 and 30 %, respectively (Table 5). Treatment of plants with bacterial

Table 5. The effect of endophytes of *Bacillus* spp. on the vitality of aphids and endurance of *S. graminum*-infested wheat plants

Isolate/Strain	Aphid indices of vitality (antibiosis)		Plant endurance	
	Mortality, %	Propagation coefficient	Growth rate of the 1st leaf, % of control	Growth rate of the 2nd leaf, % of control
Control (Water)	10.9 ± 1.5[a]	2.47 ± 0.15[a]	79.2 ± 6.1[a]	69.9 ± 5.1[a]
B. subtilis 26D	31.5 ± 2.2[b]	1.32 ± 0.10[b]	114.7 ± 7.3[b]	142.0 ± 12.9[b]
B. subtilis Tas2.1	22.6 ± 1.1[c]	1.54 ± 0.12[c]	110.0 ± 6.6[b]	111.0 ± 4.7[c]
B. subtilis Tas8.2	28.7 ± 1.4[b]	1.25 ± 0.09[b]	113.2 ± 8.1[b]	116.0 ± 5.8[c]
B. subtilis 11VM	24.3 ± 3.4[c]	2.10 ± 0.13[d]	103.2 ± 5.6[c]	115.0 ± 9.2[c]
B. subtilis Ttl2	22.4 ± 2.5[c]	1.60 ± 0.11[c]	107.8 ± 5.8[c]	122.5 ± 12.3[d]
Bacillus sp. Stl7	35.5 ± 3.8[d]	1.95 ± 0.12[d]	97.8 ± 4.6[d]	117.6 ± 9.5[c]
B. thuringiensis B-6066	36.3 ± 3.5[d]	2.08 ± 0.15[d]	103.6 ± 2.5[c]	121.0 ± 7.1[d]

Note. Growth rate of the 1st or 2nd leaf of control, non-treated with bacterial strains and non-infested with aphids is 100 %. The same Latin letters in one column indicate that the values aren't statistically different according to Duncan's test ($n = 30$, $p \leq 0.05$).

Table 6. Effect of lipopeptide-rich fractions (LRFs) of three *Bacillus* spp. strains on the vitality of aphids and endurance of *S. graminum*-infested wheat plants

LRF from strain	Aphid indices of vitality (antibiosis)		Plant endurance	
	Mortality, %	Propagation coefficient	Growth rate of the 1st leaf, % of control	Growth rate of the 2nd leaf, % of control
Control	10.9 ± 1.5[a]	2.47 ± 0.15[a]	79.2 ± 6.1[a]	69.9 ± 5.1[a]
LRF of *B. subtilis* 26D (surfactin)	24.9 ± 2.3[b]	1.3 ± 0.09[b]	107.0 ± 5.7[b]	102.1 ± 5.0[b]
LRF of *B. subtilis* 11VM (iturin)	20.9 ± 2.6[c]	1.2 ± 0.08[b]	98.1 ± 6.2[c]	102.3 ± 8.0[b]
LRF of *B. thuringiensis* B-6066 (fengycin + surfactin)	21.5 ± 3.9[c]	0.9 ± 0.03[c]	113.0 ± 8.2[b]	95.3 ± 3.0[c]
LRF of *B. subtilis* Tas8.2 (surfactin)	18.6 ± 4.9[c]	2.2 ± 0.30[d]	99.5 ± 3.6[c]	93.6 ± 5.4[c]
LRF of *B. subtilis* Tas2.1 (surfactin)	14.1 ± 5.1[d]	1.5 ± 0.20[b]	113.0 ± 4.0[b]	100.0 ± 2.8[b]

Note. Growth rate of the 1st or 2nd leaf of control, non-treated with bacterial suspensions and non-populated with aphids is 100 %. The same Latin letters in one column indicate that the values aren't statistically different according to Duncan's test ($n = 30$, $p \leq 0.05$).

strains and isolates increased plant resistance to Greenbug aphid by accelerating leaf growth by 10–20 % compared to the control and by 30–50 % compared to plants infested with aphids (Table 5).

Since the effect of bacteria on plants and pests depends on the synthesis of various metabolites, we tested the indirect effect of LRF from five bacterial strains presented in Table 6 on the aphid indices of vitality and endurance of wheat plants.

Major lipopeptides in the LRFBs26D and LRFBs11VM were surfactin and iturin, respectively, which was confirmed by HPLC (Rumyantsev et al., 2023). LRFBt B-6066 presumably contained a mixture of fengycin and surfactin and LRFBsTas8.2, LRFBsTas2.1 contained surfactin (Fig. 1). Growth-promoting concentrations of LRFBs26D, LRFBs 11VM, LRFBt B-6066, LRFBsTas8.2, and LRFBsTas2.1 increased plant tolerance to the pest and increased aphid mortality, but to a lesser extent than bacterial strains (Table 6).

However, the propagation coefficient of aphids decreased much more during feeding on LRF-treated plants than on plants treated with the *B. subtilis* 11VM and *B. thuringiensis* B-6066 strains. LRFBt B-6066 had the greatest influence on the propagation coefficient of aphids, which indicates the role

of fengycin in the indirect effect on aphid indices of vitality (Table 6). Thus, the results of this work show that lipopeptides, besides the direct insecticidal effect (Rumyantsev et al., 2023), manifest an indirect effect on the pest.

The plant-mediated effect of endophytes of *Bacillus* spp. and LRFs on changes in the redox status of *S. graminum*-infested wheat plants

The plant-mediated effect of endophytes of *Bacillus* spp. and their LRFs on plant endurance and indices of vitality of aphids may be connected with the start of induced systemic resistance (ISR) in plants (Rashid, Chung, 2017; Veselova et al., 2019). During the development of ISR, bacteria can affect the accumulation of ROS, both locally and systemically (Rashid, Chung, 2017).

The infestation of non-bacterial control plants by aphids led to a decrease in the content of hydrogen peroxide (Fig. 3*a*, *b*), the absence of an increase in peroxidase activity (Fig. 3*c*, *d*) and an increase in catalase activity (Fig. 3*e*, *f*) 24 and 72 hours post aphid infestation and was accompanied by low aphid mortality and low plant endurance (Table 5). In wheat plants treated with strains and isolates of *Bacillus* spp. and infested

С.Д. Румянцев, В.Ю. Алексеев, А.В. Сорокан, Г.Ф. Бурханова
Е.А. Черепанова, И.В. Максимов, С.В. Веселова

Эндофитные липопептид-синтезирующие *Bacillus* ssp.
для защиты пшеницы от *Schizaphis graminum*

2024
28 • 3
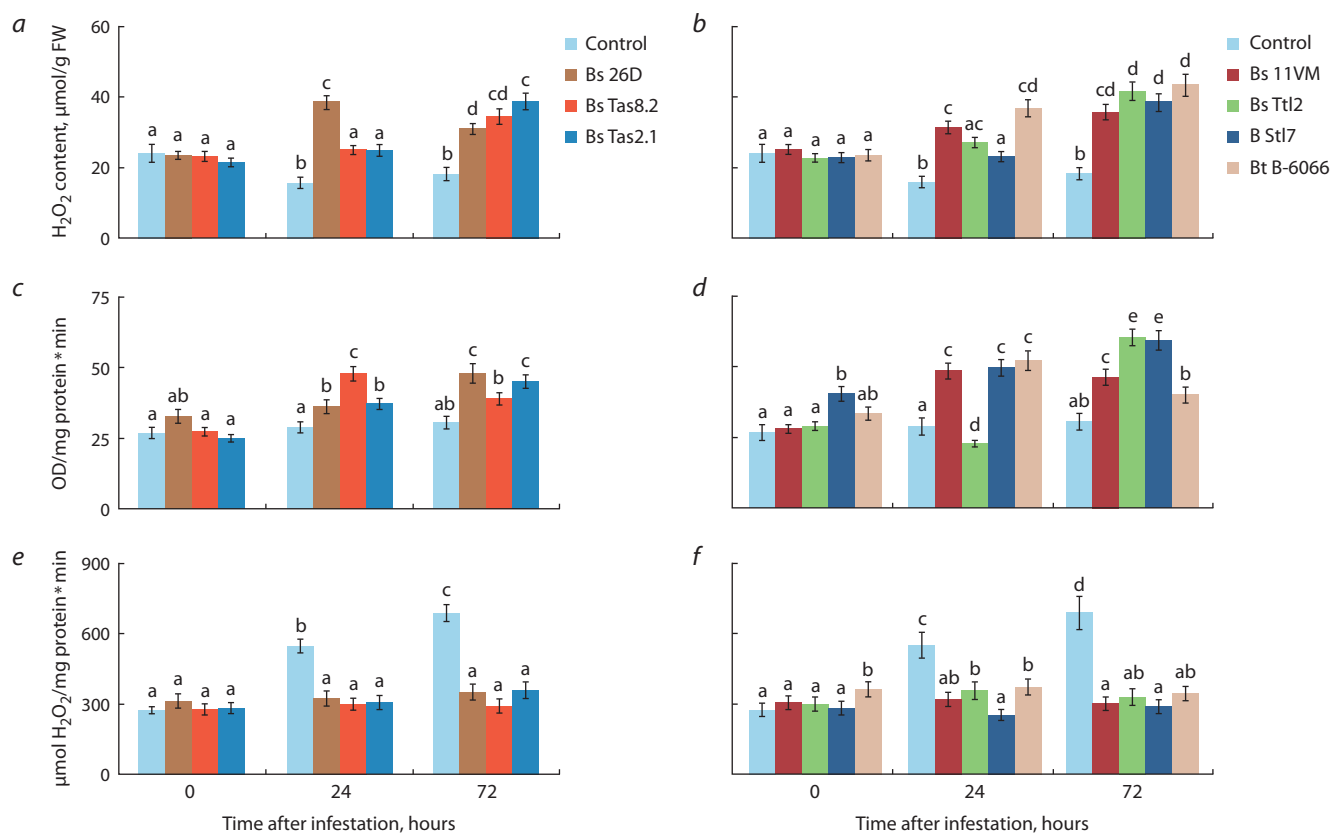
Fig. 3. The effect of endophytes of *Bacillus* spp. on the content of hydrogen peroxide (H_2O_2) (*a*, *b*), activity of peroxidase (*c*, *d*), and activity of catalase (*e*, *f*) of *S. graminum*-infested wheat plants.

Designations in the figure: 0 h – plants uninfested by aphids; Control – unbacterized plants; Bs 26D, Bs Tas8.2, Bs Tas2.1, Bs 11VM, Bs Ttl2, B Stl7 and Bt B-6066 – plants treated by the appropriate strain or isolate. Figures present means ± SE (*n* = 9). Columns of each histogram marked with the same Latin letters indicate that the values aren't statistically different according to Duncan's test (*p* ≤ 0.05).

with *S. graminum*, a sharp accumulation of H_2O_2, an increase in POD activity, no change in CAT activity compared to the control ones were found (Fig. 3).

The accumulation of H_2O_2 that was observed in bacterized plants of colonized aphids was associated with high pest mortality (Table 5, Fig. 3*a*, *b*). Treatment with strains *B. subtilis* 26D, *B. subtilis* 11VM, and *B. thuringiensis* B-6066 had the greatest effect on H_2O_2 accumulation 24 hours after aphid infestation. All strains and isolates equally increased the content of H_2O_2 after 72 hours post aphid infestation (Fig. 3*a*, *b*).

Treatment with strains *B. thuringiensis* B-6066, *B. subtilis* 11VM, *B. subtilis* Tas8.2 and *Bacillus* sp. Stl7 isolate increased POD activity earlier than treatment with strains of *B. subtilis* 26D, *B. subtilis* Ttl2 and *B. subtilis* Tas2.1 (Fig. 3*c*, *d*). The first bacteria mentioned acted 24 hours after plant infestation by aphids, and the second bacteria mentioned activated POD 72 hours after plant infestation by aphids (Fig. 3*c*, *d*). LRFs affected components of the pro-/antioxidant system of plants in the same way as bacterial strains (Fig. 4). However, LRFBs26D, LRFBt B-6066, and LRFBs11VM significantly induced the accumulation of H_2O_2 only 72 hours after plant colonization with the pest (Fig. 4*a*), unlike bacteria that induced H_2O_2 accumulation after 24 hours of feeding (Fig. 3).

LRFBs11VM and LRFBs26D increased POD activity in plants infested with aphids, as well as in plants treated with bacterial strains *B. subtilis* 11VM and *B. subtilis* 26D (Fig. 4*b*,

Fig. 3). LRFBt B-6066 increased POD activity later than treatment with the bacterial strain, only 72 hours after aphid infestation (Fig. 4*b*). Treatment of wheat plants with LRFs did not lead to an increase in CAT activity during aphid feeding (Fig. 4*c*). Such results may indicate the possible role of lipopeptides in the induction of systemic resistance against Greenbug aphid in wheat.

The plant-mediated effect of endophytes of *Bacillus* spp. and LRFs on changes in the expression of defense-related genes of PR proteins of wheat plants populated by *S. graminum*

Another indicator of the formation of systemic resistance in plants is considered to be an increase in the expression of defense-related genes of pathogenesis-related (PR) proteins, which is regulated by intermediate products of cell signaling systems (for example, H_2O_2) and phytohormones (Pieterse et al., 2014). The expression of defense-related *Pr* genes, salicylate (SA)-regulated and ethylene/jasmonate (JA)-regulated markers have been studied to test the bacteria-mediated activation of systemic resistance in *S. graminum*-infested plants. Proteins PR1, PR2 (glucanases) are markers of the SA signaling pathway. PR3 proteins (chitinases) are considered ethylene (ET)-regulated markers, and PR6 proteins (proteinase inhibitors) are considered JA-regulated markers. Proteins of PR9 (peroxidases) are both SA-responsive and JA-responsive

Fig. 4. Effect of lipopeptide-rich fractions (LRFs) of the *B. subtilis* 26D (LRFBs26D), *B. subtilis* 11VM (LRFBs11VM) and *B. thuringiensis* B-6066 (LRFBt B-6066) strains on the content of hydrogen peroxide (H₂O₂) (*a*), activity of peroxidase (*b*), and activity of catalase (*c*) of *S. graminum*-infested wheat plants.

Designations in the figure: 0 h – plants uninfested by aphids; Control – unbacterized plants; LRFBs26D, LRFBs11VM and LRFBt B-6066 – plants treated by the appropriate LRFs 24 h before aphid infestation. Figures present means ± SE (*n* = 9). Columns of each histogram marked with the same Latin letters indicate that the values aren't statistically different according to Duncan's test (*p* ≤ 0.05).

pathogenesis-related proteins (Pieterse et al., 2014). In this work, in the moderately susceptible cv. SY, a slight increase of transcripts level of the *Pr3* and *Pr6* genes, markers of the ET- and JA-signaling pathways, respectively, and an increase of the expression levels of the *Pr9* gene 72 hours after aphid colonization were found (Table 7).

The effect of bacterial treatment on the expression of *Pr* genes had a different pattern. All seven bacterial strains and isolates increased the transcripts level of the *Pr9* gene in aphid-infested plants compared to the control (Table 7). Six strains and isolates, excluding the *B. subtilis* Tas2.1 strain, increased the expression levels of the *Pr3* gene, an ET-regulated marker, in *S. graminum*-infested plants. However, only two strains, *B. thuringiensis* B-6066 and *B. subtilis* Ttl2, influenced the expression levels of the *Pr3* gene more strongly than others (Table 7).

Only four strains of *B. subtilis* 26D, 11VM, Tas8.2 and Tas2.1 significantly increased the expression levels of SA-regulated markers genes *Pr1* and *Pr2* in *S. graminum*-infested plants compared to the control. Only one strain of *B. thuringiensis* B-6066 increased the expression levels of the *Pr6* gene, a marker of the JA-signaling pathway, in *S. graminum*-infested plants (Table 7).

LRFs affected the expression of defense-related *Pr* genes of plants in the same way as bacterial strains, however, the degree of influence of LP was much higher (Fig. 5). Treatment with LRFBs26D, in which the major lipopeptide was surfactin, affected the accumulation of mRNA levels of the *Pr1* and *Pr2* genes in *S. graminum*-infested plants more than treatment with the *B. subtilis* 26D strain (Fig. 5*a*). Treatment with LRFBs11VM, in which the major lipopeptide was iturin and which also contained fengycin, increased the expression levels of the *Pr1* and *Pr3* genes in *S. graminum*-infested plants twice as much as treatment with the *B. subtilis* 11VM strain (Fig. 5*a*, *b*). The effect of LRFBt B-6066 on the expression of *Pr* genes resembled the effect of the *B. thuringiensis* B-6066 strain (Fig. 5, Table 7).

Table 7. The effect of endophytes of the *Bacillus* spp. on changes in the expression of *Pr* genes of wheat plants infested by *S. graminum*

Isolate/Strain	Genes				
	TaPr1	*TaPr2*	*TaPr3*	*TaPr6*	*TaPr9*
Control	100 ± 5[a]	100 ± 4[a]	100 ± 62[a]	100 ± 3[a]	100 ± 7[a]
Aphid	66 ± 3[a]	126 ± 5[a]	143 ± 5[ac]	270 ± 18[b]	200 ± 16[b]
Bs 26D + Aphid	382 ± 23[c]	223 ± 15[b]	240 ± 12[b]	180 ± 6[c]	375 ± 22[c]
Bs Tas2.1 + Aphid	80 ± 4[a]	200 ± 18[b]	130 ± 4[ac]	140 ± 4[d]	180 ± 15[b]
Bs Tas8.2 + Aphid	250 ± 17[d]	300 ± 24[c]	160 ± 13[c]	123 ± 5[a]	260 ± 6[d]
Bs 11VM + Aphid	170 ± 15[e]	120 ± 5[a]	200 ± 18[b]	70 ± 10[a]	405 ± 29[e]
Bs Ttl2 + Aphid	80 ± 2[a]	85 ± 4[a]	300 ± 19[d]	83 ± 3[a]	402 ± 32[e]
B Stl7 + Aphid	110 ± 3[a]	100 ± 6[a]	250 ± 17[b]	90 ± 4[a]	452 ± 37[e]
Bt B-6066 + Aphid	70 ± 3[ab]	140 ± 5[d]	350 ± 19[d]	380 ± 9[e]	180 ± 17[b]

N o t e. The same Latin letters in one column indicate that the values aren't statistically different according to Duncan's test (*n* = 9, *p* ≤ 0.05).

С.Д. Румянцев, В.Ю. Алексеев, А.В. Сорокан, Г.Ф. Бурханова
Е.А. Черепанова, И.В. Максимов, С.В. Веселова

Эндофитные липопептид-синтезирующие *Bacillus* ssp.
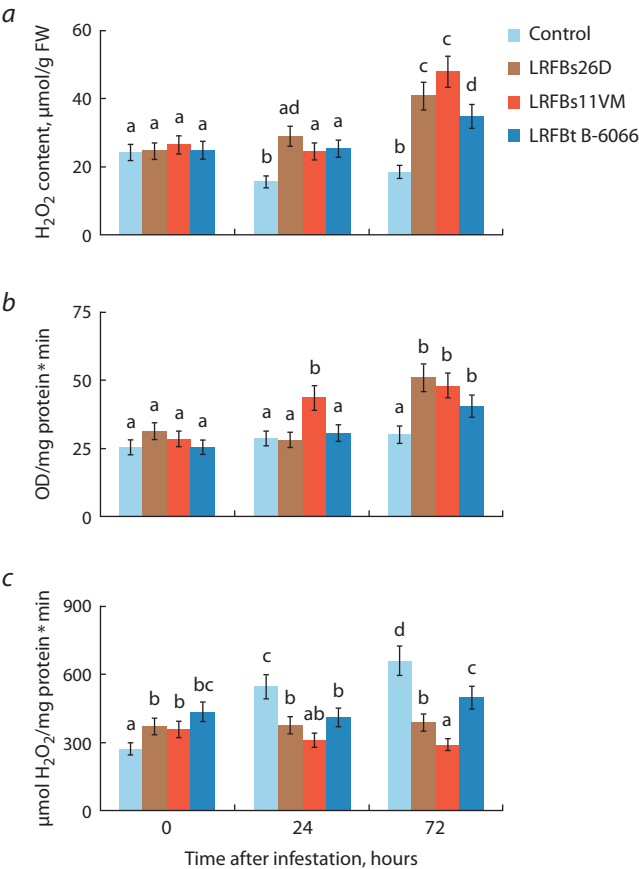для защиты пшеницы от *Schizaphis graminum*

2024
28•3
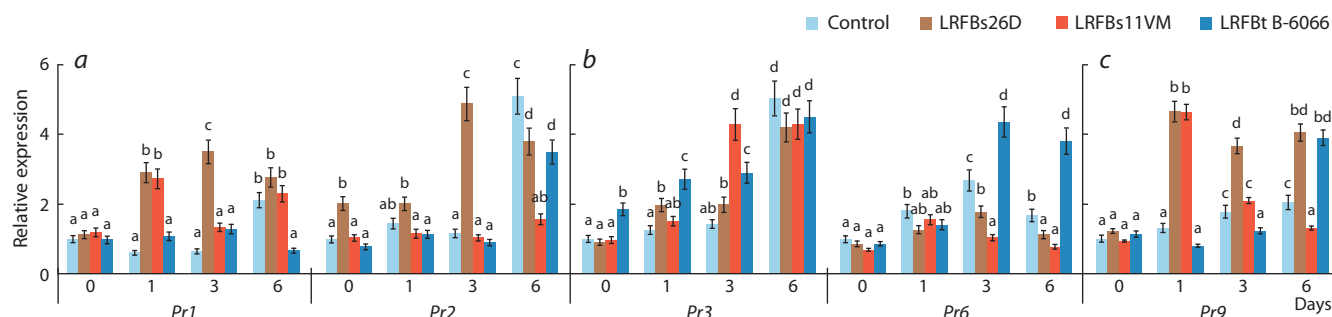
Fig. 5. Effect of lipopeptide-rich fractions (LRFs) of the *B. subtilis* 26D (LRFBs26D), *B. subtilis* 11VM (LRFBs11VM) and *B. thuringiensis* B-6066 (LRFBt B-6066) strains on the relative expression of the *Pr1* and *Pr2* genes (*a*), *Pr3* and *Pr6* genes (*b*) and *Pr9* gene (*c*) in *S. graminum*-infested wheat plants.

Designations in the figure: 0 h – plants uninfested by aphids; Control – unbacterized plants. Figures present means ± SE (*n* = 9). Columns of each histogram marked with the same Latin letters indicate that the values aren't statistically different according to Duncan's test (*p* ≤ 0.05).

However, it is worth noting that LRFBt B-6066 contained two lipopeptides – surfactin and fengycin. Treatment with LRFBt B-6066 increased the transcripts level of the *Pr3* gene as LRFBs11VM, increased the mRNA content of the *Pr9* gene as LRFBs26D and, in addition, only LRFBt B-6066 affected the expression of the *Pr6* gene in *S. graminum*-infested plants (Fig. 5). Importantly, the expression of some *Pr* genes induced by LRFs was activated later than during the treatment with the corresponding bacterial strains, 6 days after plant colonization by aphids (Fig. 5, Table 7). Thus, the results of this work show that lipopeptides have elicitor activity and induce the expression of defense-related *Pr* genes in aphid-infested plants.

Discussion

In this research, ten endophyte isolates of the genus *Bacillus* from the collection of the Laboratory of Biochemistry of Plant Immunity of the Institute of Biochemistry and Genetics UFRC RAS were studied. Although the bacteria have been isolated from the inner tissues of various plants, many of them have been tested for their ability to colonize the inner tissues of other host plants (Veselova et al., 2019, 2022; Sorokan et al., 2020; Rumyantsev et al., 2023). All studied strains and isolates were found to have lipopeptide synthetase genes (Fig. 1) and all strains and isolates showed aphicidal activity (Table 4), which was due to the synthesis of lipopeptides as the results showed (Fig. 2).

In this work, using LRF isolated from *Bacillus* spp. strains, it was proven that the aphidicidal activity of bacterial strains against Greenbug aphid was due to lipopeptides – surfactin, iturin and fengycin (Fig. 2). This coincides with the results of other authors. Nowadays, the insecticidal activity of surfactin, plipastatin (fengycin family), bacillopeptin and iturin against some species of phloem-feeding insects has been shown (Rashid et al., 2018; Rodríguez et al., 2018; Denoirjean et al., 2022). Our studies have recently shown that surfactin and iturin exhibit aphicidal activity against Greenbug aphid (Rumyantsev et al., 2023). In addition, the results of our recent work showed that commercial surfactin (Sigma, USA) exhibited the same aphicidal activity as LRF from the *B. subtilis* 26D strain (Maksimov et al., 2020). In this work, the aphicidal effect of fengycin was demonstrated for the first time (Fig. 2).

This work also shows that bacterial strains, isolates, and LRFs of three *Bacillus* spp. strains had an indirect effect on

the indices of vitality of aphids and endurance of *S. graminum*-populated wheat plants (Tables 5, 6). The weaker effect of bacteria on the mortality of aphids under the indirect effect compared to the direct effect was possibly due to the different degree of plant tissue colonization of the strains and isolates, which we showed in another work using the *B. subtilis* 11VM strain as an example (Rumyantsev et al., 2023). Thus, when testing for endophyticity, the *B. subtilis* 26D strain showed the greatest ability to reproduce in the plant tissues, the other strains and isolates studied in this work reproduced in the tissues of plants by an order of magnitude less, and the *Bacillus* sp. Stl7 isolate reduced the number of cells by two orders of quantity compared to the *B. subtilis* 26D strain (Veselova et al., 2019, 2022; Sorokan et al., 2020; Rumyantsev et al., 2023).

LRFs increased plant tolerance, but to a weaker extent than bacterial strains and isolates (Tables 3, 4). The influence of bacteria on plant growth may be associated with the synthesis of hormone-like compounds by bacteria and the effect on the availability of nutrients for plants (Eid et al., 2021). Also, the effect on plant growth may be indirect through the synthesis of metabolites with biocidal activity, which reduce the infection load on plants, and may trigger systemic resistance in plants (Eid et al., 2021). Presumably, the effect of LRFs on plant growth was indirect and was related to the stimulation of systemic resistance in plants.

Verification of the indirect action of bacterial strains and isolates and LRFs showed that both bacteria and LRFs are able to change the redox status of plants inhabited by aphids (Fig. 3, 4) and cause an oxidative burst, which subsequently induces the expression of defense-related *Pr* genes (Rashid et al., 2018; Tunsagool et al., 2019; Oukala et al., 2021). Thus, the generation of ROS during attack by phloem-feeding insects is discussed as a resistance response against pests (Koch et al., 2016; Veselova et al., 2019). The jump in the ROS generation, including H_2O_2, can lead both to direct damage to aphids and their death, and to the circumstantial effect of H_2O_2 through signaling regulation of resistance and gene expression (Rashid, Chung, 2017; Rashid et al., 2018). In addition, bacterial strains and isolates, and LRFs affected the activity of redox enzymes – POD and CAT in aphid-infested plants (Fig. 3, 4). Low catalase activity was found in aphid resistant crop phenotypes (Zhu-Salzman et al., 2004). An

increase in POD activity under the influence of bacteria led
to an improvement in the strategy of plant resistance against
insects (Rashid et al., 2018; Veselova et al., 2019; Ling et
al., 2022). To date, the role of lipopeptides in the regulation
of ROS generation and the work of redox enzymes has been
studied only during infection of plants with pathogenic fungi
(Farzand et al., 2019; Tunsagool et al., 2019). These works
showed the positive effect of fengycin, surfactin and iturin on
the activity of peroxidases in plants during the attack of fungal
pathogens (Farzand et al., 2019; Tunsagool et al., 2019). This
work demonstrates for the first time the ability of strains and
isolates *B. subtilis* Ttl2, *Bacillus* sp. Stl7 and *B. thuringiensis* B-6066, which synthesize fengycin, regulate components
of the pro-/antioxidant system of aphid-infested plants.

Bacterial strains and isolates and LRFs induced the expression of defense-related *Pr* genes, markers of hormonal
signaling pathways such as JA, SA and ethylene (Table 7,
Fig. 5). All three hormonal signaling pathways are known to
play a role in plant defense against phloem-feeding insects
and other pests (Morkunas et al., 2011; Pangesti et al., 2016).
B. subtilis induced resistance against the whitefly *Bemisia
tabaci* on tomato plants by activating SA- and JA-responsive
genes. Rhizobacteria *Pseudomonas simiae* WCS417r induced
Arabidopsis defense reaction against *Mamestra brassicae* by
activating the synthesis of camalexin and aliphatic glucosinolates, which is regulated by the ORA59-branch of the JA/ethylene signaling pathway (Pangesti et al., 2016). A series of
studies have shown that the ethylene signaling pathway is
required for the polymerization of phloem proteins, which
block phloem pores and therefore prevent aphids feeding (Fu
et al., 2014; Lu et al., 2023).

Unfortunately, there are very few works on the activation of
resistance against insects by lipopeptides. Thus, it was shown
that the bacillopeptin of the *B. velezensis* YC7010 strain,
which induces the deposition of lignin and callose in plants,
increased the resistance of rice against *Nilaparvata lugens*
(brown planthopper) (Rashid et al., 2018). Nowadays, the role
of lipopeptides in the activation of plant resistance against
various pathogens through the induction of JA/ethylene-,
ABA-, SA- and auxin-dependent response is well studied
(Tunsagool et al., 2019; Jiang et al., 2021). Our results showed
that lipopeptides surfactin, fengycin and iturin activated the
expression of defense-related *Pr* genes of the SA-, JA- and
ethylene-regulated markers in wheat against the *S. graminum*.
Our results suggest a role of fengycin in inducing the expression of ethylene-dependent genes (Fig. 5), which is consistent
with results obtained during studies of resistance to pathogen
(Waewthongrak et al., 2014). This work demonstrates for the
first time the elicitor role of fengycin in triggering the systemic
resistance of wheat plants against *S. graminum*.

Conclusion
In the ten studied strains and isolates of endophytes of the
genus *Bacillus*, lipopeptide synthetase genes were found,
and all bacteria had aphicidal activity. This study shows that
lipopeptides play a role in the defense of plants from phloem-feeding insects through a direct and an indirect mechanism of
action. We have discovered new promising strains and isolates
of endophytes of the genus *Bacillus*, which can become the

basis for future biocontrol agents against aphids. The search
for new bacteria active against phloem-feeding insects can be
conducted by the presence of lipopeptide synthetase genes in
the bacterial genome.

References
Alekseev V.Y., Veselova S.V., Rumyantsev S.D., Burkhanova G.F., Cherepanova E.A., Maksimov I.V. Bacteria of the genus *Bacillus* and their lipopeptides enhance endurance of wheat plants to the greenbug aphid *Schizaphis graminum* Rond. *AIP Conf. Proc.* 2021; 2388(1):030001. DOI 10.1063/5.0071841

Andrić S., Meyer T., Rigolet A., Prigent-Combaret C., Höfte M., Balleux G., Steels S., Hoff G., De Mot R., McCann A., De Pauw E., Arias A.A., Ongena M. Lipopeptide interplay mediates molecular interactions between soil *Bacilli* and *Pseudomonads*. *Microbiol. Spectr.* 2021;9(3):e0203821. DOI 10.1128/spectrum.02038-21

Bindschedler L.V., Minibayeva F., Gardner S.L., Gerrish C., Davies D.R., Bolwell G.P. Early signaling events in the apoplastic oxidative burst in suspension cultured French bean cells involve cAMP and Ca^{2+}. *New Phytol.* 2001;151(1):185-194. DOI 10.1046/j.1469-8137.2001.00170.x

Denoirjean T., Ameline A., Couty A., Dubois F., Coutte F., Doury G. Effects of surfactins, *Bacillus* lipopeptides, on the behavior of an aphid and host selection by its parasitoid. *Pest Manag. Sci.* 2022; 78(3):929-937. DOI 10.1002/ps.6702

Eid A.M., Fouda A., Abdel-Rahman M.A., Salem S.S., Elsaied A., Oelmüller R., Hijri M., Bhowmik A., Elkelish A., Hassan S.E.-D. Harnessing bacterial endophytes for promotion of plant growth and biotechnological applications: an overview. *Plants.* 2021;10(5):935. DOI 10.3390/plants10050935

Farzand A., Moosa A., Zubair M., Khan A.R., Massawe V.C., Tahir H.A.S., Sheikh S.M.M., Ayaz M., Gao X. Suppression of *Sclerotinia sclerotiorum* by the induction of systemic resistance and regulation of antioxidant pathways in tomato using fengycin produced by *Bacillus amyloliquefaciens* FZB42. *Biomolecules.* 2019;9(10): 613. DOI 10.3390/biom9100613

Fu M., Xu M., Zhou T., Wang D., Tian S., Han L., Dong H., Zhang C. Transgenic expression of a functional fragment of harpin protein Hpa1 in wheat induces the phloem-based defence against English grain aphid. *J. Exp. Bot.* 2014;65(6):1439-1453. DOI 10.1093/jxb/ert488

Jiang M., Pang X., Liu H., Lin F., Lu F., Bie X., Lu Z., Lu Y. Iturin A induces resistance and improves the quality and safety of harvested cherry tomato. *Molecules.* 2021;26(22):6905. DOI 10.3390/molecules26226905

Koch K.G., Chapman K., Louis J., Heng-Moss T., Sarath G. Plant tolerance: a unique approach to control hemipteran pests. *Front. Plant Sci.* 2016;7:1363. DOI 10.3389/fpls.2016.01363

Ling S., Zhao Y., Sun S., Zheng D., Sun X., Zeng R., Chen D., Song Y. Enhanced anti-herbivore defense of tomato plants against *Spodoptera litura* by their rhizosphere bacteria. *BMC Plant Biol.* 2022; 22(1):254. DOI 10.1186/s12870-022-03644-3

Lu K., Zhang L., Qin L., Chen X., Wang X., Zhang M., Dong H. Importin β1 mediates nuclear entry of EIN2C to confer the phloem-based defense against aphids. *Int. J. Mol. Sci.* 2023;24(10):8545. DOI 10.3390/ijms24108545

Maksimov I.V., Sorokan' A.V., Cherepanova E.A., Yarullina L.G. Effects of salicylic and jasmonic acids on the components of pro/antioxidant system in potato plants infected with late blight. *Russ. J. Plant Physiol.* 2011;58(2):299-306. DOI 10.1134/S1021443711010109

Maksimov I.V., Blagova D.K., Veselova S.V., Sorokan A.V., Burkhanova G.F., Cherepanova E.A., Sarvarova E.R., Rumyantsev S.D., Alekseev V.Yu., Khayrullin R.M. Recombinant *Bacillus subtilis* 26DCryChS line with gene *Btcry1Ia* encoding Cry1Ia toxin from *Bacillus thuringiensis* promotes integrated wheat defense against pathogen *Stagonospora nodorum* Berk. and greenbug *Schizaphis*

С.Д. Румянцев, В.Ю. Алексеев, А.В. Сорокан, Г.Ф. Бурханова
Е.А. Черепанова, И.В. Максимов, С.В. Веселова

Эндофитные липопептид-синтезирующие *Bacillus* ssp.
для защиты пшеницы от *Schizaphis graminum*

2024
28•3

graminum Rond. *Biol. Control*. 2020;144:104242. DOI 10.1016/
j.biocontrol.2020.104242

Miljaković D., Marinković J., Balešević-Tubić S. The significance
of *Bacillus* spp. in disease suppression and growth promotion of
field and vegetable crops. *Microorganisms*. 2020;8(7):1037. DOI
10.3390/microorganisms8071037

Morkunas I., Mai V.C., Gabrys B. Phytohormonal signaling in plant
responses to aphid feeding. *Acta Physiol. Plant*. 2011;33(6):2057-
2073. DOI 10.1007/s11738-011-0751-7

Oukala N., Aissat K., Pastor V. Bacterial endophytes: the hidden ac-
tor in plant immune responses against biotic stress. *Plants*. 2021;
10(5):1012. DOI 10.3390/plants10051012

Pangesti N., Reichelt M., van de Mortel J.E., Kapsomenou E., Gershen-
zon J., van Loon J.J., Dicke M., Pineda A. Jasmonic acid and ethyle-
ne signaling pathways regulate glucosinolate levels in plants during
rhizobacteria-induced systemic resistance against a leaf-chewing
herbivore. *J. Chem. Ecol*. 2016;42(12):1212-1225. DOI 10.1007/
s10886-016-0787-7

Pieterse C.M., Zamioudis C., Berendsen R.L., Weller D.M., Van
Wees S.C., Bakker P.A. Induced systemic resistance by beneficial
microbes. *Annu. Rev. Phytopathol*. 2014;52:347-375. DOI 10.1146/
annurev-phyto-082712-102340

Radchenko E.E., Abdullaev R.A., Anisimova I.N. Genetic resources
of cereal crops for aphid resistance. *Plants*. 2022;11(11):1490. DOI
10.3390/plants11111490

Rani S., Kumar P., Dahiya P., Maheshwari R., Dang A.S., Suneja P.
Endophytism: a multidimensional approach to plant–prokaryo-
tic microbe interaction. *Front. Microbiol*. 2022;13:861235. DOI
10.3389/fmicb.2022.861235

Rashid M.H., Chung Y.R. Induction of systemic resistance against in-
sect herbivores in plants by beneficial soil microbes. *Front. Plant
Sci*. 2017;8:1816. DOI 10.3389/fpls.2017.01816

Rashid M.H., Kim H.-J., Yeom S.-I., Yu H.-A., Manir M.M., Moon S.-S.,
Kang Y.J., Chung Y.R. *Bacillus velezensis* YC7010 enhances plant
defenses against brown planthopper through transcriptomic and me-
tabolic changes in rice. *Front. Plant Sci*. 2018;9:1904. DOI 10.3389/
fpls.2018.01904

Rodríguez M., Marín A., Torres M., Béjar V., Campos M., Sampe-
dro I. Aphicidal activity of surfactants produced by *Bacillus atro-
phaeus* L193. *Front. Microbiol*. 2018;9:3114-3123. DOI 10.3389/
fmicb.2018.03114

Rumyantsev S.D., Alekseev V.Y., Sorokan A.V., Burkhanova G.F.,
Cherepanova E.A., Garafutdinov R.R., Maksimov I.V., Veselo-
va S.V. Additive effect of the composition of endophytic bacteria
Bacillus subtilis on systemic resistance of wheat against greenbug
aphid *Schizaphis graminum* due to lipopeptides. *Life*. 2023;13(1):
214. DOI 10.3390/life13010214

Sorokan A., Cherepanova E., Burkhanova G., Veselova S., Rumyan-
tsev S., Alekseev V., Mardanshin I., Sarvarova E., Khairullin R.,
Benkovskaya G., Maksimov I. Endophytic *Bacillus* spp. as a pro-
spective biological tool for control of viral diseases and non-vector
Leptinotarsa decemlineata Say. in *Solanum tuberosum* L. *Front.
Microbiol*. 2020;11:569457. DOI 10.3389/fmicb.2020.569457

Tunsagool P., Leelasuphakul W., Jaresitthikunchai J., Phaonakrop N.,
Roytrakul S., Jutidamrongphan W. Targeted transcriptional and pro-
teomic studies explicate specific roles of *Bacillus subtilis* iturin A,
fengycin, and surfactin on elicitation of defensive systems in man-
darin fruit during stress. *PLoS One*. 2019;14(5):e0217202. DOI
10.1371/journal.pone.0217202

Veselova S.V., Nuzhnaya T.V., Maksimov I.V. The effect of 1-methyl-
cyclopropene on the components of pro- and antioxidant systems of
wheat and the development of defense reactions in fungal pathoge-
nesis. *Appl. Biochem. Microbiol*. 2014;50(5):516-523. DOI 10.1134/
S0003683814050111

Veselova S.V., Burkhanova G.F., Rumyantsev S.D., Blagova D.K.,
Maksimov I.V. Strains of *Bacillus* spp. regulate wheat resistance to
greenbug aphid *Schizaphis graminum* Rond. *Appl. Biochem. Micro-
biol*. 2019;55(1):41-47. DOI 10.1134/S0003683819010186

Veselova S.V., Sorokan A.V., Burkhanova G.F., Rumyantsev S.D.,
Cherepanova E.A., Alekseev V.Y., Sarvarova E.R., Kasimova A.R.,
Maksimov I.V. By modulating the hormonal balance and ribonucle-
ase activity of tomato plants *Bacillus subtilis* induces defense re-
sponse against potato virus X and potato virus Y. *Biomolecules*.
2022;12(2):288. DOI 10.3390/biom12020288

Waewthongrak W., Leelasuphakul W., McCollum G. Cyclic lipopep-
tides from *Bacillus subtilis* ABS–S14 elicit defense-related gene
expression in citrus fruit. *PLoS One*. 2014;9(10):e109386. DOI
10.1371/journal.pone.0109386

Xia Y., Liu J., Chen C., Mo X., Tan Q., He Y., Wang Z., Yin J., Zhou G.
The multifunctions and future prospects of endophytes and their
metabolites in plant disease management. *Microorganisms*. 2022;
10(5):1072. DOI 10.3390/microorganisms10051072

Zhu-Salzman K., Salzman R.A., Ahn J.-E., Koiwa H. Transcriptional
regulation of sorghum defense determinants against a phloem-feed-
ing aphid. *Plant Physiol*. 2004;134(1):420-431. DOI 10.1104/pp.103.
028324

Influence of leptin administration to pregnant mice on fetal gene expression and adaptation to sweet and fatty food in adult offspring of different sexes

E.I. Denisova (iD), E.N. Makarova (iD) (✉)

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ enmakarova@gmail.com

Abstract. Elevated leptin in pregnant mice improves metabolism in offspring fed high-calorie diet and its influence may be sex-specific. Molecular mechanisms mediating leptin programming action are unknown. We aimed to investigate programming actions of maternal leptin on the signaling function of the placenta and fetal liver and on adaptation to high-calorie diet in male and female offspring. Female C57BL/6J mice received leptin injections in mid-pregnancy. Gene expression was assessed in placentas and in the fetal brain and liver at the end of pregnancy. Metabolic parameters and gene expression in the liver, brown fat and hypothalamus were assessed in adult male and female offspring that had consumed sweet and fatty diet (SFD: chow, lard, sweet biscuits) for 2 weeks. Females had lower blood levels of leptin, glucose, triglycerides and cholesterol than males. Consuming SFD, females had increased *Ucp1* expression in brown fat, while males had accumulated fat, decreased blood triglycerides and liver *Fasn* expression. Leptin administration to mothers increased *Igf1* and *Dnmt3b* expression in fetal liver, decreased post-weaning growth rate, and increased hypothalamic *Crh* expression in response to SFD in both sexes. Only in male offspring this administration decreased expression of *Fasn* and *Gck* in the mature liver, increased fat mass, blood levels of glucose, triglycerides and cholesterol and *Dmnt3a* expression in the fetal liver. The results suggest that the influence of maternal leptin on the expression of genes encoding growth factors and DNA methyltransferases in the fetal liver may mediate its programming effect on offspring metabolic phenotypes.

Key words: adaptation to high-calorie food; developmental programming; leptin; mice; pregnancy.

Введение лептина беременным мышам влияет на экспрессию генов у плодов и адаптацию к сладкой и жирной пище у взрослых потомков разного пола

Е.И. Денисова (iD), Е.Н. Макарова (iD) (✉)

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ enmakarova@gmail.com

Аннотация. Повышенный уровень лептина в период беременности у самок мышей оказывает благоприятное действие на метаболические показатели их зрелого потомства при потреблении последним высококалорийной пищи, и это влияние может зависеть от пола. Молекулярные механизмы, опосредующие программирующее действие лептина, неизвестны. Целью представленной работы было изучение программирующего действия материнского лептина на сигнальную функцию плацент и печени плодов, а также на адаптацию к высококалорийной диете у потомства в зависимости от пола. Самкам мышей линии C57BL/6J вводили лептин в середине беременности. В конце беременности в плацентах, мозге и печени плодов оценивали экспрессию генов. У взрослого потомства обоего пола оценивали метаболические показатели и экспрессию генов в печени, буром жире и гипоталамусе после двухнедельного потребления стандартной либо сладко-жирной диеты (СЖД: гранулы стандартного корма, сало, сладкое печенье). У самок наблюдался более низкий уровень лептина, глюкозы, триглицеридов и холестерина в крови, чем у самцов. Потребление СЖД увеличивало экспрессию гена *Ucp1* в буром жире у самок, тогда как у самцов накапливался жир, снижались уровень тригли-

Е.И. Денисова
Е.Н. Макарова

Влияние материнского лептина
на плоды и взрослое потомство

2024
28•3

церидов в крови и экспрессия гена *Fasn* в печени. Введение лептина матерям увеличивало экспрессию генов *Igf1* и *Dnmt3b* в печени плодов, снижало скорость роста после отъема от матери и повышало экспрессию *Crh* в гипоталамусе в ответ на СЖД у взрослых потомков обоих полов. Только у самцов введение лептина матерям снижало экспрессию генов *Fasn* и *Gck* в печени, увеличивало жировую массу, уровни глюкозы, триглицеридов и холестерина в крови, а также экспрессию гена *Dmnt3a* в печени плодов. Полученные результаты позволяют предположить, что влияние материнского лептина на экспрессию генов, кодирующих факторы роста и ДНК-метилтрансферазы в печени плодов, может опосредовать его программирующий эффект на метаболический фенотип потомства.

Ключевые слова: адаптация к высококалорийной пище; программирование развития; лептин; мыши; беременность.

Introduction

Obesity and related metabolic diseases are one of the major problems in modern medicine. The potentiating effect of maternal obesity on the development of obesity in the offspring is considered as one of the reasons for the widespread prevalence of obesity (Shrestha et al., 2020; Schoonejans et al., 2021). In this regard, the study of the possible mechanisms responsible for mediating the effects of early-life environment on susceptibility to obesity later in life is of particular relevance.

The adipocyte hormone leptin can have a programming effect on the development of offspring. It was shown in laboratory models that elevated blood levels of leptin in pregnant females, whether caused by genetic disorders or leptin administration, may have a beneficial effect on glucose metabolism and obesity in offspring fed a high-calorie diet (Stocker, Cawthorne, 2008; Pollock et al., 2015; Talton et al., 2016; Denisova et al., 2021). It was also shown that the programming effects of maternal leptin can be different in offspring of different sexes (Nilsson et al., 2003; Makarova et al., 2013). The study of the molecular and physiological mechanisms that mediate the programming effect of leptin may contribute to the elaboration of methods for correcting individual development to reduce the risk of metabolic disease.

In most cases, the development of obesity is promoted by the consumption of high-calorie sweet and fatty food. Adaptation to the consumption of this type of food is expressed in a decrease in the amount of food consumed, storage of excess energy in adipose tissue, and an increase in energy expenditure (Duca et al., 2014). These adaptive responses are associated with changes in the expression of orexigenic and anorexigenic neuropeptides in the hypothalamus (Cone, 2005), activation of thermogenesis in brown adipose tissue (Even, 2011), and changes in the activity of enzymes related to glucose and lipid metabolism in the liver and other organs (Akieda-Asai et al., 2013). Ability to adapt to the consumption of high-calorie foods may affect the rate and degree of obesity development. However, the effect of maternal leptin on adaptation to sweet and fatty foods has not yet been investigated.

The programming effect of maternal leptin on the development of offspring can be mediated via epigenetic modifications, including methylation of regulatory regions of genes and changes in the expression of signaling factors that affect the growth and maturation of organs and tissues in fetuses (Reynolds et al., 2017). Insulin-like growth factors 1 and 2 (IGF1, IGF2) play a significant role in the somatic development of the fetus (Petry et al., 2010; Xiagedeer et al., 2020; Hattori et al., 2021). These factors are synthesized and secreted into the blood of the fetus by both placenta and fetal liver (Nawathe et al., 2016). The effect of maternal leptin on the signal-

ing function of the placenta and fetal liver has not yet been studied.

The aim of this study is to investigate the effect of increased leptin levels in pregnant females on the signaling function of the placenta and fetal liver and on the adaptation to the consumption of high-calorie sweet and fatty foods in mature offspring of different sexes in mice.

Materials and methods

Animals and experimental design. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Independent Ethics Committee of the Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences (protocol number 76, 07.04.2021).

Experiments were conducted with C57BL/6J mice housed at the vivarium of the Institute of Cytology and Genetics, Novosibirsk, Russia. The animals were kept at a 12-h daylight cycle with free access to water and standard chow for the conventional maintenance and breeding of rodents (BioPro Company, Novosibirsk, Russia). Mature females were mated to males of the same strain. Mating was confirmed by the presence of a copulation plug. The appearance of the plug signified day 0 of pregnancy. The females were administered 0.2 mg/kg of recombinant murine leptin (Peprotech, United Kingdom) or the same volume of normal saline on days 11, 12, and 13 of pregnancy. The injections were done subcutaneously in the shoulder area. It has been shown that during this period, sexual differentiation begins in fetuses (Hacker et al., 1995) and there is a peak in the formation of hypothalamic neurons that regulate energy intake and expenditure (Ishii, Bouret, 2012). As we showed earlier, the food intake of females reduces in response to leptin administration, and the offspring demonstrate sensitivity to its programming effect during this period of pregnancy (Denisova et al., 2021).

To study the effect of leptin administration on the fetal growth and expression of genes in fetuses and placentas, 6 leptin-treated and 6 control females were sacrificed at the pregnancy day (PD) 18 by displacement of the cervical vertebrae, fetuses and placentas were removed and weighed. Samples of placentas and fetal liver and brain were placed in liquid nitrogen. To measure gene expression, two tissue samples of the placentas and fetuses of each sex were selected from each litter and combined in equal representation, taking into account the RNA concentration after RNA isolation.

In another group, the mated females were monitored to record parturition and the number of pups, and the day of delivery was designated as postpartum day (PPD) 0. Females with a litter of less than 6 pups did not participate in the further experiment. If there were more than 7 pups in the litter, it was

Table 1. TaqMan Gene Expression Assays used for relative quantitative real-time PCR

Protein	Function	Gene	Gene expression assay ID
Agouti-related neuropeptide	Orexigenic neuropeptide	*Agrp*	Mm00475829_g1
Beta-actin		*Actb*	Mm00607939_s1
Carnitine palmitoyltransferase 1a	Beta-oxidation of long-chain fatty acids (liver)	*Cpt1a*	Mm01231183_m1
Carnitine palmitoyltransferase 1b	Beta-oxidation of long-chain fatty acids (muscles, BAT)	*Cpt1b*	Mm00487191_g1
Corticotropin-releasing hormone	Hypothalamic signaling	*Crh*	Mm01293920_s1
DNA methyltransferase 3 alpha	De-novo DNA methylation	*Dnmt3a*	Mm00432881_m1
DNA methyltransferase 3 beta	De-novo DNA methylation	*Dnmt3b*	Mm01240113_m1
Fatty acid synthase	Fatty acid synthesis	*Fasn*	Mm00662319_m1
Fibroblast growth factor 21	Influence on carbohydrate and lipid metabolism	*Fgf21*	Mm00840165_g1
Glucokinase	Glucose phosphorylation	*Gck*	Mm00439129_m1
Glucose-6-phosphatase, catalytic	Glucose-6-phosphate dephosphorylating	*G6pc*	Mm00839363_m1
Insulin receptor	Insulin signaling	*Insr*	Mm01211875_m1
Insulin-like growth factor 1	Fetal growth and development	*Igf1*	Mm00439560_m1
Insulin-like growth factor 2	Fetal growth and development	*Igf2*	Mm00439564_m1
Insulin-like growth factor 2 receptor	Attenuation of IGF2 signaling	*Igf2r*	Mm00439576_m1
Klotho beta	Enables FGF21 binding activity	*Klb*	Mm00473122_m1
Leptin receptor	Leptin signaling	*Lepr*	Mm00440181_m1
Peroxisome proliferator-activated receptor alpha	Regulation of lipid metabolism	*Ppara*	Mm0040939_m1
Phosphoenolpyruvate carboxykinase 1, cytosolic	Regulation of gluconeogenesis	*Pck1*	Mm01247058_m1
Pro-opiomelanocortin	Anorexigenic signaling	*Pomc*	Mm00435874_m1
Peptidylprolyl isomerase A		*Ppia*	Mm02342430_g1
Pyruvate kinase liver and red blood cell	Regulation of glycolysis	*Pklr*	Mm00443090_m1
Solute carrier family 2 (facilitated glucose transporter), member 4 (GLUT4)	Glucose transporter activated by insulin	*Slc2a4*	Mm00436615_m1
Sodium-coupled neutral amino acid transporter 1	Amino acid transport	*Slc38a1*	Mm00506391_m1
Sodium-coupled neutral amino acid transporter 2	Amino acid transport	*Slc38a2*	Mm00628416_m1
Sodium-coupled neutral amino acid transporter 4	Amino acid transport	*Slc38a4*	Mm00459056_m1
Uncoupling protein 1 (mitochondrial, proton carrier)	Thermogenesis	*Ucp1*	Mm01244861_m1
Uncoupling protein 3 (mitochondrial, proton carrier)	Mitochondrial anion carrier protein	*Ucp3*	Mm01163394_m1

adjusted to 7 on PPD 0. There were 9 leptin-treated litters and 8 control litters. The females and pups were weighed on PPDs 0, 7, 14, 21, and 28. The offspring were weaned from their mothers at PPD 28.

To assess the effect of maternal leptin on the metabolic parameters of mature offspring, two males and two females from each litter were housed individually after weaning. At the age of 10 weeks, some of the offspring begun to receive a sweet and fatty diet (SFD): sweet butter cookies and lard were added to standard chow, and the other part of the animals remained on standard diet (SD). There were 8 experimental groups with 6–7 animals in each group: males and females consuming SFD and males and females consuming SD born to control mothers and males and females consuming SFD and males and females consuming SD born to leptin-treated mothers. The weight of standard chow, fat and cookies eaten per week was measured, and energy intake was calculated (lard – 8 kcal/g, cookies – 4.58 kcal/g, and standard chow – 3 kcal/g). The total amount of energy consumed was calculated and related to body weight.

After 2 weeks of SFD eating, the animals were decapitated, the weight of the liver, interscapular brown fat, and subcutaneous and intraperitoneal fat were measured. To assess the effect of leptin on blood biochemical parameters and gene expression, blood samples were collected, liver, muscle, brown fat and hypothalamus samples were placed in liquid nitrogen and then stored at –80 °C.

Plasma assays. Concentrations of leptin and FGF21 were measured using Mouse Leptin ELISA Kit (EMD Millipore, St. Charles, MO, USA) and Quantikine® ELISA Mouse/Rat FGF-21 Immunoassay (R&D Systems, Minneapolis, USA).

Concentrations of glucose, triglycerides, and cholesterol were measured colorimetrically using Fluitest GLU, Fluitest TG, and Fluitest CHOL (Analyticon® Biotechnologies AG Am Mühlenberg 10, 35,104 Lichtenfels, Germany), respectively.

Е.И. Денисова
Е.Н. Макарова

Влияние материнского лептина
на плоды и взрослое потомство

2024
28•3

Relative quantitative real-time PCR. Gene expression was measured using relative quantitative real-time PCR. Total RNA was isolated from tissue samples using the ExtractRNA kit (Evrogen, Moscow, Russia) according to the manufacturer's instructions. First-strand cDNA was synthesized using Moloney murine leukemia virus (MMLV) reverse transcriptase (Evrogen, Moscow, Russia) and oligo(dT) as a primer. TaqMan gene expression assays (Thermo Fisher Scientific, Waltham, MA USA) indicated in Table 1 were used for relative quantitative real-time PCR with β-actin (*Actb*) and cyclophilin (*Ppia*) as an endogenous control.

Sequence amplification and fluorescence detection were performed on a QuantStudio™ system. Relative quantification was performed by the comparative threshold cycle (CT) method.

Statistical analyses. Data were analyzed with the STATISTICA 10.0 program. Descriptive statistic was used to determine means and standard error (SE) of the mean. Data on body weight and food intake were analyzed using Repeated Measures ANOVA with factors "maternal treatment" (administration of leptin or saline), "sex", and "age" (from 4 to 10 weeks) for offspring when kept on a standard diet. When kept on a sweet and fatty diet, data on energy intake were analyzed using Repeated Measures ANOVA with factors "diet" (SD and SFD), "maternal treatment" and "age" (from 10 to 12 weeks) and data on weight gain were analyzed using two-way ANOVA with factors "diet" and "maternal treatment" separately for male and female offspring. Morphometric, metabolic and hormonal parameters and gene expression were analyzed initially by three-way ANOVA with factors "maternal treatment," "diet," and "sex" and then separately by two-way ANOVA in offspring consuming SD or SFD with factors "sex" and "maternal treatment," or in males and females with factors "maternal treatment" and "diet". To identify the effect of leptin administration on the weight of fetuses and placentas and gene expression in fetuses and placentas, two-way ANOVA

was used with factors "sex" and "maternal treatment". To assess intergroup differences, post hoc Newman–Keuls test was used. The comparisons between single parameters were performed with a two-tailed Student's *t*-test. The results on the graphs are presented as mean±SE. Significance was determined as *p* < 0.05.

Results

The effect of leptin administration to pregnant mice on body weight and energy intake in offspring of different sexes when kept on SD

The administration of leptin to pregnant females had no effect on body weight (BW) of the offspring at birth and during the period of maternal care (PPDs 1–28); no sex differences in BW were observed during this period either.

After weaning, males as compared to females had a higher growth rate and were significantly heavier (Fig. 1*a*). The administration of leptin to mothers affected the dynamics of weight gain in both males and females; it reduced the growth rate of the offspring in the first two weeks after weaning (Fig. 1*a*). Females consumed more energy per unit of body weight than males (Fig. 1*b*), leptin administration to mothers had no effect on offspring energy intake.

The effect of leptin administration to pregnant mice on energy intake and body weight in offspring of different sexes when kept on SFD

Energy consumed with SFD changed dramatically in the course of the experiment: it increased sharply in comparison with the control in the first week, and returned to normal in the second week in mice of both sexes (Fig. 2*a*). The leptin administration to mothers had no effect on the dynamics of energy intake with SFD in the offspring. At the same time, there were sex differences in BW changes resulting from SFD consumption (*p* < 0.05, "sex" × "diet", 3-way ANOVA): SFD
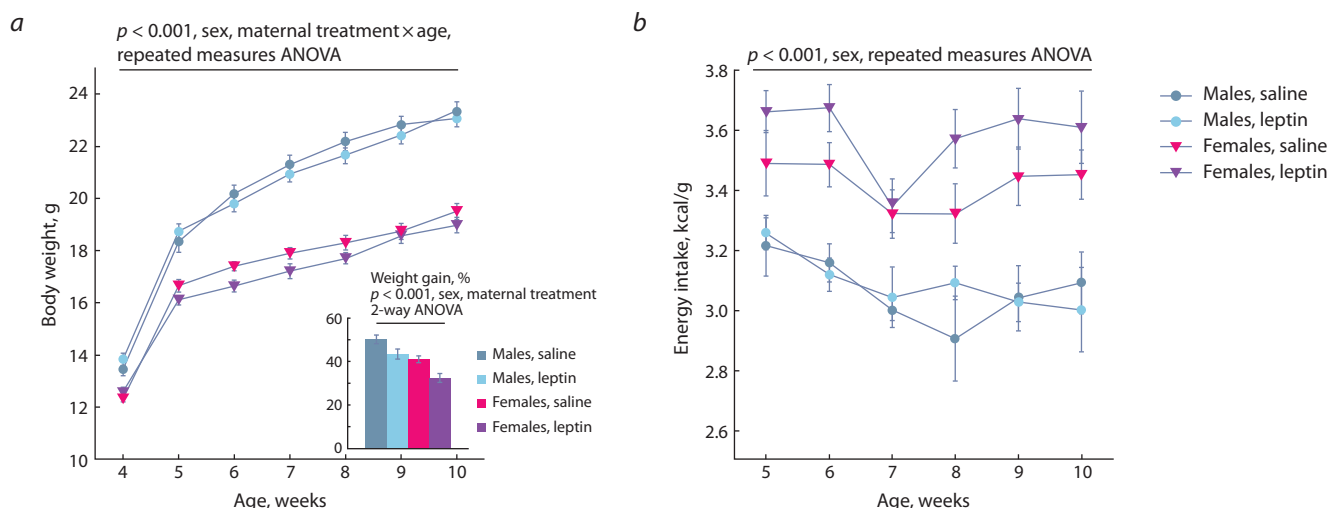
Fig. 1. The effect of leptin administration to female mice at 11–13 days of pregnancy on weight gain during the first two weeks after weaning (*a*) and body weight (*a*) and weekly energy intake related to body weight (*b*) at the age of 4–10 weeks in offspring of different sexes when consuming a standard diet.

Data are means ± SE from 12–14 animals in every group. Weight gain was calculated as the difference in weight in the first two weeks after weaning divided by weight at the weaning and expressed as a percentage.
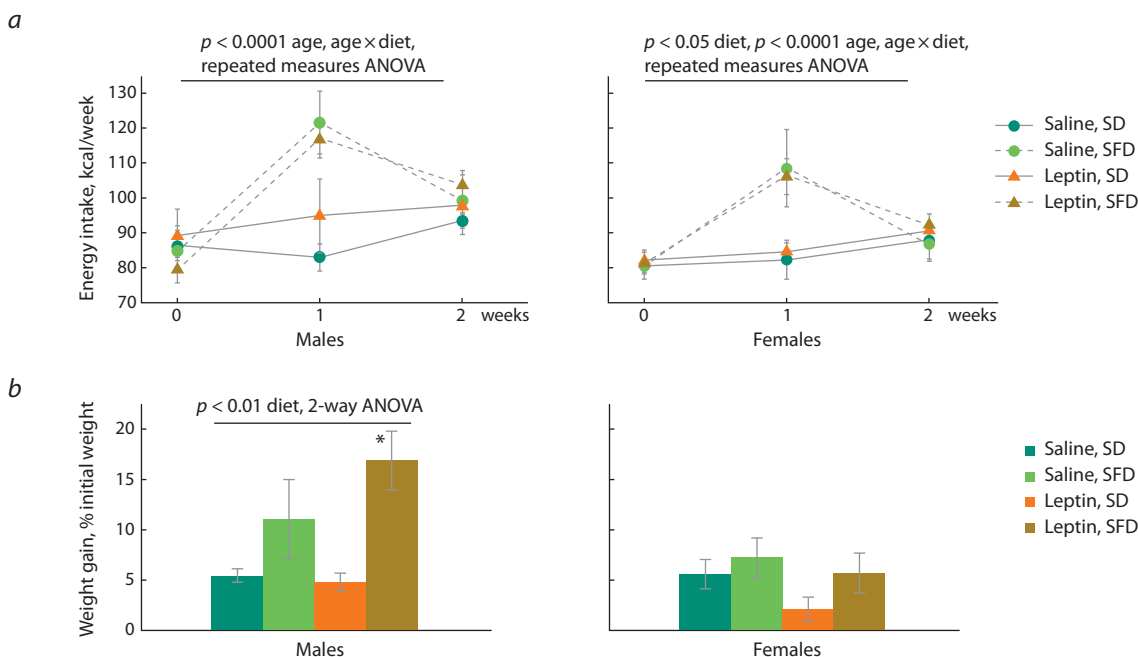
Fig. 2. Influence of leptin administration to pregnant mice on energy intake (*a*) and weight gain (*b*) in male and female offspring consuming standard or sweet and fatty diet.

Data are means ± SE from 6–7 animals in every group. * $p < 0.05$, SFD *vs.* SD, post hoc Newman–Keuls test.

did not affect weight gain in females, and increased weight gain in males, especially in the offspring of leptin-treated mothers (Fig. 2*b*).

Influence of leptin administration to pregnant mice on the metabolic characteristics in offspring of different sexes when kept on SD or SFD

When offspring consumed SD, sex differences were observed in many morphometric and biochemical parameters. Two-way ANOVA with factors "sex" and "maternal treatment" showed that females compared with males had decreased absolute and relative weights of brown adipose tissue (BAT) ($p < 0.001$, absolute, $p < 0.05$, relative, "sex") and intraperitoneal white adipose tissue (WAT) ($p < 0.01$, absolute, $p < 0.05$, relative, "sex") (Table 2), and lowered levels of glucose ($p < 0.05$, "sex"), cholesterol ($p < 0.01$, "sex"), triglycerides ($p < 0.001$, "sex") and leptin ($p < 0.05$, "sex") in the blood (Table 3). Leptin administration to pregnant mothers was associated with an increase in blood triglyceride levels ($p < 0.05$, "maternal treatment"), and this increase reached statistically significant values in male offspring ($p < 0.05$, post hoc Newman–Keuls test).

A two-week intake of SFD reduced the absolute and relative weight of the liver, increased the absolute and relative weight of BAT, as well as visceral and subcutaneous WAT, and increased the blood levels of glucose, cholesterol and leptin in both males and females (Tables 2, 3). Only the change in blood triglyceride levels in response to the consumption of SFD depended on sex: triglyceride levels decreased in males and did not change in females (Table 3). At the same time, in females, the mass of visceral WAT and the concentration of glucose, cholesterol, and leptin in the blood were lower than in males, regardless of the diet consumed (Tables 2, 3). Leptin administration to mothers had a sex-specific effect on

the mass of subcutaneous WAT and blood glucose, cholesterol, and triglyceride levels. When the effect of maternal leptin was analyzed separately in males and females (two-way ANOVA with factors "diet" and "maternal treatment"), it was observed only in males. Regardless of the diet, male offspring of leptin-treated mothers had more subcutaneous fat mass ($p < 0.05$, "maternal treatment") and elevated blood levels of glucose ($p < 0.05$, "maternal treatment"), triglycerides ($p < 0.05$, "maternal treatment") and cholesterol (at the trend level, $p < 0.07$, "maternal treatment") than males born to control mothers.

Influence of leptin administration to pregnant mice on gene expression in the liver, BAT and muscles in male and female offspring consuming SFD or SD

When mice were kept on a standard diet, sex differences were observed in the expression of some of the studied genes in the liver and brown fat. In the liver, the mRNA level of glucose-6-phosphatase (*G6pc*) in females was lower than in males ($p < 0.05$, "sex", two-way ANOVA, SD, Fig. 3*e*). In BAT, the FGF21 mRNA level in females was lower than in males, and the level of insulin receptor mRNA was higher ($p < 0.05$, "sex", for both cases, two-way ANOVA, SD, Fig. 4*a, e*). Leptin administration to mothers reduced the expression of *Fasn* ($p < 0.05$, "maternal treatment", two-way ANOVA, SD, Fig. 3*c*) and *Gck* ($p < 0.05$, "maternal treatment", two-way ANOVA, SD, Fig. 3*g*) in the liver on a standard diet, and this decrease was more pronounced in males, reaching statistically significant values in them (Fig. 3*c, g*).

In the liver, SFD consumption resulted in activation of *Fgf21* gene expression and inhibition of *Pck1* gene expression in both males and females (Fig. 3*a, h*), and inhibition of *Fasn* gene expression only in males ($p < 0.01$, "diet", two-way ANOVA, males, Fig. 3*c*). At the same time, in males,

Е.И. Денисова
Е.Н. Макарова

Влияние материнского лептина
на плоды и взрослое потомство

2024
28•3

Table 2. Influence of leptin administration to pregnant mice on the absolute and relative weight
of the liver, BAT, and visceral and subcutaneous WAT in male and female offspring consuming SD or SFD

Parameter	Males				Females				p, ANOVA
	SD		SFD		SD		SFD		
	Saline	Leptin	Saline	Leptin	Saline	Leptin	Saline	Leptin	
Weight, g									
Liver	1.21±0.03	1.22±0.07	1.05±0.06	1.06±0.03	1.06±0.04	0.89±0.01	0.92±0.05	0.91±0.06	<0.01 sex, <0.01 diet
BAT	0.10±0.01	0.09±0.01	0.14±0.02	0.14±0.02	0.07±0.01	0.06±0.00	0.11±0.02	0.08±0.01	<0.001 sex, <0.001 diet
WAT visceral	0.41±0.08	0.47±0.04	0.86±0.13	1.21±0.22#	0.29±0.10	0.15±0.01	0.61±0.13	0.59±0.14*	<0.01 sex, <0.01 diet
WAT sub-cutaneous	0.63±0.06	0.89±0.13	1.18±0.12	1.65±0.22	0.58±0.09	0.49±0.03	1.14±0.17	0.97±0.22	<0.01 sex, diet <0.05 sex×mat. tr.
Index, %									
Liver	4.87±0.08	4.85±0.22	4.18±0.18	4.01±0.09	5.09±0.13	4.66±0.07	4.45±0.15	4.45±0.17	<0.001 diet
BAT	0.40±0.02	0.36±0.03	0.55±0.06	0.53±0.04	0.32±0.04	0.30±0.02	0.51±0.07	0.41±0.03	<0.001 diet, <0.05 sex
WAT visceral	1.64±0.31	1.87±0.15	3.37±0.39	4.41±0.67	1.39±0.46	0.77±0.05	2.92±0.60	2.75±0.54	<0.01 sex, <0.01 diet
WAT sub-cutaneous	2.51±0.20	3.52±0.47	4.68±0.35	6.08±0.62	2.78±0.39	2.55±0.15	5.48±0.82	4.58±0.86	<0.001 diet <0.05 sex×mat. tr.

Note. Data are means±SE from 6–7 animals in every group. Data were analyzed using three-way ANOVA with factors "sex", "diet", and "maternal treatment" (mat. tr.). * $p < 0.05$ females *vs.* males, # $p < 0.05$ SFD *vs.* SD, post hoc Newman–Keuls test.

Table 3. Influence of leptin administration to pregnant mice on hormonal and metabolic characteristics
in male and female offspring consuming SD or FSD

Parameter	Males				Females				p, ANOVA
	SD		SFD		SD		SFD		
	Saline	Leptin	Saline	Leptin	Saline	Leptin	Saline	Leptin	
Glucose, mM	15.8±1.7	17.4±0.5	16.2±0.9	18.9±0.6	13.4±0.7	14.0±0.5	15.2±0.6	15.2±0.6*	<0.001 sex, <0.05 diet
Cholesterol, mM	1.4±0.1	1.5±0.1	2.3±0.1	2.7±0.2	1.2±0.1	1.2±0.03	2.6±0.4	2.0±0.1*	<0.001, diet, <0.05 sex×mat. tr.
Triglycerides, mM	1.5±1.1	1.9±0.1	1.0±0.1	1.2±0.1#	0.9±0.1*	1.1±0.1	1.3±0.3	1.1±0.1	<0.01 sex, <0.001 sex×diet
Leptin, ng/ml	2.8±1.1	3.5±0.6	9.1±2.4#	13.8±2.3+	1.6±0.7	1.3±0.2	4.9±1.6	7.1±1.6*	<0.01 sex, <0.001 diet
FGF21, ng/ml			5.0±1.6	4.9±1.4			8.2±1.6	3.7±1.6	

Note. Data are means±SE from 6–7 animals in every group. Data were analyzed by three-way ANOVA with factors "sex", "diet" and "maternal treatment" (mat. tr.).
* $p < 0.05$ females *vs.* males, # $p < 0.05$ SFD *vs.* SD; + $p < 0.05$ males, leptin *vs.* saline, post hoc Newman–Keuls test.

leptin administration to mothers changed the response of the *Fasn* gene to SFD consumption: in the offspring of control mothers, *Fasn* gene expression significantly decreased, while in the offspring of leptin-treated mothers, it did not change (Fig. 3*c*). Leptin administration to mothers also had a sex-specific effect on the expression of the glucokinase gene in the liver – it decreased in males regardless of the diet and did not significantly change in females (Fig. 3*g*).

In BAT, SFD consumption increased *Fgf21* and *Cpt1* gene expression (Fig. 4*a*, *c*), decreased *Slc2a4* gene expression (Fig. 4*f*), had a down-regulating effect on *Klb* expression (Fig. 4*b*) in mice of both sexes, and increased *Ucp1* gene expression only in females (Fig. 4*d*). Leptin administration to mothers had no effect on the expression of the studied genes in BAT.

In the muscles, the expression of genes related to insulin sensitivity (*Slc2a4*, *Insr*) and β-oxidation (*Cpt1b*, *Ucp3*) were studied. The expression of these genes did not depend on sex and diet, and leptin administration to mothers had no effect on the expression of these genes.

Fig. 3. Influence of leptin administration to pregnant mice on liver gene expression in male and female offspring consuming SFD or SD.

* $p < 0.05$ SD, males, leptin *vs.* saline; # $p < 0.05$ SFD *vs.* SD, post hoc Newman–Keuls test. Data are means ± SE from 6–7 animals in every group.

Fig. 4. Influence of leptin administration to pregnant mice on gene expression in BAT in male and female offspring consuming SFD or SD.

Data are means ± SE from 6–7 animals in every group.

Е.И. Денисова
Е.Н. Макарова

Влияние материнского лептина
на плоды и взрослое потомство

2024
28 • 3

Fig. 5. Influence of leptin administration to pregnant mice on gene expression in hypothalamus in male and female offspring consuming SFD or SD.
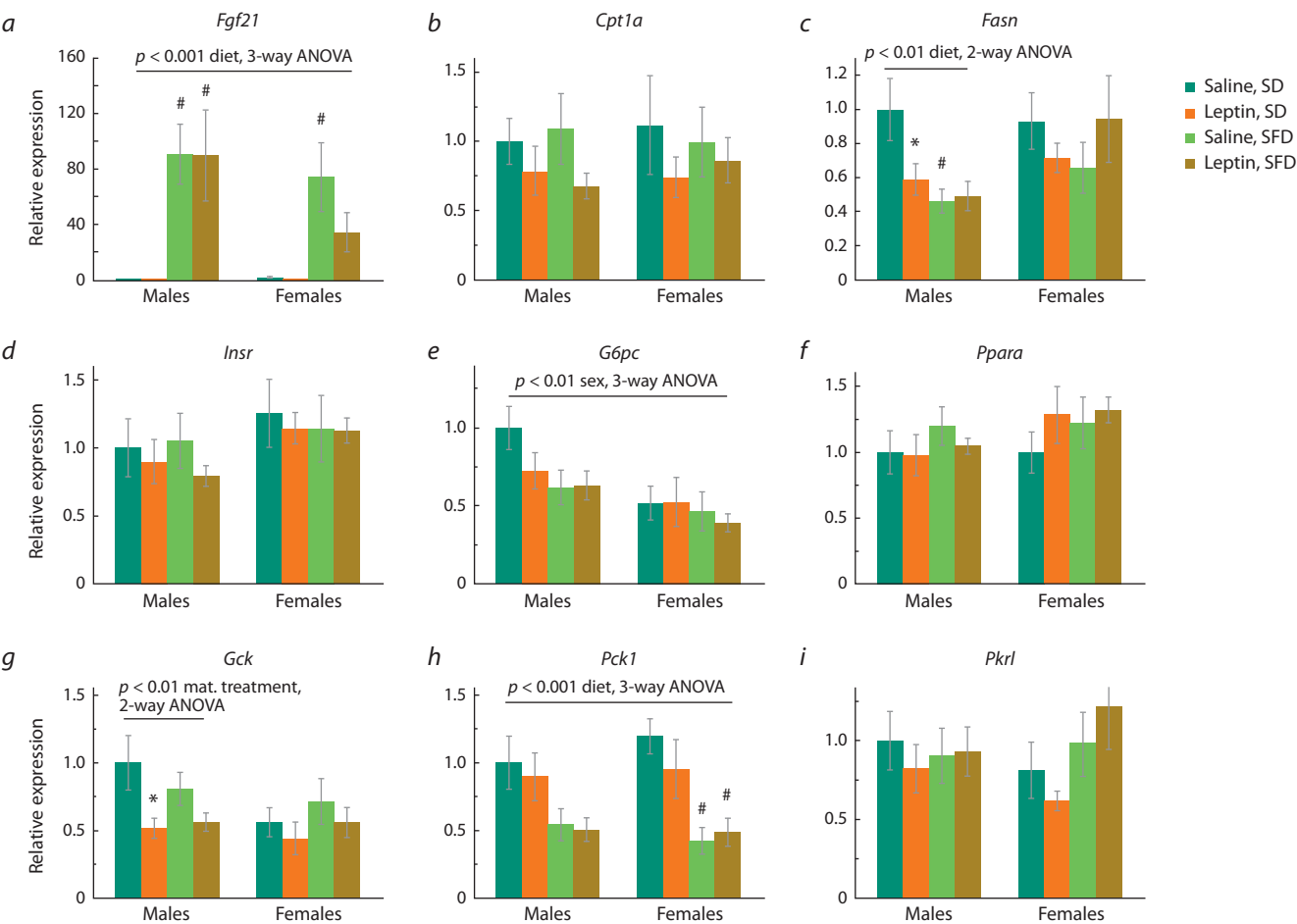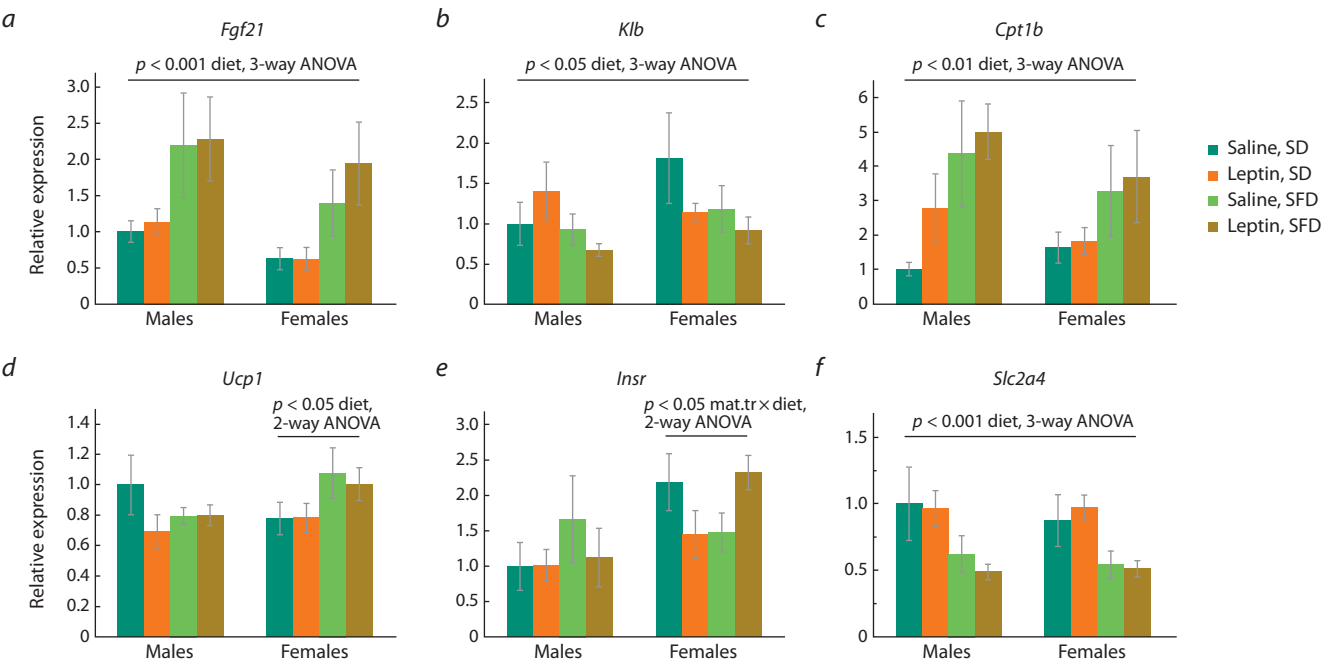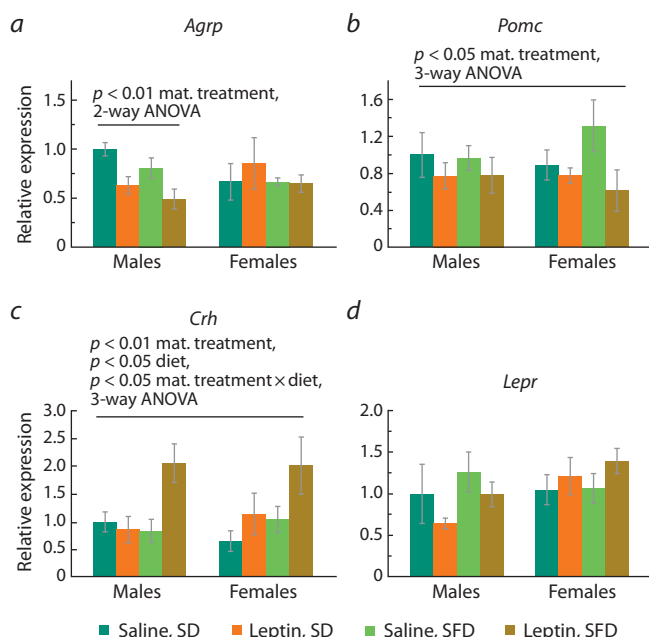
Data are means ± SE from 6–7 animals in every group.

Influence of leptin administration to pregnant mice on hypothalamic gene expression in male and female offspring consuming SFD or SD

When kept on SD, males and females did not differ in the expression of the studied genes in the hypothalamus. Leptin administration to mothers had a down-regulating effect on *Pomc* gene expression regardless of animal sex and diet (Fig. 5*b*), reduced *Agrp* gene expression only in males (Fig. 5*a*) on both SD and SFD, and altered the response of the *Crh* gene to SFD intake. In mice of both sexes born to leptin-treated mothers, the expression of the *Crh* gene increased when SFD was consumed, while in the offspring of control females it did not change (Fig. 5*c*). Expression of *Agrp*, *Pomc*, and *Lepr* did not change in response to SFD consumption.

The results presented suggest that maternal leptin has a programming effect on the metabolic phenotype of the offspring, including influence on the central mechanisms supporting energy homeostasis, and gene expression in the liver and brown fat, and males are more sensitive to the programming action of maternal leptin.

Influence of leptin administration to pregnant mice on the weight of placentas and fetuses in offspring of different sexes

Leptin administration to mothers at mid-pregnancy did not affect fetus viability: control and leptin-treated mothers did not differ in litter size (8.7 ± 0.2, $n = 6$, control mothers, and 9.0 ± 0.2, $n = 6$, leptin-treated mothers). At the end of the embryonic period, male and female fetuses did not differ in weight, and leptin administration to mothers did not have a delayed effect on fetal weight (Fig. 6*b*). Male placentas weighed more than female placentas (Fig. 6*a*). Leptin administration to mothers had no effect on placental or fetal weight.
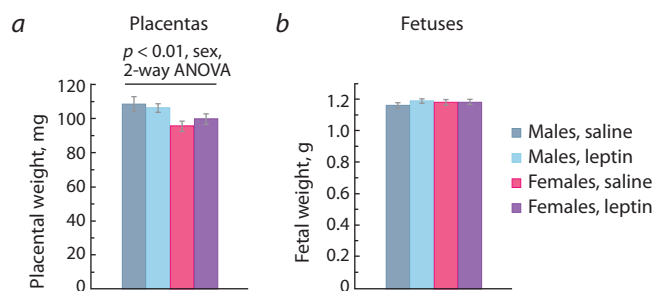
Fig. 6. The effect of leptin administration to female mice at 11–13 days of pregnancy on the weight of placentas (*a*) and fetuses (*b*) of different sex at the end of pregnancy (PD 18).

Data are means ± SE from 32 male and 20 female offspring of control mothers and 29 male and 24 female offspring of leptin-treated mothers.

Influence of leptin administration to pregnant mice on gene expression in placentas, and in the brain and liver of fetuses of different sexes

In the control, female fetus placentas differed from male fetus placentas by increased expression of the *Igf1* gene ($p < 0.05$, Student's *t*-test). Administration of leptin to pregnant mice affected the placental expression of this gene differently in male and female fetuses ($p < 0.05$, "sex" × "maternal treatment", two-way ANOVA): it increased *Igf1* expression in male placentas and decreased in female placentas (Fig. 7*a*). As a result, the sex differences in *Igf1* expression observed in the control group disappeared when leptin was administered to mothers.

The expression of the *Igf2r* gene and, at the level of a trend, the *Slc38a2* (SNAT2) gene ($p = 0.054$, two-way ANOVA) in placentas (Fig. 7*a*) depended on the sex of the fetuses: it was higher in females than in males, and leptin administration to pregnant females had no effect on the expression of these genes.

Sex differences in the expression of the genes studied in the fetal brain and the effect of leptin administration to pregnant females on the expression of these genes were not found (Fig. 7*b*).

Sex differences in the expression of the genes studied in the liver were not found. Leptin administration to pregnant females had an up-regulating effect on the liver expression of the *Igf1* and *Dnmt3b* genes in the fetuses of both sexes and a multidirectional effect (up-regulating in males and down-regulating in females) on the liver expression of the *Dnmt3a* gene (Fig. 7*c*). As a result, *Dnmt3a* gene expression in male fetuses was higher than in female fetuses after leptin administration to mothers.

Thus, administration of leptin to females during pregnancy has a delayed effect on the expression of genes encoding growth factors and DNA methyltransferases in the fetal liver.

Discussion

In the present work, we assessed the effect of maternal leptin on adaptation to high-calorie food in adult offspring, as well as on the signaling function of placentas and fetal liver depending on offspring sex. Sex has a significant effect on obesity-induced metabolic alterations (Hwang et al., 2010), and, in addition, there is sexual dimorphism in the response of offspring to maternal influences not only in the postnatal
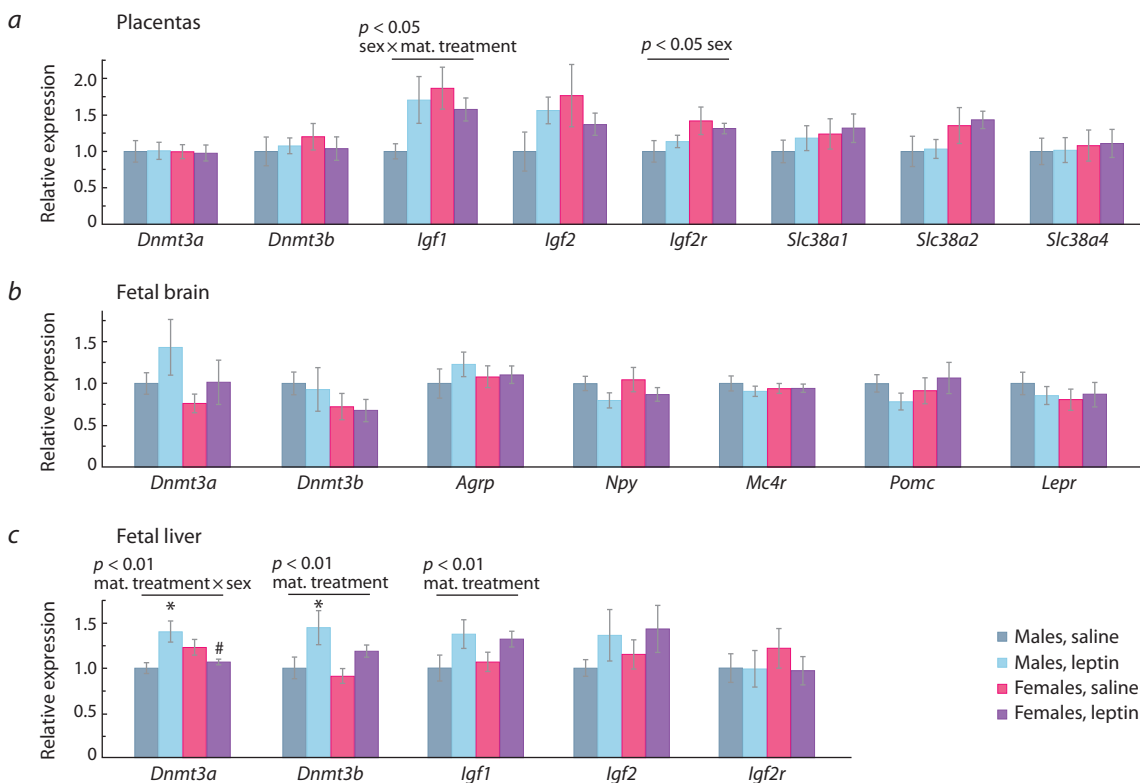
Fig. 7. Influence of leptin administration to female mice at the days 11–13 of pregnancy on gene expression in placentas (*a*), fetal brain (*b*) and fetal liver (*c*) in male and female fetuses at the end of pregnancy (PD 18).

Data are means ± SE from 6 samples in every group. * $p < 0.05$, male fetuses, leptin *vs.* saline; # $p < 0.05$, leptin, females *vs.* males, post hoc Newman–Keuls test.

period of life, but also in fetuses and placentas (Dearden et al., 2018; Yu et al., 2021). It suggests that the programming effect of maternal leptin may be sex-specific.

Male and female offspring differed in metabolic characteristics consuming SD and in response to SFD intake. Compared to males, females had reduced fat mass and reduced blood glucose, cholesterol, and leptin concentrations regardless of the diet consumed, which is consistent with the results of other authors (Freire-Regatillo et al., 2020). SFD consumption was accompanied by an increase in the intake of energy in the offspring of both sexes, but the utilization of this excess energy depended on the sex. In males, when switching to SFD, the mass of white fat increased, the expression of the *Fasn* gene encoding the enzyme for the synthesis of fatty acids decreased in the liver, and the level of triglycerides in the blood decreased. These results are consistent with data obtained in other studies on male mice (Voigt et al., 2013; Casimiro et al., 2021; Kakall et al., 2021) and suggest that in males, excessive consumption of fat at the initial stages of fatty food eating inhibits lipogenesis in the liver and enhances lipid uptake by tissues and lipid storage in adipose tissue. In females, the mass of adipose tissue, liver expression of *Fasn*, and blood triglyceride level did not change in response to SFD but the expression of the *Ucp1* gene in BAT increased, which indicates an increase in thermogenesis and energy dissipation in the form of heat. Thus, males and females demonstrate different adaptive strategies in relation to excess energy intake with SFD.

In other respects, the hormonal and metabolic changes induced by the intake of SFD were similar in males and females and were aimed at reducing food intake, lowering blood glucose levels, and activating fat utilization. In offspring of both sexes, energy intake declined to normal levels in the second week of SFD intake, which may be due to an increase in leptin levels, because leptin reduces food intake (Morton, 2007). In both males and females, the mass of BAT increased and BAT expression of the *Cpt1* gene increased and that of the *Slc2a4* gene (GLUT4) decreased, which points to intensification of lipid utilization. In addition, liver mass decreased and liver *Pck1* gene expression decreased, which indicates the suppression of gluconeogenesis. The expression of the *Fgf21* gene increased in the liver and brown fat. This hormone increases insulin sensitivity, activates fat oxidation, and influences food choice, increasing the propensity to consume a balanced diet (Flippo, Potthoff, 2021). These results are consistent with data obtained by other authors. It has been shown in mice and rats that the initial stages of adaptation to the consumption of a high-calorie diet are characterized by an increase in energy expenditure, an increase in the level of leptin in the blood, an increase in the mass of brown fat, UCP1 protein expression and fatty acid oxidation in brown fat, an increase in fat utilization, a decrease in liver weight, a decrease in the expression of the *Slc2a4* gene (GLUT 4) in adipocytes (So et al., 2011; Andrich et al., 2018; Kakall et al., 2021).

Leptin administration to pregnant females had a delayed effect on both the metabolic phenotype of the offspring in the

Е.И. Денисова
Е.Н. Макарова

Влияние материнского лептина
на плоды и взрослое потомство

2024
28•3

postnatal period, and on fetuses and placentas. Leptin administration to mothers reduced offspring growth rate in the first weeks after weaning. These results are consistent with the results obtained previously, demonstrating that hyperleptinemia during pregnancy reduces the weight of the offspring during their growth after weaning (Makarova et al., 2013; Pollock et al., 2015). In this work, we have shown for the first time that leptin administration to pregnant females has an up-regulating effect on the level of IGF1 mRNA in the liver of fetuses at the end of pregnancy. IGF1 has multisystem effects on fetal development (Hellström et al., 2016), and it is possible that the programming effect of maternal leptin on postnatal metabolic traits and offspring growth is partly mediated by its influence on *Igf1* expression in fetuses.

The programming effect of maternal leptin was more pronounced in male offspring: only in males, administration of leptin to mothers increased fat mass, plasma concentrations of glucose, cholesterol, and triglycerides and decreased the expression of the *Agrp* gene in the hypothalamus and the genes for glucokinase and fatty acid synthase in the liver. Sex differences in the response to elevated maternal leptin were also observed at the prenatal stage of development: only in male fetuses, administration of leptin to mothers increased the expression of the *Dnmt3a* gene in the liver. DNMT3a mediates *de novo* methylation (Jurkowska et al., 2011) and maternal influence on fetal liver expression of this enzyme may have delayed effects on mature liver gene expression. In turn, changes in the expression of genes encoding enzymes in the liver can affect the metabolic parameters of the blood. Thus, a decrease in the expression of the glucokinase gene may be the cause of an increased blood level of glucose in males born to leptin-treated mothers, since glucokinase is a major contributor to glucose homeostasis (Massa et al., 2011), and a decrease in the expression of the *Gck* gene is accompanied by an increase in the level of glucose in the blood (Magnuson et al., 2003).

Despite the pronounced sex differences in metabolic characteristics and the sex-specific effect of maternal leptin on the metabolic phenotype of the offspring, the programming effect of maternal leptin on adaptation to SFD consumption did not depend on the offspring sex. Leptin administration to mothers did not pronouncedly affect the metabolic response and transcriptional changes in the liver and brown fat caused by SFD consumption, but affected the central mechanisms regulating energy intake and expenditure. In both sexes, administration of leptin to mothers doubled the expression of the *Crh* gene in the hypothalamus when SFD was consumed. Hypothalamic corticotrophin-releasing hormone (CRH) coordinates energy intake and expenditure with metabolic and behavioral response to stress (Richard et al., 2000). CRH in the hypothalamus has an anorexigenic effect and increases energy expenditure (Radahmadi et al., 2021). Decreased sensitivity of CRH neurons increases susceptibility to obesity in mice (Zhu et al., 2020). Since the increase in *Crh* gene expression was not accompanied by changes in food intake and body weight, it can be assumed that maternal leptin affected the response of hypothalamic–pituitary–adrenal axis to metabolic stress caused by SFD consumption. The nature of these influences requires additional research.

In addition, leptin administration to mothers affected the hypothalamic expression of orexigenic (*Agrp*) neuropeptide

in males and anorexigenic (*Pomc*) neuropeptide in males and females. It is assumed that prenatal programming of the metabolic phenotype is mediated via epigenetic modifications of the central systems that regulate energy intake and expenditure (Dearden, Ozanne, 2015). Thus, it has been shown in laboratory models and humans that the metabolic state of mothers during pregnancy (malnutrition, overeating) affects methylation of the gene encoding proopiomelanocortin and, accordingly, its expression in the hypothalamus in the offspring (Candler et al., 2019). In rats, maternal consumption of high-calorie diet significantly increased basal CRH mRNA expression in the paraventricular nucleus of hypothalamus (Niu et al., 2019). Our results indicate that leptin may be the factor mediating maternal influences on the central regulation of energy homeostasis.

Although we found no sex-dependent programming effects of maternal leptin on adaptation to SFD eating, its sex-specific influence on liver gene expression and metabolic characteristics may promote formation of sex differences in the development of diet-induced obesity in offspring.

Conclusion

Males differ from females in metabolic features associated with glucose and lipid metabolism, as well as adaptation to excess energy intake with a high-calorie diet. Leptin administration to pregnant female mice sex-specifically affects liver gene expression and metabolic characteristics in adult offspring. This sex-specific programming effect may be associated with sex-specific influence of maternal leptin on expression of the *Dnmt3a* gene in fetal liver. Regardless of sex, maternal leptin had a programming effect on the activity of the hypothalamic CRH system during adaptation to SFD consumption.

References

Akieda-Asai S., Koda S., Sugiyama M., Hasegawa K., Furuya M., Miyazato M., Date Y. Metabolic features of rats resistant to a high-fat diet. *Obes. Res. Clin. Pract.* 2013;7(4):e243-e250. DOI 10.1016/j.orcp.2013.01.004

Andrich D.E., Melbouci L., Ou Y., Leduc-Gaudet J.-P., Chabot F., Lalonde F., Lira F.S., Gaylinn B.D., Gouspillou G., Danialou G., Comtois A.-S., St-Pierre D.H. Altered feeding behaviors and adiposity precede observable weight gain in young rats submitted to a short-term high-fat diet. *J. Nutr. Metab.* 2018;2018:1498150. DOI 10.1155/2018/1498150

Candler T., Kühnen P., Prentice A.M., Silver M. Epigenetic regulation of *POMC*; implications for nutritional programming, obesity and metabolic disease. *Front. Neuroendocrinol.* 2019;54:100773. DOI 10.1016/j.yfrne.2019.100773

Casimiro I., Stull N.D., Tersey S.A., Mirmira R.G. Phenotypic sexual dimorphism in response to dietary fat manipulation in C57BL/6J mice. *J. Diabetes Complications.* 2021;35(2):107795. DOI 10.1016/j.jdiacomp.2020.107795

Cone R.D. Anatomy and regulation of the central melanocortin system. *Nat. Neurosci.* 2005;8(5):571-578. DOI 10.1038/nn1455

Dearden L., Ozanne S.E. Early life origins of metabolic disease: developmental programming of hypothalamic pathways controlling energy homeostasis. *Front. Neuroendocrinol.* 2015;39:3-16. DOI 10.1016/j.yfrne.2015.08.001

Dearden L., Bouret S.G., Ozanne S.E. Sex and gender differences in developmental programming of metabolism. *Mol. Metab.* 2018;15: 8-19. DOI 10.1016/j.molmet.2018.04.007

Denisova E.I., Savinkova M.M., Makarova E.N. Influence of leptin administration to pregnant female mice on obesity development, taste preferences, and gene expression in the liver and muscles of their

male and female offspring. *Vavilov J. Genet. Breed.* 2021;25(6): 669-676. DOI 10.18699/VJ21.076

Duca F.A., Sakar Y., Lepage P., Devime F., Langelier B., Doré J., Covasa M. Replication of obesity and associated signaling pathways through transfer of microbiota from obese-prone rats. *Diabetes.* 2014;63(5):1624-1636. DOI 10.2337/db13-1526

Even P.C. Identification of behavioral and metabolic factors predicting adiposity sensitivity to both high fat and high carbohydrate diets in rats. *Front. Physiol.* 2011;2:96. DOI 10.3389/fphys.2011.00096

Flippo K.H., Potthoff M.J. Metabolic messengers: FGF21. *Nat. Metab.* 2021;3(3):309-317. DOI 10.1038/s42255-021-00354-2

Freire-Regatillo A., Fernández-Gómez M.J., Díaz F., Barrios V., Sánchez-Jabonero I., Frago L.M., Argente J., García-Segura L.M., Chowen J.A. Sex differences in the peripubertal response to a short-term, high-fat diet intake. *J. Neuroendocrinol.* 2020;32(1):e12756. DOI 10.1111/jne.12756

Hacker A., Capel B., Goodfellow P., Lovell-Badge R. Expression of *Sry*, the mouse sex determining gene. *Development.* 1995;121(6): 1603-1614. DOI 10.1242/dev.121.6.1603

Hattori Y., Takeda T., Fujii M., Taura J., Yamada H., Ishii Y. Attenuation of growth hormone production at the fetal stage is critical for dioxin-induced developmental disorder in rat offspring. *Biochem. Pharmacol.* 2021;186:114495. DOI 10.1016/j.bcp.2021.114495

Hellström A., Ley D., Hansen-Pupp I., Hallberg B., Löfqvist C., Marter L., Weissenbruch M., Ramenghi L.A., Beardsall K., Dunger D., Hård A., Smith L.E.H. Insulin-like growth factor 1 has multisystem effects on foetal and preterm infant development. *Acta Paediatr.* 2016;105(6):576-586. DOI 10.1111/apa.13350

Hwang L.-L., Wang C.-H., Li T.-L., Chang S.-D., Lin L.-C., Chen C.-P., Chen C.-T., Liang K.-C., Ho I.-K., Yang W.-S., Chiou L.-C. Sex differences in high-fat diet-induced obesity, metabolic alterations and learning, and synaptic plasticity deficits in mice. *Obesity.* 2010; 18(3):463-469. DOI 10.1038/oby.2009.273

Ishii Y., Bouret S.G. Embryonic birthdate of hypothalamic leptin-activated neurons in mice. *Endocrinology.* 2012;153(8):3657-3667. DOI 10.1210/en.2012-1328

Jurkowska R.Z., Jurkowski T.P., Jeltsch A. Structure and function of mammalian DNA methyltransferases. *Chembiochem.* 2011;12(2): 206-222. DOI 10.1002/cbic.201000195

Kakall Z.M., Gopalasingam G., Herzog H., Zhang L. Dynamic regional alterations in mouse brain neuronal activity following short-term changes in energy balance. *Obesity.* 2021;29(10):1650-1663. DOI 10.1002/oby.23253

Magnuson M.A., She P., Shiota M. Gene-altered mice and metabolic flux control. *J. Biol. Chem.* 2003;278(35):32485-32488. DOI 10.1074/jbc.R300020200

Makarova E.N., Chepeleva E.V., Panchenko P.E., Bazhan N.M. Influence of abnormally high leptin levels during pregnancy on metabolic phenotypes in progeny mice. *Am. J. Physiol. Integr. Comp. Physiol.* 2013;305(11):R1268-R1280. DOI 10.1152/ajpregu.00162. 2013

Massa M.L., Gagliardino J.J., Francini F. Liver glucokinase: an overview on the regulatory mechanisms of its activity. *IUBMB Life.* 2011;63(1):1-6. DOI 10.1002/iub.411

Morton G.J. Hypothalamic leptin regulation of energy homeostasis and glucose metabolism. *J. Physiol.* 2007;583(2):437-443. DOI 10.1113/jphysiol.2007.135590

Nawathe A.R., Christian M., Kim S.H., Johnson M., Savvidou M.D., Terzidou V. Insulin-like growth factor axis in pregnancies affected by fetal growth disorders. *Clin. Epigenetics.* 2016;8(1):11. DOI 10.1186/s13148-016-0178-5

Nilsson C., Swolin-Eide D., Ohlsson C., Eriksson E., Ho H., Bjorntorp P., Holmang A. Reductions in adipose tissue and skeletal growth

in rat adult offspring after prenatal leptin exposure. *J. Endocrinol.* 2003;176(1):13-21. DOI 10.1677/joe.0.1760013

Niu X., Wu X., Ying A., Shao B., Li X., Zhang W., Lin C., Lin Y. Maternal high fat diet programs hypothalamic-pituitary-adrenal function in adult rat offspring. *Psychoneuroendocrinology.* 2019;102:128-138. DOI 10.1016/j.psyneuen.2018.12.003

Petry C.J., Evans M.L., Wingate D.L., Ong K.K., Reik W., Constância M., Dunger D.B. Raised late pregnancy glucose concentrations in mice carrying pups with targeted disruption of $H19^{\Delta 13}$. *Diabetes.* 2010;59(1):282-286. DOI 10.2337/db09-0757

Pollock K.E., Stevens D., Pennington K.A., Thaisrivongs R., Kaiser J., Ellersieck M.R., Miller D.K., Schulz L.C. Hyperleptinemia during pregnancy decreases adult weight of offspring and is associated with increased offspring locomotor activity in mice. *Endocrinology.* 2015;156(10):3777-3790. DOI 10.1210/en.2015-1247

Radahmadi M., Izadi M.S., Rayatpour A., Ghasemi M. Comparative study of CRH microinjections into PVN and CeA nuclei on food intake, ghrelin, leptin, and glucose levels in acute stressed rats. *Basic Clin. Neurosci. J.* 2021;12(1):133-148. DOI 10.32598/bcn.12.1. 2346.1

Reynolds C.M., Segovia S.A., Vickers M.H. Experimental models of maternal obesity and neuroendocrine programming of metabolic disorders in offspring. *Front. Endocrinol.* 2017;8:245. DOI 10.3389/fendo.2017.00245

Richard D., Huang Q., Timofeeva E. The corticotropin-releasing hormone system in the regulation of energy balance in obesity. *Int. J. Obes. Relat. Metab. Disord.* 2000;24(S2):S36-S39. DOI 10.1038/sj.ijo.0801275

Schoonejans J.M., Blackmore H.L., Ashmore T.J., Aiken C.E., Fernandez-Twinn D.S., Ozanne S.E. Maternal metformin intervention during obese glucose-intolerant pregnancy affects adiposity in young adult mouse offspring in a sex-specific manner. *Int. J. Mol. Sci.* 2021;22(15):8104. DOI 10.3390/ijms22158104

Shrestha D., Ouidir M., Workalemahu T., Zeng X., Tekola-Ayele F. Placental DNA methylation changes associated with maternal prepregnancy BMI and gestational weight gain. *Int. J. Obes.* 2020;44(6): 1406-1416. DOI 10.1038/s41366-020-0546-2

So M., Gaidhu M.P., Maghdoori B., Ceddia R.B. Analysis of time-dependent adaptations in whole-body energy balance in obesity induced by high-fat diet in rats. *Lipids Health Dis.* 2011;10(1):99. DOI 10.1186/1476-511X-10-99

Stocker C.J., Cawthorne M.A. The influence of leptin on early life programming of obesity. *Trends Biotechnol.* 2008;26(10):545-551. DOI 10.1016/j.tibtech.2008.06.004

Talton O.O., Pennington K.A., Pollock K.E., Bates K., Ma L., Ellersieck M.R., Schulz L.C. Maternal hyperleptinemia improves offspring insulin sensitivity in mice. *Endocrinology.* 2016;157(7): 2636-2648. DOI 10.1210/en.2016-1039

Voigt A., Agnew K., van Schothorst E.M., Keijer J., Klaus S. Short-term, high fat feeding-induced changes in white adipose tissue gene expression are highly predictive for long-term changes. *Mol. Nutr. Food Res.* 2013;57(8):1423-1434. DOI 10.1002/mnfr.201200671

Xiagedeer B., Kang C., Hou X., Hu H., Xiao Q., Hao W. Chlormequat chloride promotes rat embryonic growth and GH-IGF-1 axis. *Toxicology.* 2020;429:152326. DOI 10.1016/j.tox.2019.152326

Yu P., Chen Y., Ge C., Wang H. Sexual dimorphism in placental development and its contribution to health and diseases. *Crit. Rev. Toxicol.* 2021;51(6):555-570. DOI 10.1080/10408444.2021.1977237

Zhu C., Xu Y., Jiang Z., Tian J.B., Cassidy R.M., Cai Z.L., Shu G., Xu Y., Xue M., Arenkiel B.R., Jiang Q., Tong Q. Disrupted hypothalamic CRH neuron responsiveness contributes to diet-induced obesity. *EMBO Rep.* 2020;21(7):e49210. DOI 10.15252/embr.2019 49210

Urine metabolic profile in rats with arterial hypertension of different genesis

A.A. Sorokoumova [1] ✉, A.A. Seryapina [1] ✉, Yu.K. Polityko [1, 3], L.V. Yanshole [2], Yu.P. Tsentalovich [2], M.A. Gilinsky [3], A.L. Markel [1, 4]

[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] Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia
[4] Novosibirsk State University, Novosibirsk, Russia

✉ evanesalisa@yandex.ru; seryapina@bionet.nsc.ru

Abstract. The diversity of pathogenetic mechanisms underlying arterial hypertension leads to the necessity to devise a personalized approach to the diagnosis and treatment of the disease. Metabolomics is one of the promising methods for personalized medicine, as it provides a comprehensive understanding of the physiological processes occurring in the body. The metabolome is a set of low-molecular substances available for detection in a sample and representing intermediate and final products of cell metabolism. Changes in the content and ratio of metabolites in the sample mark the corresponding pathogenetic mechanisms by highlighting them, which is especially important for such a multifactorial disease as arterial hypertension. To identify metabolomic markers for hypertensive conditions of different origins, three forms of arterial hypertension (AH) were studied: rats with hereditary AH (ISIAH rat strain); rats with AH induced by L-NAME administration (a model of endothelial dysfunction with impaired NO production); rats with AH caused by the administration of deoxycorticosterone in combination with salt loading (hormone-dependent form – DOCA-salt AH). WAG rats were used as normotensive controls. 24-hour urine samples were collected from all animals and analyzed by quantitative NMR spectroscopy for metabolic profiling. Then, potential metabolomic markers for the studied forms of hypertensive conditions were identified using multivariate statistics. Analysis of the data obtained showed that hereditary stress-induced arterial hypertension in ISIAH rats was characterized by a decrease in the following urine metabolites: nicotinamide and 1-methylnicotinamide (markers of inflammatory processes), N-acetylglutamate (nitric oxide cycle), isobutyrate and methyl acetoacetate (gut microbiota). Pharmacologically induced forms of hypertension (the L-NAME and DOCA+NaCl groups) do not share metabolomic markers with hereditary AH. They are differentiated by N,N-dimethylglycine (both groups), choline (the L-NAME group) and 1-methylnicotinamide (the group of rats with DOCA-salt hypertension).

Key words: arterial hypertension; ISIAH rats; L-NAME; DOCA-salt hypertension; urine metabolomic markers.

Метаболомный профиль мочи крыс с артериальной гипертонией разного генеза

А.А. Сорокоумова [1] ✉, А.А. Серяпина [1] ✉, Ю.К. Политыко [1, 3], Л.В. Яньшоле [2], Ю.П. Centaлович [2], М.А. Гилинский [3], А.Л. Маркель [1, 4]

[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] Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия
[4] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ evanesalisa@yandex.ru; seryapina@bionet.nsc.ru

Аннотация. Многообразие патогенетических механизмов, лежащих в основе артериальной гипертонии, приводит к необходимости разработки персонализированного подхода к диагностике и терапии заболевания. Одним из перспективных методов для персонализированной медицины является метаболомика, которая позволяет получить комплексное представление о физиологических процессах, происходящих в организме. Метаболом – это совокупность низкомолекулярных веществ, определяемых в образце и являющихся промежуточными и конеч-

ными продуктами метаболизма клеток. Изменения в содержании и соотношении метаболитов в исследуемом образце маркируют соответствующие патогенетические механизмы, выделяя их, что особенно важно для такого мультифакторного заболевания, как артериальная гипертония. Для идентификации метаболомных маркеров гипертензивных состояний разного генеза были исследованы три разные формы артериальной гипертонии (АГ): крысы с наследственной АГ (линия крыс НИСАГ/ISIAH); крысы с АГ, индуцированной введением L-NAME (модель эндотелиальной дисфункции с нарушением продукции NO); крысы с АГ, вызванной введением дезоксикортикостерона в сочетании с солевой нагрузкой (гормон-зависимая форма – DOCA-солевая АГ). В качестве нормотензивного контроля были использованы крысы линии WAG. У всех животных собрали образцы суточной мочи, метаболомный профиль которой проанализировали методом количественной ЯМР-спектроскопии. Затем с помощью методов многомерной статистики выявили потенциальные метаболомные маркеры исследуемых форм гипертензивных состояний. Анализ полученных данных показал, что для наследственной стресс-индуцированной артериальной гипертонии у крыс линии НИСАГ характерно снижение содержания следующих метаболитов в моче: никотинамида и 1-метилникотинамида (маркеры воспалительных процессов), N-ацетилглутамата (цикл оксида азота), изобутирата и метилацетоацетата (микробиота кишечника). Фармакологически индуцированные формы АГ (группы L-NAME и DOCA+NaCl) не имеют общих с наследственной АГ метаболомных маркеров. Их отличают один общий маркер, N,N-диметилглицин, и два специфических – холин (для группы L-NAME) и 1-метилникотинамид (для группы крыс с DOCA-солевой артериальной гипертонией).

Ключевые слова: артериальная гипертония; крысы НИСАГ (ISIAH); L-NAME; DOCA-солевая гипертония; метаболомные маркеры мочи.

Introduction

Arterial hypertension (AH) is a complex multifactorial disease, simultaneously affecting various systems of the body. Pathogenetic diversity of AH and interaction between different underlying mechanisms determine the necessity to consider a lot of factors when developing methods of prevention and treatment. At the moment, multi-stage protocols for the treatment of hypertension have been devised and applied, taking into account lifestyle, stage of the disease, concomitant pathologies, etc. (Carey et al., 2022). However, the tasks of personalized medicine are still relevant, including use of integrated approaches in diagnostics for identifying the distinctive "set" of mechanisms involved in the development of hypertension for a particular patient, and, accordingly, prescribing individual treatment. The so-called "omics" technologies are well suited for this purpose, as they provide a kind of "snapshot" of the organism and its systems for further analysis (at the level of genome, transcriptome, proteome, metabolome, etc.).

To study the metabolic pathways involved in the pathogenesis of various hypertensive conditions, we used three experimental models of hypertension. The first model was the ISIAH rat strain (Inherited Stress-Induced Arterial Hypertension), obtained from an outbred population of Wistar rats through long-term selection for increased blood pressure (BP) under psycho-emotional stress (Markel, 1992). This model reproduces primary (essential) human hypertension quite accurately. The second model – AH caused by dysfunction of the vascular endothelium – was induced pharmacologically by the administration of L-NAME (an inhibitor of NO synthesis) (Biancardi et al., 2007). Endothelial dysfunction associated with impaired nitric oxide synthesis is also one of the common mechanisms of BP rising. The third model of AH was also induced pharmacologically by the administration of a synthetic mineralocorticoid – DOCA (deoxycorticosterone acetate) along with additional salt loading (Basting, Lazartigues, 2017). The combination of elevated mineralocorticoid levels and salt loading is another possible cause of hypertension in humans (Gupta, 2011).

When developing methods to identify biochemical markers for different forms of hypertension, attention should be paid to the availability and non-invasiveness of the proposed technologies. One of the most accessible methods is the analysis of urine samples. Metabolomic studies of urine are currently of interest to researchers, and methods for analyzing and interpreting such data are actively discussed (Zhang et al., 2012; Bouatra et al., 2013).

The purpose of our study is to evaluate the metabolomic profile of 24-hour urine in rats representing three different forms of hypertensive conditions, in comparison with normotensive controls.

Materials and methods

Experimental animals. We studied 3–4-month-old male rats of the ISIAH strain with a hereditary form of hypertension, together with two groups of rats with pharmacologically induced forms of AH: a group of rats treated with a NO synthesis blocker – L-NAME, and a group of rats with hormone-dependent DOCA+NaCl hypertension. WAG rats were used as a normotensive control.

To model NO-deficient hypertension with endothelial dysfunction, WAG rats were orally administered with a solution of endothelial NO synthase inhibitor (L-NAME, N_ω-nitro-l-arginine methyl ester; Sigma Aldrich, USA) at a dose of 30 mg/kg of body weight for two weeks (Fürstenau et al., 2008). To obtain hormone-dependent DOCA-salt hypertension, WAG rats were subcutaneously injected with DOCA (deoxycorticosterone acetate; Sigma Aldrich, USA) at a dose of 25 mg/kg of body weight once every 4 days with a constant salt loading – 1 % NaCl solution in drinking water – for three weeks (Chan et al., 2006). As a result, four experimental groups of animals were formed, three with hypertension and one normotensive, 10 males in each.

All animals were kept under standard conditions at the vivarium of the Institute of Cytology and Genetics SB RAS (air temperature 22–24 °C, light:dark cycle 12:12 hours), receiving standard chow (Chara, Russia) and free access to drinking water. All procedures with experimental 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-Institutional Commission

А.А. Сорокоумова, А.А. Серяпина, Ю.К. Политыко
Л.В. Яньшоле, Ю.П. Централович, М.А. Гилинский, А.Л. Маркель

Метаболомный профиль мочи крыс
с артериальной гипертонией разного генеза

2024
28 · 3

on Biological Ethics at the Institute of Cytology and Genetics SB RAS (protocol No. 127 of 09/08/2022).

Blood pressure (BP) monitoring was performed twice a week throughout the experiment on a device for non-invasive BP measurement (BIOPAC, USA) using the tail-cuff method with preliminary adaptation of the animals to this procedure for 3–4 days. Also, simultaneously with BP measurements, the rats were weighed regularly.

Collection of 24-hour urine samples. Animals were placed in individual rodent metabolic cages (Techniplast, Italy), where they adapted to new conditions for 3 days. Over the next 3 days, each day at the same time, urine samples were collected and the volume of water drunk was recorded. The collected urine was stored at –70 °C. Further analysis of the samples obtained was carried out at the Center for Shared Use "Mass Spectrometric Research" of the International Tomography Center SB RAS, in the Laboratory of Proteomics and Metabolomics.

Extraction of metabolites from urine samples. To obtain a non-protein extract of rat urine metabolites, the following sample preparation protocol was used: the optimal ratio of urine volumes to extracting solution was urine/methanol ratio = 1/4. 400 µl of cold methanol (–20 °C) was added to 100 µl of urine. The samples were mixed in a vortex centrifuge and placed on a shaker for 15 minutes at 1,300 rpm, then centrifuged at 12,000 rpm at 4 °C for 30 minutes, followed by the collection of supernatant. The supernatant was dried on a vacuum evaporator and stored at –70 °C. Lyophilized extracts were diluted in 600 µl of deuterated phosphate buffer (50 mM, pH 7.4) supplemented with internal standard DSS (sodium 3-(trimethylsilyl)propane-1-sulfonate, 20 µM).

NMR spectra were recorded on an AVANCE III HD 700 MHz NMR spectrometer (Bruker BioSpin, Germany), equipped with an Ascend cryomagnet with a field of 16.44 Tesla and a TXI 1H-13C/15N/D ZGR 5 mm probe. Detection parameters corresponded to those described previously (Zelentsova et al., 2020). MestReNova v 12.0 program was used to process the spectra and integrate the signals. Metabolites were identified using the Human Metabolome Database (https://hmdb.ca/) and our own data on the metabolic profiling of hu-

man and animal biological fluids (Tsentalovich et al., 2020; Fomenko et al., 2022).

Statistical data processing was performed using Statistica 12 software package (StatSoft, Inc., 2014) and Metabo Analyst 5.0 web platform (https://www.metaboanalyst.ca/); multivariate analysis (principal component method) and non-parametric methods (Mann–Whitney U test with Bonferroni correction for multiple comparisons) were applied.

Results

A comparative analysis of physiological parameters of the studied animals showed that experimental groups of rats did not have significant differences in body weight: WAG – 326.1 ± 12.87 g (BP – 135.9 ± 1.21 mmHg), ISIAH – 325.9 ± 6.44 g (BP – 205.9 ± 2.12 mmHg), L-NAME – 326.9 ± 4.71 g (BP – 192.0 ± 2.96 mmHg), DOCA – 328.2 ± 6.18 g (BP – 184.2 ± 1.19 mmHg). However, daily water intake, daily diuresis and glomerular filtration rate were significantly increased in rats of the DOCA group compared to the control (Fig. 1). Increased diuresis and water intake were also observed in ISIAH rats. Based on these intergroup differences, for metabolomic analysis of daily urine, the concentrations of metabolites obtained by NMR spectroscopy (nmol/ml) were recalculated into daily urine excretion of metabolites (nmol/day), considering the diuresis level in each rat on the day of sample collection.

Multivariate analysis of metabolomic data using the principal component method revealed three main axes (PC1, PC2, PC3), respectively responsible for 45.2, 16.5 and 10.0 % of the total variation in the content of the studied metabolites in 24-hour urine samples. The distribution of experimental groups in the coordinates of the principal components is shown in Fig. 2. A distinct separation of the experimental groups along the axis of the first principal component (responsible for 45.2 % of the variation in the studied parameters) is observed: normotensive WAG rats and ISIAH rats with hereditary hypertension are virtually combined into one group. Rats with pharmacological forms of hypertension – L-NAME and DOCA – form another separate group together. From this we may conclude that "natural" genetic hypertension is in sharp
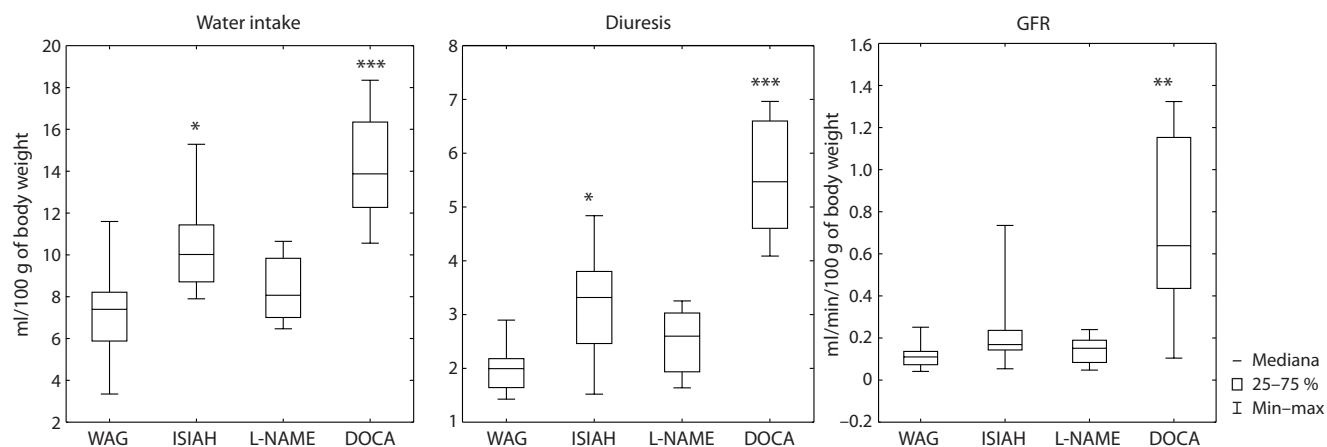

Fig. 1. Intergroup differences in daily water intake, daily diuresis and glomerular filtration rate (GFR) in ISIAH, L-NAME, DOCA rats relative to control WAG rats.

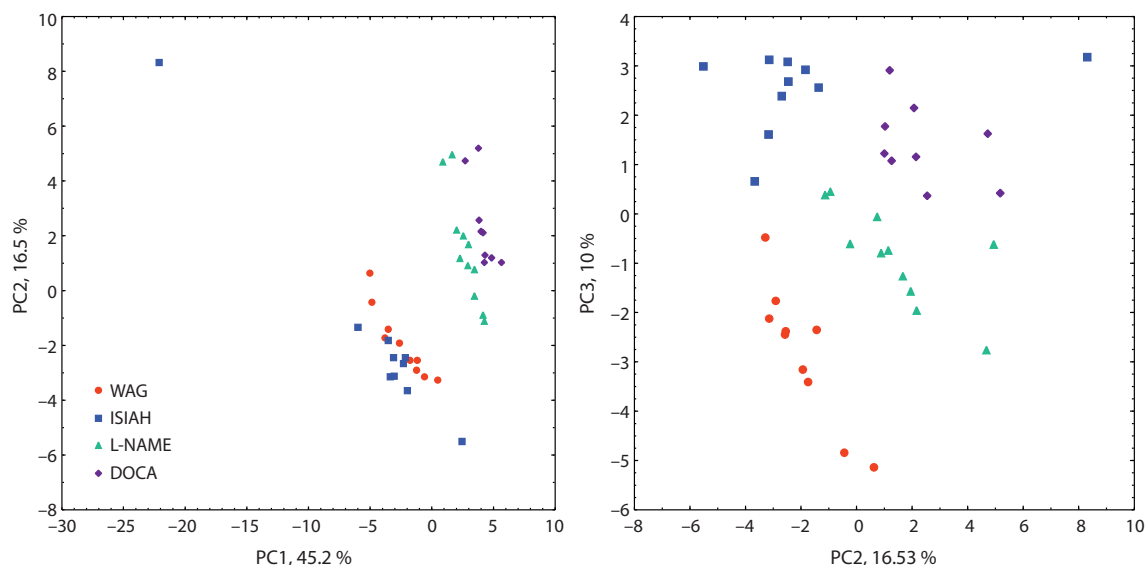Mann–Whitney test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Fig. 2. Distribution of groups of normotensive (WAG) and hypertensive (ISIAH, L-NAME, DOCA) rats in the coordinates of the principal component axes (PC1, PC2, PC3).

contrast to the two pharmacologically induced forms of AH, so urine metabolomic markers correspond not so much to elevated BP levels as to the pathogenesis of different forms of AH.

A relatively small percentage of variability (16.5 %) in metabolomic parameters is described by the second principal component. The projection of the experimental groups onto the second component does not make it possible to separate the compared groups in accordance to any set of metabolomic markers. At the same time, the residual variability of parameters (10 %) for the third principal component, when considered, showed the possibility of a distinct separation of normotensive rats (WAG) and rats with hereditary arterial hypertension (ISIAH). Thus, metabolomic markers that correlate with the third principal component can serve as diagnostic indicators of hereditary stress-dependent forms of hypertension.

Fig. 3 shows the loadings of metabolites along the axes of the first and third principal components. For the first principal component, most of the parameters correlate with the combined group – normotensive WAG rats + hypertensive ISIAH rats, – while choline, N,N-dimethylglycine, N6-acetyllysine, 1-methylnicotinamide and formate correlate with pharmacologically induced forms of AH. Thus, the metabolic profile of hereditary AH (at least in the early stages of its development) is closer to normal than to those of pharmacologically induced models of AH. Markers that distinguish ISIAH rats from normotensive controls (WAG rats) may be divided into those that correlate positively and those that correlate negatively with the hereditary form of hypertension, based on their loadings along the axis of the third principal component. Positively correlated compounds include acetate, cytosine, glycine, lactate; negatively correlated are cytidine, isobutyrate, 1-methylnicotinamide, 2′-deoxyuridine, uracil, nicotinamide, citrate, methyl acetoacetate, N-acetylglutamate.

In addition to assessing the metabolite loadings along the axes of the principal components, an analysis of intergroup differences in the content of metabolites in the daily urine of the studied animals was also performed (see the Table). Thus, a list of 12 urine metabolites was formed, the levels of which differed from the controls in animals with various forms of AH, also contributing the most to the separation of experimental groups in the coordinates of the principal component axes.

To identify possible associations between metabolites that made the greatest contribution to the observed intergroup differences, Pearson's correlation analysis with partial clustering was performed (Fig. 4). The highest correlation coefficients ($r > 0.7$) were observed between choline, N,N-dimethylglycine and N6-acetyllysine (the correlation was positive); these three metabolites also correlated negatively with methyl acetoacetate ($r < -0.5$). Citrate and 1-methylnicotinamide positively correlated with each other, as well as with N,N-dimethylglycine, N6-acetyllysine and cytidine ($r > 0.5$). Nicotinamide, 2′-deoxyuridine, isobutyrate, N-acetylglutamate and uracil were also positively correlated with each other ($r > 0.5$). Also, 2′-deoxyuridine positively correlated with cytidine with a correlation coefficient of $r = 0.64$.

Thus, certain associations between urine metabolites were found, which may provide markers for three hypertensive conditions that are different in their genesis.

Discussion

Choline and homocysteine metabolism
Choline participates in lipid metabolism, in the formation of cell membranes, and in the synthesis of the neurotransmitter acetylcholine (Zeisel, 2000). Choline oxidase and betaine aldehyde dehydrogenase oxidize choline to betaine. Betaine, in its turn, is a methyl group donor for betaine homocysteine methyltransferase that is involved in the remethylation of homocysteine to methionine resulting in the production of N,N-dimethylglycine (an alternative pathway for homocysteine utilization in the folate cycle). Normally, homocysteine should not accumulate in the body; its blood level elevation

А.А. Сорокоумова, А.А. Серяпина, Ю.К. Политыко
Л.В. Яньшоле, Ю.П. Центалович, М.А. Гилинский, А.Л. Маркель

Метаболомный профиль мочи крыс
с артериальной гипертонией разного генеза

2024
28 • 3
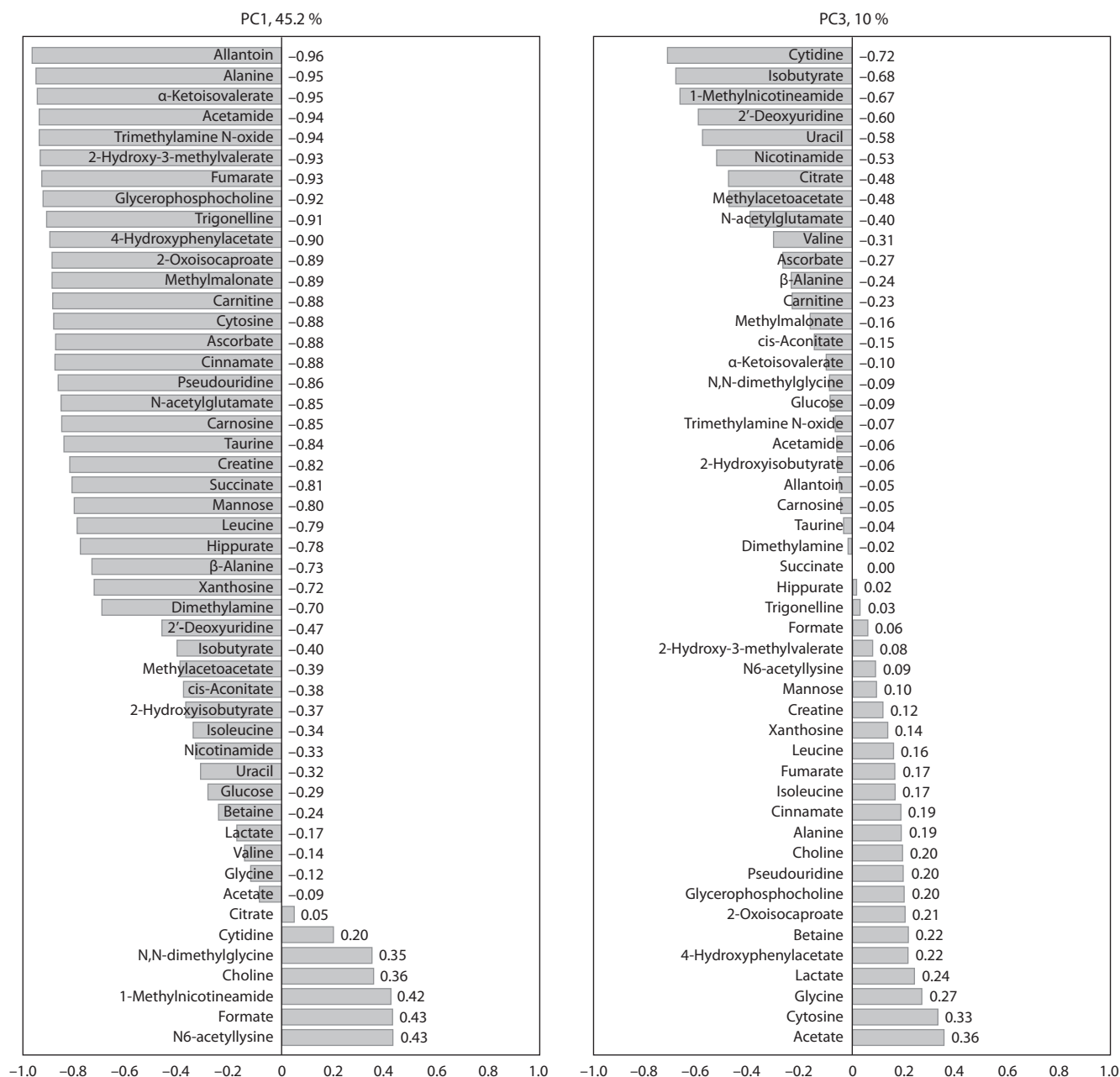
Fig. 3. PCA loadings of urine metabolite levels (PC1, PC3).

Significant differences in the content of individual metabolites in the urine of hypertensive rats of three groups (ISIAH, L-NAME, DOCA) from the controls

Metabolites	Hypertensive rat groups			Metabolites	Hypertensive rat groups		
	ISIAH	L-NAME	DOCA		ISIAH	L-NAME	DOCA
Choline		***		2'-Deoxyuridine	**		
N,N-dimethylglycine		*	*	Uracil	***		
N6-acetyllysine		***	*	Nicotinamide	**		
1-Methylnicotinamide	***		***	Citrate	***		
Cytidine	***			Methyl acetoacetate	**		
Isobutyrate	***			N-acetylglutamate	**		

N o t e. Color indicates an increase (orange) or a decrease (blue) in metabolite levels when compared to control WAG rats. Mann–Whitney test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
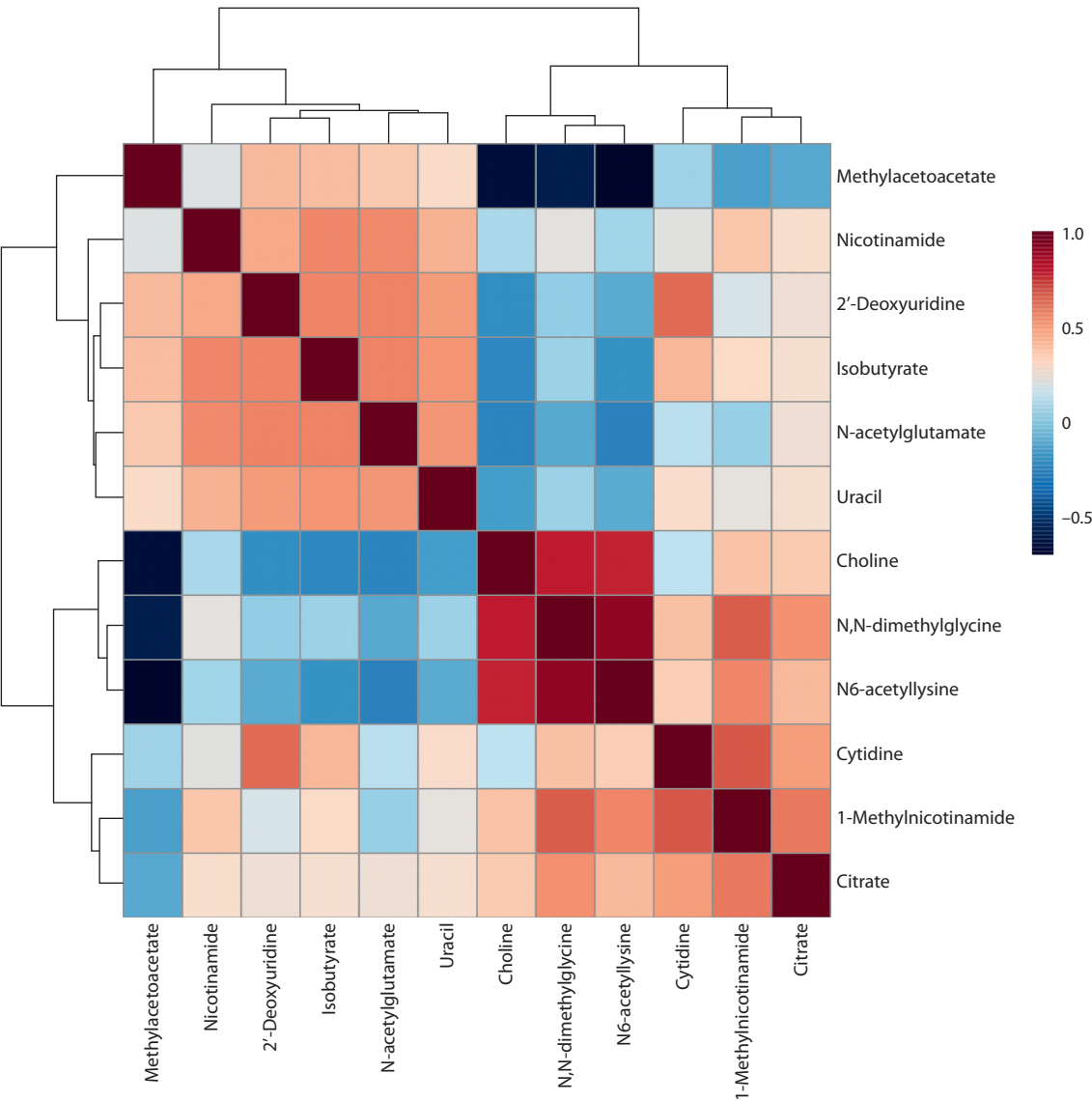
Fig. 4. Pearson's correlation coefficients between the studied parameters (urinary metabolite levels).

increases the risk of neurodegenerative and cardiovascular diseases (Wald et al., 2002). An increase in the homocysteine blood concentration causes damage of endothelium, activation of platelet aggregation, and formation of atherosclerotic plaques (Paré et al., 2009; Ganguly, Alam, 2015). Elevated blood levels of N,N-dimethylglycine correlate with increased homocysteine in patients with chronic renal failure (McGregor et al., 2001). Since L-NAME administration inhibits nitric oxide synthesis and leads to endothelial dysfunction, observed simultaneous increase in the levels of choline and N,N-dimethylglycine in the urine of rats in this group is of interest for further research. In the literature, there is evidence of opposite effects of exogenous homocysteine and L-NAME administration – homocysteine increased the expression of NO synthase, which was inhibited by L-NAME (Celotto et al., 2010). At the same time, administration of L-NAME normalized the levels of homocysteine and its metabolites in the blood plasma of rats with cholestasis induced by bile duct ligation (Ebrahimkhani et al., 2005).

N6-acetyllysine is increased in urine of L-NAME and DOCA rats. At this moment, the biological role of this compound has not been determined; however, there is some evidence of its association with the complications of type 1 and type 2 diabetes (Niewczas et al., 2017; Xu et al., 2023). Nevertheless, the lack of description of any pathogenetic mechanisms do not allow us to consider N6-acetyllysine as a potential biomarker of hypertension. High correlation coefficients were observed between N6-acetyllysine, choline and N,N-dimethylglycine. Association of choline and N,N-dimethylglycine with homocysteine and regulation of NO synthase was described above – perhaps the role of these compounds in the development of hypertension caused by endothelial dysfunction will be clarified in the future.

Nicotinamide metabolism
Nicotinamide is converted to 1-methylnicotinamide by the liver enzyme nicotinamide-N-methyltransferase. Nicotinamide-N-methyltransferase also promotes remethylation of

А.А. Сорокоумова, А.А. Серяпина, Ю.К. Политыко
Л.В. Яньшоле, Ю.П. Centaлович, М.А. Гилинский, А.Л. Маркель

Метаболомный профиль мочи крыс
с артериальной гипертонией разного генеза

2024
28•3

homocysteine to S-adenosylmethionine (another pathway for homocysteine utilization) (Hong et al., 2018), positively correlates with obesity and insulin resistance (Kannt et al., 2015), and presumably regulates the expression of fructose-1,6-bisphosphatase involved in the process of gluconeogenesis (Visinoni et al., 2008). Nicotinamide has been shown to prevent cytochrome C release and caspase induction, thereby maintaining mitochondrial membrane potential and exerting a cytoprotective effect on the endothelium of small cerebral vessels (Chong et al., 2002).

There is also evidence that nicotinamide may exhibit anti-inflammatory activity by inhibiting the expression of thromboplastin and CD11b antigen (Ungerstedt et al., 2003). Intravenous administration of 1-methylnicotinamide had an antithrombotic effect by activating the prostacyclin and cyclooxygenase pathways for the inflammation development (Chlopicki et al., 2007). Urine levels of nicotinamide and 1-methylnicotinamide were reduced in ISIAH rats, while in the DOCA-salt group, 1-methylnicotinamide was increased, which may indicate the role of inflammation in both hereditary AH and pharmacologically induced AH pathogenesis.

Pyrimidine metabolism

Serious disorders of pyrimidine metabolism, as a rule, are associated with dysfunction of enzymes, most often of dihydropyrimidine dehydrogenase or dihydropyrimidinase; they occur in early childhood, being systemic in nature and manifesting themselves in mental retardation and seizures (Nyhan, 2005). ISIAH rats do not have such symptoms, although there is a decrease in the levels of cytidine, 2′-deoxyuridine and uracil in their urine compared to normotensive controls. Interpretation of these data is difficult due to only a small number of individual studies: for example, in patients with chronic renal failure, reduced levels of 1-methyladenosine, 1-methylguanosine, N2,N2-dimethylguanosine and N4-acetylcytidine renal excretion were found (Niwa et al., 1998). There is also evidence that cyticoline (cytidine-5′-diphosphocholine) synthesized from cytidine and choline may have choline-like effects on membrane metabolism and cholinergic signaling (Yilmaz et al., 2008). However, these results are not sufficient to propose the products of pyrimidine metabolism in urine as markers of a hypertensive state.

Urea and nitric oxide cycle

N-acetylglutamate is an important participant in the urea cycle; it is synthesized in mitochondria from acetyl CoA and glutamate by the enzyme N-acetylglutamate synthase. A deficiency of N-acetylglutamate synthase or N-acetylglutamate itself causes disturbances in the urea cycle and accumulation of free ammonium ions in the blood – hyperammonemia (Tuchman et al., 2008). Increased serum ornithine concentrations have previously been observed in ISIAH rats (Seryapina et al., 2023), which, combined with decreased urinary N-acetylglutamate, suggests that disturbances in nitric oxide synthesis play a significant role in the hypertensive status of ISIAH rats.

Tricarboxylic acid cycle

Citrate is involved in the tricarboxylic acid (TCA) cycle, and a decrease in its urine concentration correlated with the development of hypertension in a study involving volunteers (Chachaj et al., 2020); however, in addition to citrate, the levels of other metabolites participating in the Krebs cycle were also changed: fumarate and trans-aconitate were decreased, methyl malonate was increased. In this study, in ISIAH rats, only citrate urinary level was reduced, so it seems incorrect to claim a serious disturbance of the TCA cycle. Citrate is also known to prevent the crystallization of calcium salts and the formation of kidney stones, therefore, low urine citrate may be linked to the disturbance in the renal mechanism of calcium excretion; besides, there are studies showing association of low citrate excretion with an increased insulin resistance (Cupisti et al., 2007). However, it is difficult to determine the mechanism of citrate decrease in this study, so it seems inappropriate to propose it as a marker of hereditary stress-related hypertension.

Short-chain fatty acids metabolism

Isobutyrate and methyl acetoacetate are derivatives of the so-called short-chain fatty acids (SCFAs), which are mainly produced by the gut microbiota. A decrease in their production causes intestinal inflammation and dysfunction, and kidney failure, which in turn contributes to increased blood pressure (Kim et al., 2018; Felizardo et al., 2019). It has been shown that short-chain fatty acids can bind to various G protein-coupled receptors. These receptors are located in many tissues and interact with their ligands in different ways (Chen et al., 2020). The effects of SCFAs include modulation of cytokine synthesis, regulation of differentiation and activation of macrophages, neutrophils and T-lymphocytes, and reduction of TNF-α and IL-12 production (Corrêa-Oliveira et al., 2016). In the urine of ISIAH rats, reduced contents of isobutyrate and methyl acetoacetate are observed. Apparently, their hypertensive status is similar in the parameters of the gut microbiota to SHR rats with spontaneous hypertension, in which a reduced number of bacteria producing acetate and butyrate was found (Yang et al., 2015).

Conclusion

Thus, in ISIAH rats with hereditary stress-dependent hypertension, judging by the characteristics of urine metabolites, disturbances in the nitric oxide cycle (decreased levels of N-acetylglutamate), changes in the function of gut microbiota (decreased isobutyrate and methyl acetoacetate), and the participation of inflammatory processes in formation of hypertensive status (decrease in nicotinamide and 1-methylnicotinamide levels) may be suggested. The data obtained complement our previous study (Seryapina et al., 2023): when comparing serum metabolic profiles of ISIAH and WAG rats, alterations in the nitric oxide cycle were also found (increased ornithine blood level in ISIAH rats), as well as changes in content and ratio of SCFAs (increased isobutyrate and decreased 2-hydroxyisobutyrate), and decreased concentrations of betaine and tryptophan, which have anti-inflammatory properties. Therefore, a decrease in the urinary N-acetylglutamate, isobutyrate, methyl acetoacetate, nicotinamide and 1-methylnicotinamide, in combination with the changes in the serum metabolome listed above, may be considered as a set of potential markers for hereditary stress-dependent hypertension in ISIAH rats.

Pharmacologically induced forms of AH (L-NAME and DOCA+NaCl) are identified by different metabolomic markers (increased urinary levels of choline, N,N-dimethylglycine, 1-methylnicotinamide). These two groups are positioned in the coordinates of the first two principal components in such a way that they actually do not overlap with the groups of control rats and ISIAH rats with hereditary AH. This occurs despite the fact that a common metabolic link is suggested between ISIAH and L-NAME rats, associated with impairment of endothelial function, and, possibly, NO synthesis. The results obtained demonstrate substantial metabolic differences between the "naturally" developing hereditary form of AH and two others caused by external pharmacological influences, which, in fact, made it possible to identify a set of specific metabolomic markers of hereditary, or "primary", hypertension, distinguishing it from symptomatic, or "secondary" ones.

References

Basting T., Lazartigues E. DOCA-salt hypertension: an update. *Curr. Hypertens. Rep.* 2017;19(4):32. DOI 10.1007/s11906-017-0731-4

Biancardi V.C., Bergamaschi C.T., Lopes O.U., Campos R.R. Sympathetic activation in rats with L-NAME-induced hypertension. *Braz. J. Med. Biol. Res.* 2007;40(3):401-408. DOI 10.1590/S0100-879X2006005000077

Bouatra S., Aziat F., Mandal R., Guo A.C., Wilson M.R., Knox C., Bjorndahl T.C., Krishnamurthy R., Saleem F., Liu P., Dame Z.T., Poelzer J., Huynh J., Yallou F.S., Psychogios N., Dong E., Bogumil R., Roehring C., Wishart D.S. The human urine metabolome. *PLoS One.* 2013;8(9):e73076. DOI 10.1371/journal.pone.0073076

Carey R.M., Moran A.E., Whelton P.K. Treatment of hypertension: a review. *JAMA.* 2022;328(18):1849-1861. DOI 10.1001/jama.2022.19590

Celotto A.C., Fukada S.Y., Laurindo F.R.M., Haddad R., Eberlin M.N., de Oliveira A.M. Chronic hyperhomocysteinemia impairs vascular function in ovariectomized rat carotid arteries. *Amino Acids.* 2010; 38(5):1515-1522. DOI 10.1007/s00726-009-0368-y

Chachaj A., Matkowski R., Gröbner G., Szuba A., Dudka I. Metabolomics of interstitial fluid, plasma and urine in patients with arterial hypertension: new insights into the underlying mechanisms. *Diagnostics.* 2020;10(11):936. DOI 10.3390/diagnostics10110936

Chan V., Hoey A., Brown L. Improved cardiovascular function with aminoguanidine in DOCA-salt hypertensive rats. *Br. J. Pharmacol.* 2006;148(7):902-908. DOI 10.1038/sj.bjp.0706801

Chen X.F., Chen X., Tang X. Short-chain fatty acid, acylation and cardiovascular diseases. *Clin. Sci.* (*Lond.*). 2020;134(6):657-676. DOI 10.1042/CS20200128

Chlopicki S., Swies J., Mogielnicki A., Buczko W., Bartus M., Lomnicka M., Adamus J., Gebicki J. 1-Methylnicotinamide (MNA), a primary metabolite of nicotinamide, exerts anti-thrombotic activity mediated by a cyclooxygenase-2/prostacyclin pathway. *Br. J. Pharmacol.* 2007;152(2):230-239. DOI 10.1038/sj.bjp.0707383

Chong Z.Z., Lin S.H., Maiese K. Nicotinamide modulates mitochondrial membrane potential and cysteine protease activity during cerebral vascular endothelial cell injury. *J. Vasc. Res.* 2002;39(2):131-147. DOI 10.1159/000057762

Corrêa-Oliveira R., Fachi J.L., Vieira A., Sato F.T., Vinolo M.A.R. Regulation of immune cell function by short-chain fatty acids. *Clin. Transl. Immunol.* 2016;5(4):e73. DOI 10.1038/cti.2016.17

Cupisti A., Meola M., D'Alessandro C., Bernabini G., Pasquali E., Carpi A., Barsotti G. Insulin resistance and low urinary citrate excretion in calcium stone formers. *Biomed. Pharmacother.* 2007;61(1): 86-90. DOI 10.1016/j.biopha.2006.09.012

Ebrahimkhani M.R., Sadeghipour H., Dehghani M., Kiani S., Payabvash S., Riazi K., Honar H., Pasalar P., Mirazi N., Amanlou M.,

Farsam H., Dehpour A.R. Homocysteine alterations in experimental cholestasis and its subsequent cirrhosis. *Life Sci.* 2005;76(21):2497-2512. DOI 10.1016/j.lfs.2004.12.009

Felizardo R.J.F., Watanabe I.K.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

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

Fürstenau C.R., Trentin D. da S., Gossenheimer A.N., Ramos D.B., Casali E.A., Barreto-Chaves M.L.M., Sarkis J.J.F. Ectonucleotidase activities are altered in serum and platelets of L-NAME-treated rats. *Blood Cells Mol. Dis.* 2008;41(2):223-229. DOI 10.1016/j.bcmd.2008.04.009

Ganguly P., Alam S.F. Role of homocysteine in the development of cardiovascular disease. *Nutr. J.* 2015;14(1):6. DOI 10.1186/1475-2891-14-6

Gupta V. Mineralocorticoid hypertension. *Indian J. Endocrinol. Metab.* 2011;15(Suppl.4):S298-S312. DOI 10.4103/2230-8210.86972

Hong S., Zhai B., Pissios P. Nicotinamide N-methyltransferase interacts with enzymes of the methionine cycle and regulates methyl donor metabolism. *Biochemistry.* 2018;57(40):5775-5779. DOI 10.1021/acs.biochem.8b00561

Kannt A., Pfenninger A., Teichert L., Tönjes A., Dietrich A., Schön M.R., Klöting N., Blüher M. Association of nicotinamide-N-methyltransferase mRNA expression in human adipose tissue and the plasma concentration of its product, 1-methylnicotinamide, with insulin resistance. *Diabetologia.* 2015;58(4):799-808. DOI 10.1007/s00125-014-3490-7

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

Markel A.L. Development of a new strain of rats with inherited stressinduced arterial hypertension. In: Sassard J. (Ed.) Genetic Hypertension. London, UK: John Libbey & Company, 1992;218: 405-407

McGregor D.O., Dellow W.J., Lever M., George P.M., Robson R.A., Chambers S.T. Dimethylglycine accumulates in uremia and predicts elevated plasma homocysteine concentrations. *Kidney Int.* 2001; 59(6):2267-2272. DOI 10.1046/j.1523-1755.2001.00743.x

Niewczas M.A., Mathew A.V., Croall S., Byun J., Major M., Sabisetti V.S., Smiles A., Bonventre J.V., Pennathur S., Krolewski A.S. Circulating modified metabolites and a risk of ESRD in patients with type 1 diabetes and chronic kidney disease. *Diabetes Care.* 2017; 40(3):383-390. DOI 10.2337/dc16-0173

Niwa T., Takeda N., Yoshizumi H. RNA metabolism in uremic patients: accumulation of modified ribonucleosides in uremic serum: technical note. *Kidney Int.* 1998;53(6):1801-1806. DOI 10.1046/j.1523-1755.1998.00944.x

Nyhan W.L. Disorders of purine and pyrimidine metabolism. *Mol. Genet. Metab.* 2005;86(1-2):25-33. DOI 10.1016/j.ymgme.2005.07.027

Paré G., Chasman D.I., Parker A.N., Zee R.R.Y., Mälarstig A., Seedorf U., Collins R., Watkins H., Hamsten A., Miletich J.P., Ridker P.M. Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13974 participants in the Women's Genome Health Study. *Circ. Cardiovasc. Genet.* 2009;2(2):142-150. DOI 10.1161/CIRCGENETICS.108.829804

Seryapina A.A., Malyavko A.A., Polityko Y.K., Yanshole L.V., Tsentalovich Y.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

А.А. Сорокоумова, А.А. Серяпина, Ю.К. Политыко
Л.В. Яньшоле, Ю.П. Центалович, М.А. Гилинский, А.Л. Маркель

Метаболомный профиль мочи крыс
с артериальной гипертонией разного генеза

2024
28·3

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

Tuchman M., Lee B., Lichter-Konecki U., Summar M.L., Yudkoff M., Cederbaum S.D., Kerr D.S., Diaz G.A., Seashore M.R., Lee H.S., McCarter R.J., Krischer J.P., Batshaw M.L. Cross-sectional multicenter study of patients with urea cycle disorders in the United States. *Mol. Genet. Metab.* 2008;94(4):397-402. DOI 10.1016/j.ymgme.2008.05.004

Ungerstedt J.S., Heimersson K., Söderström T., Hansson M. Nicotinamide inhibits endotoxin-induced monocyte tissue factor expression. *J. Thromb. Haemost.* 2003;1(12):2554-2560. DOI 10.1046/j.1538-7836.2003.00463.x

Visinoni S., Fam B.C., Blair A., Rantzau C., Lamont B.J., Bouwman R., Watt M.J., Proietto J., Favaloro J.M., Andrikopoulos S. Increased glucose production in mice overexpressing human fructose-1,6-bisphosphatase in the liver. *Am. J. Physiol. Endocrinol. Metab.* 2008;295(5):E1132-E1141. DOI 10.1152/ajpendo.90552.2008

Wald D.S., Law M., Morris J.K. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *Br. Med. J.* 2002;325(7374):1202-1206. DOI 10.1136/bmj.325.7374.1202

Xu J., Chen Q., Cai M., Han X., Lu H. Ultra-high performance liquid chromatography coupled to tandem mass spectrometry-based metabolomics study of diabetic distal symmetric polyneuropathy. *J. Diabetes Investig.* 2023;14(9):1110-1120. DOI 10.1111/jdi.14041

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

Yilmaz M.S., Coskun C., Suzer O., Yalcin M., Mutlu D., Savci V. Hypotensive effects of intravenously administered uridine and cytidine in conscious rats: involvement of adenosine receptors. *Eur. J. Pharmacol.* 2008;584(1):125-136. DOI 10.1016/j.ejphar.2008.01.044

Zeisel S.H. Choline: an essential nutrient for humans. *Nutrition.* 2000;16(7-8):669-671. DOI 10.1016/S0899-9007(00)00349-X

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 A., Sun H., Wu X., Wang X. Urine metabolomics. *Clin. Chim. Acta.* 2012;414:65-69. DOI 10.1016/j.cca.2012.08.016

Phylogenetic and pangenomic analyses of members of the family *Micrococcaceae* related to a plant-growth-promoting rhizobacterium isolated from the rhizosphere of potato (*Solanum tuberosum* L.)

S.Yu. Shchyogolev [1] ✉, G.L. Burygin [1, 2], L.A. Dykman [1], L.Yu. Matora [1]

[1] Institute of Biochemistry and Physiology of Plants and Microorganisms – Subdivision of the Saratov Federal Scientific Centre
   of the Russian Academy of Sciences, Saratov, Russia
[2] Saratov State Vavilov Agrarian University, Saratov, Russia

✉ shegolev_s@ibppm.ru

Abstract. We report the results of taxonomic studies on members of the family *Micrococcaceae* that, according to the 16S rRNA, internal transcribed spacer 1 (ITS1), average nucleotide identity (ANI), and average amino acid identity (AAI) tests, are related to *Kocuria rosea* strain RCAM04488, a plant-growth-promoting rhizobacterium (PGPR) isolated from the rhizosphere of potato (*Solanum tuberosum* L.). In these studies, we used whole-genome phylogenetic tests and pangenomic analysis. According to the ANI > 95 % criterion, several known members of *K. salina*, *K. polaris*, and *K. rosea* (including *K. rosea* type strain ATCC 186^T) that are related most closely to isolate RCAM04488 in the ITS1 test should be assigned to the same species with appropriate strain verification. However, these strains were isolated from strongly contrasting ecological and geographical habitats, which could not but affect their genotypes and phenotypes and which should be taken into account in evaluation of their systematic position. This contradiction was resolved by a pangenomic analysis, which showed that the strains differed strongly in the number of accessory and strain-specific genes determining their individuality and possibly their potential for adaptation to different ecological niches. Similar results were obtained in a full-scale AAI test against the UniProt database (about 250 million records), by using the AAI-profiler program and the proteome of *K. rosea* strain ATCC 186^T as a query. According to the AAI > 65 % criterion, members of the genus *Arthrobacter* and several other genera belonging to the class *Actinomycetes*, with a very wide geographical and ecological range of sources of isolation, should be placed into the same genus as *Kocuria*. Within the paradigm with vertically inherited phylogenetic markers, this could be regarded as a signal for their following taxonomic reclassification. An important factor in this case may be the detailing of the gene composition of the strains and the taxonomic ratios resulting from analysis of the pangenomes of the corresponding clades.

Key words: *Arthrobacter*; *Kocuria*; *Micrococcaceae*; pangenome; PGPR; phylogenetic analysis; strain verification; *Solanum tuberosum* L.

Филогенетический и пангеномный анализ представителей семейства *Micrococcaceae*, родственных стимулирующей рост растений ризобактерии, изолированной из ризосферы картофеля (*Solanum tuberosum* L.)

С.Ю. Щеголев [1] ✉, Г.Л. Бурыгин [1, 2], Л.А. Дыкман [1], Л.Ю. Матора [1]

[1] Институт биохимии и физиологии растений и микроорганизмов Федерального исследовательского центра
   «Саратовский научный центр Российской академии наук», Саратов, Россия
[2] Саратовский государственный университет генетики, биотехнологии и инженерии им. Н.И. Вавилова, Саратов, Россия

✉ shegolev_s@ibppm.ru

Аннотация. Исследованы представители семейства *Micrococcaceae*, родственные, согласно тестам 16S рРНК, ITS1 (транскрибируемый межгенный спейсер), средней нуклеотидной идентичности (ANI) и средней аминокислотной идентичности (AAI), штамму RCAM04488 *Kocuria rosea* – стимулирующей рост растений ризобактерии

С.Ю. Щеголев, Г.Л. Бурыгин
Л.А. Дыкман, Л.Ю. Матора

Филогенетический и пангеномный анализ
представителей семейства *Micrococcaceae*

2024
28 • 3

(PGPR), изолированному из ризосферы картофеля (*Solanum tuberosum* L.) с использованием полногеномных филогенетических тестов и пангеномного анализа. Согласно критерию ANI > 95 %, ряд известных представителей видов *K. salina*, *K. polaris* и *K. rosea* (включая типовой штамм *K. rosea* ATCC 186^{T}), наиболее близкородственных изоляту в тесте ITS1, должны быть приписаны к одному и тому же виду с соответствующей верификацией штаммов. Однако указанные штаммы были выделены из весьма контрастных по экологии и географии мест обитания, что не могло не сказаться на их генотипе и фенотипе и должно быть так или иначе учтено в оценках их систематического положения. Данное противоречие проясняют результаты пангеномного анализа, продемонстрировавшие существенные различия в этих штаммах количества акцессорных и штамм-специфичных генов, определяющих их индивидуальность и, возможно, потенциал для адаптации к различным экологическим нишам с соответствующими фенотипическими признаками. Аналогичные результаты получены в тесте AAI в полномасштабном варианте его применения против базы данных UniProt (около 250 млн записей) с использованием программы AAI-profiler и протеома штамма *K. rosea* ATCC 186^{T} в качестве запроса. Согласно критерию AAI > 65 %, в один и тот же род с *Kocuria* должны быть объединены представители рода *Arthrobacter* и некоторых других родов, относящихся к классу актиномицетов, с весьма широким географическим и экологическим спектром источников их выделения. В рамках парадигмы о вертикально наследуемых филогенетических маркерах это можно трактовать как сигнал для их последующей таксономической переквалификации. Важным фактором при этом может быть детализация генного состава штаммов и таксономических соотношений, получаемых в результате анализа пангеномов соответствующих клад.

Ключевые слова: *Arthrobacter*; *Kocuria*; *Micrococcaceae*; пангеном; PGPR; филогенетический анализ; верификация штаммов; *Solanum tuberosum* L.

Introduction

The paper (Potanina et al., 2017) presented the results of phylogenetic studies on the plant-growth-promoting (Kargapolova et al., 2017) bacterial strain *Kocuria rosea* RCAM04488, isolated from surface-sterilized roots of potato (*Solanum tuberosum* L. 'Kondor'). For the genotypic taxonomic identification of this isolate, sequences of the 16S rRNA gene (GenBank MF754147.1) and of the ITS1 transcribed intergenic spacer (GenBank MF765458.1) were obtained. By using 16S rRNA (Potanina et al., 2017), the evolutionary proximity of this isolate to the genera *Rothia*, *Arthrobacter*, and *Zhihengliuella*, as well as to members of the species *K. rosea* and *K. polaris*, was ascertained.

Kocuria is a genus of gram-positive bacteria of the family *Micrococcaceae*, phylum *Actinobacteria*, which are either aerobic or facultatively anaerobic. To date, 32 *Kocuria* species have been identified.

Kocuria bacteria have been found on human and animal skin and mucous membranes. They are generally considered nonpathogenic but can be detected in some urinary tract infections and in hepatobiliary, cardiovascular, nervous system, and gastrointestinal infections (Kandi et al., 2016). Although *Kocuria* can infect immunocompromised patients, they are weakly pathogenic and are highly sensitive to antibiotics (Odeberg et al., 2023).

Many *Kocuria* members, including the type species *K. rosea*, live in soil (Stackebrandt, Schumann, 2015) and are endophytes; i. e., they have been isolated from the rhizosphere and tissues of many plants. Endophytic *Kocuria* are inhibitory to several pathogenic fungi and bacteria (Cho et al., 2007; Rao et al., 2015; Andreolli et al., 2016; Candra et al., 2022; Tavarideh et al., 2022; Tedsree et al., 2022). In addition, some of them have properties of plant-growth-promoting rhizobacteria (PGPR), because they produce indole-3-acetic acid and other phytohormones and because they increase plant resistance to stress (Passari et al., 2017; Li et al., 2020).

Bacteriocins from nonlactic acid bacteria, in particular variacin from *K. varians*, can be used for the biopreservation of food (cheese and meat) products (Gálvez et al., 2010). The *K. rosea* exopolysaccharide, kocuran, is used in the production of antimicrobial coatings (Kumar, Sujitha, 2014).

A number of soil *Kocuria* can degrade some xenobiotics, in particular phthalate esters, pesticides, and salts of arsenic, copper, and other heavy metals (Kaur et al., 2015; Román-Ponce et al., 2016; Hansda et al., 2017; Mukherjee et al., 2018; Vital et al., 2019; Yastrebova, Plotnikova, 2020; González-Benítez et al., 2021; Mawang et al., 2021). Various *Kocuria* have been recovered from soils; marine sediments; meat, dairy and seafood products; beer; seawater; rocks; livestock bedding; manure; surface spring water; and other sources (Church et al., 2020).

The aim of this research was to obtain phylogenetic and genetic information on *Micrococcaceae* members related to *K. rosea* isolate RCAM04488 according to the following tests: 16S rRNA (Potanina et al., 2017), internal transcribed spacer (ITS1), average nucleotide identity (ANI), and average amino acid identity (AAI). Whole-genome phylogenetic tests and pangenomic analysis were applied to the known results of whole-genome DNA sequencing of these strains.

Materials and methods

In our phylogenetic and pangenomic studies, we used the published genomes of the bacterial strains under study, brought into consideration as a result of the use of the bioinformatic resources mentioned below. The characteristics of the genomes are summarized in Results and Discussion and in Supplementary Materials.

The blastn program[1] was used in taxonomic analysis with the genetic sequence of the ITS1 intergenic spacer of *K. rosea* RCAM04488 (GenBank MF765458.1). Strain RCAM04488 is part of the Russian Collection of Agricultural Microorganisms (https://arriam.ru/kollekciya-kul-tur1, accessed 09/06/2023; RCAM04488) and of the Collection of Rhizosphere Microor-

---

[1] Standard Nucleotide BLAST. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome. Accessed 09/06/2023.

ganisms, Institute of Biochemistry and Physiology of Plants and Microorganisms, Russian Academy of Sciences (IBPPM RAS) (http://collection.ibppm.ru, accessed 09/06/2023; IBPPM604).

The average nucleotide identity (ANI) test[2] (Goris et al., 2007; Rodriguez-R, Konstantinidis, 2014; Jain et al., 2018), in its OAT modification[3] (Lee et al., 2016), was used for quantitative species/genus demarcation on the basis of whole-genome sequencing of the strains' DNA. Note that the 16S rRNA, ITS1, ANI, and AAI tests were developed within the paradigm of vertically inherited prokaryotic genotypic traits by using markers from the core component of the pangenome (Tettelin, Medini, 2020) without any account of the effects of horizontal gene transfer (HGT) in its accessory (optional) and strain-specific parts. The HGT effects largely control the variety of phenotypic traits that determine, in particular, the ability of bacteria and archaea to adapt and function in diverse, frequently changing ecological niches (Koonin, 2012). These traits are taken into account in the analysis of the systematic position of entries that is based on the polyphasic approach, which is very common in the traditional systematics of the prokaryotes (Oren, Garrity, 2014). Hence follows the obvious conventionality of phylogenetic analysis within any scheme using only vertically inherited phylogenetic markers, as do possible contradictions of its results to the traditional classification and nomenclature of the prokaryotes (Shchyogolev, 2021). This probably explains, in particular, the need for their verification with appropriate classification changes, which turned out to be relevant for about 60 % of the Genome Taxonomy Database[4] (GTDB) entries analyzed in Parks et al. (2018).

We used the PGAP program[5] (Chen et al., 2018) to obtain information on pangenome composition for selected phylogenetic groups of bacteria (clades).

We used the AAI-profiler program[6] (Medlar et al., 2018) to evaluate the whole-genome systematic position of *Kocuria* members relative to the entries from the UniProt protein property database[7] (250 million records) for *Kocuria* members found by the 16S rRNA and ITS1 tests to be closely related to *K. rosea* RCAM04488. In particular, the program detects and visualizes possible contradictions in the classification of pro- and eukaryotes and microbial contamination (Medlar et al., 2018). To visualize and analyze phylogenetic trees, we used the MEGA11 program[8].

Results and discussion

Strains closely related to *Kocuria rosea* isolate RCAM04488 in the ITS1 test

Use of blastn with the sequence of the ITS1 intergenic spacer of *K. rosea* RCAM04488 (GenBank MF765458.1) against the RefSeq Genome Database (refseq_genomes) with the *Ko-*

curia option (taxid:57493) yielded a set of 12 hits. These included ITS1 sequences from *Kocuria* members and references to the results of whole-genome DNA sequencing of all (mostly type) strains in the set (Supplementary Material 1)[9]. Of note, in the BacDive database (Reimer et al., 2022), on the web page[10], information is given on 29 entries representing the type strain of *K. rosea*, including *K. rosea* DSM 20447[T], which, according to the 16S rRNA test, is evolutionarily close to the isolate we are studying (Potanina et al., 2017). Among the results of similar studies conducted by us with the use of the resource[11] (data not shown), *K. rosea* strain ATCC 186[T] (characteristics summarized in Supplementary Material 1) is indicated as the type strain. It is also found on the *K. rosea* DSM 20447[T] Bac*Dive* web page, presented as the type strain in Trachtenberg et al. (2018), and used for comparison in pangenomic analysis and in AAI-profiler studies.

The BLAST distance tree for ITS1 of *K. rosea* RCAM04488 (GenBank MF765458.1) with the indicated 12 hits (Fig. 1) shows clustering of *K. rosea* RCAM04488 with members of *K. rosea*, *K. polaris*, and *K. salina*, in agreement with the main results of the 16S rRNA test (Potanina et al., 2017). The value of the ITS1 sequence identity $I = 99.1$ % among those marked in Figure 1 and the general structure of the cluster, the node of which is marked by a red dot in Figure 1, indicates that *K. rosea* RCAM04488 is most closely related taxonomically to *K. rosea* strain AF099C18 in this test. Strain AF099C18 belongs to the type species of the genus *Kocuria*, the members of which are found in very diverse ecological niches (Stackebrandt, Schumann, 2015) (see Introduction and Supplementary Material 1), and was isolated in Eugene (Oregon, USA) from a dust sample during the study of the effect of the finishing of indoor surfaces on bacterial viability (Hu et al., 2019).

The other members of this cluster include *K. rosea* strain DSM 20447[T], a member of the *Actinobacteria*; *K. polaris* type strain CMS 76or[T], isolated from cyanobacterial mats in McMurdo Dry Valley, Antarctica (Gundlapally et al., 2015); and *K. salina* strain CV6 29, isolated in the vicinity of Lake Schott el Djerid (Tunisia) from the roots of *Cistanche violacea*, a desert plant of the Orobanchaceae family that lives on the roots of host plants (tamarix, black saxaul). The adaptation of these strains to such a contrasting diversity of habitats and conditions can be attributed to the phenotypic traits encoded by the genes in the accessory and strain-specific parts of the *Kocuria* pangenome, to which a large contribution is probably made by HGT (Treangen, Rocha, 2011; Koonin, 2012).

At the whole-genome level with housekeeping genes, however (ANI test, orthologous genes), all four strains in the monophyletic group with sequence MF765458.1 (Fig. 1) obey the ANI > 95 % condition and, therefore, should be considered as belonging to the same species (Jain et al., 2018). This is illustrated by the ANI dendrogram (UPGMA variant) obtained by the OAT method described in Lee et al. (2016) (Supplementary Material 2). In addition to the four strains listed above, the ANI > 95 % condition, which groups the

---

[2] ANI/AAI-Matrix. http://enve-omics.ce.gatech.edu/g-matrix. Accessed 09/06/2023.

[3] OAT. https://www.ezbiocloud.net/tools/orthoani. Accessed 09/06/2023.

[4] Genome Taxonomy Database. https://gtdb.ecogenomic.org. Accessed 09/06/2023.

[5] PGAweb. http://pgaweb.vlcc.cn/analyze. Accessed 09/06/2023.

[6] AAI-profiler. http://ekhidna2.biocenter.helsinki.fi/AAI. Accessed 09/06/2023.

[7] Find your protein. https://www.uniprot.org. Accessed 09/06/2023.

[8] MEGA. https://www.megasoftware.net. Accessed 09/06/2023.

[9] Supplementary Materials 1–4 are available at:
https://vavilov-icg.ru/download/pict-2024-28/appx12.pdf

[10] *Kocuria rosea* DSM 20447 is an aerobe, mesophilic bacterium of the family *Micrococcaceae*. https://bacdive.dsmz.de/strain/7641. Accessed 09/06/2023.

[11] Search EzBioCloud Database. https://www.ezbiocloud.net. Accessed 09/06/2023.

С.Ю. Щеголев, Г.Л. Бурыгин
Л.А. Дыкман, Л.Ю. Матора

Филогенетический и пангеномный анализ
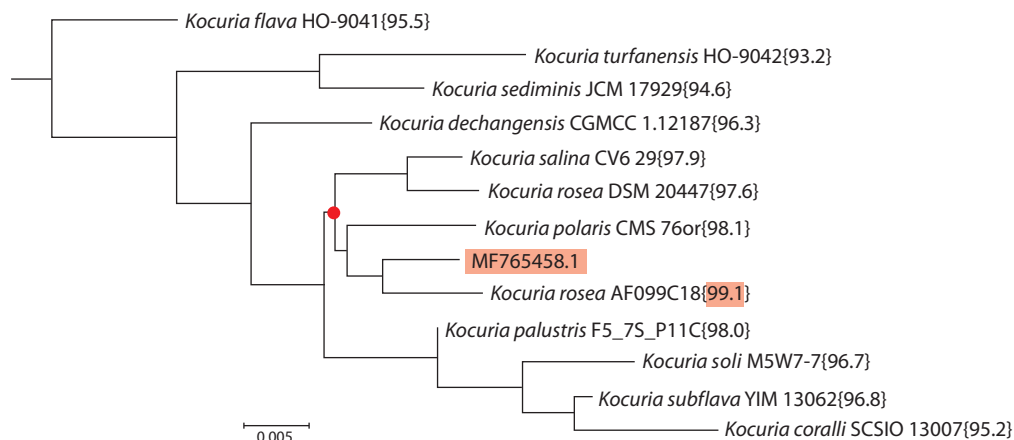представителей семейства *Micrococcaceae*

2024
28•3

Fig. 1. Distance tree for BLAST hits with the ITS1 query sequence (GenBank MF765458.1) of *K. rosea* RCAM04488 against the RefSeq Genome Database (refseq_genomes) with the *Kocuria* option (taxid:57493).

The identity values (*I*, %) between sequence MF765458.1 and hit sequences are indicated in curly brackets.
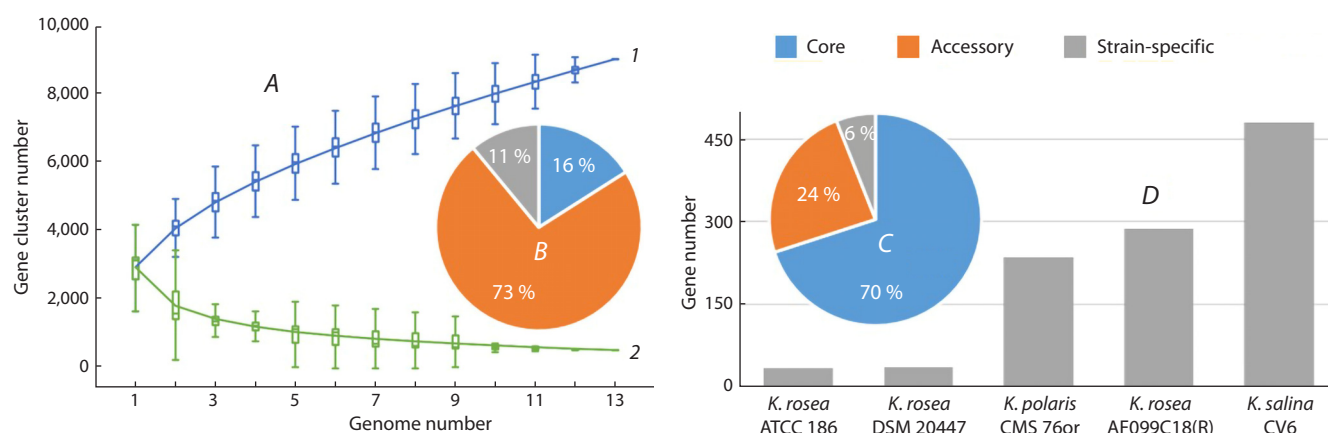

Fig. 2. Pangenomic analysis of the *Kocuria* strains included in Figure 1 and Supplementary Material 2. *A*, size of the pangenome (*1*) and core genome (*2*) versus the number of genomes being considered. *B*, pie chart of pangenome contents in Fig. *A* with core, accessory, and strain-specific genes. *C*, pie chart of the pangenome contents corresponding to the monophyletic group highlighted in Figure 1 (red dot), with core, accessory, and strain-specific genes. *D*, histogram of the number of strain-specific genes in the strains corresponding to the pangenome in Fig. *C*.

strains into the same species (Jain et al., 2018), is also obeyed by the *K. sediminis* JCM 17929[T]–*K. turfanensis* HO-9042[T] pair. Unlike 16S rRNA and ANI, the ITS1 test (Fig. 1) does not have quantitative criteria for grouping/demarcating taxa. In our case, this turned out to be possible only by combining the OrthoANI dendrogram with the heat map in Supplementary Material 2.

To elucidate those whole-genome details that determine the strains' individuality and possibly also differences in adaptation behavior in contrasting ecological niches, we performed a pangenomic analysis (Chen et al., 2018; Tettelin, Medini, 2020) of the strains included in Figure 1 and Supplementary Material 2 by using the PGAweb program. To the 12 genomes of these strains, we added the genome of *K. rosea* strain ATCC 186[T] (see above).

Figure 2*A* shows the dependences of pangenome size (curve *1*) and core genome size (curve *2*) on the number of genomes being considered for the set corresponding to Figure 1. Different colors and numbers in the pie chart of Figure 2*B*

denote the content of the core, accessory, and strain-specific genes in their total pool for a clade of 13 strains. The general appearance of curves *1* and *2* indicates that this pangenome is of the open type, which means that it allows DNA exchange with the global prokaryote gene pool through a variety of mechanisms (Chen et al., 2018; Tettelin, Medini, 2020), including HGT (Treangen, Rocha, 2011; Koonin, 2012).

To elucidate subtle differences among these five genomes, which possibly contribute substantially to their distribution across different ecological niches and other individual phenotypic traits but are not evident in the ANI test (see Supplementary Material 2), we conducted a separate pangenomic analysis for 5 strains corresponding to the monophyletic group highlighted in Figure 1 (red dot) (Fig. 2*C*, *D*). For this clade of closely related species, the analysis showed a relatively high content of core genes (70 %), a lower content of accessory genes (24 %), and a low percentage of unique (strain-specific) genes (6 %), with very marked interstrain differences in their number (Fig. 2*D*).

The largest number of unique genes is present in the endophytic strain *K. salina* CV6 (480), which is followed by *K. rosea* AF099C18 (287) and *K. polaris* CMS 76or[T] (235). The letter R in *K. rosea* strain AF099C18 in Figure 2, *D* shows its status as a representative of *K. rosea* isolate RCAM04488 at the whole-genome level, which is the most closely related to it in the 16S rRNA (Potanina et al., 2017) and ITS1 tests (Fig. 1).

Strains related to *Kocuria rosea* RCAM04488 in the AAI-profiler test

Using AAI-profiler, we made an extended whole-genome evaluation of the systematic position of the *Kocuria* members related to *K. rosea* RCAM04488 according to the 16S rRNA (Potanina et al., 2017) and ITS1 tests (see above). This was done at the level of the UniProt database, which has about 250 million records as of autumn 2023. With allowance for the close kinship between *K. rosea* strains, which is shown in Supplementary Material 2, we chose *K. rosea* type strain ATCC 186[T] as the initial one for use in AAI-profiler.

The query was the proteome of *K. rosea* ATCC 186[T] (genome assembly GCF_006094695.1), used by AAI-profiler to determine AAI between the query proteome and the proteomes of the species members in UniProt and to construct an AAI distribution diagram (Fig. 3). AAI values are plotted on the horizontal axis, and the values of the MF (matched fraction, the proportion of query proteins that have matches in the species analyzed by the program) are plotted on the vertical axis. The diagram icons correspond to the species that received the highest scores, with account taken of AAI and coverage, i. e., the sum of the sequence identity values for all query proteins with established matches.

Related species, grouped and colored on the basis of genus, form a characteristic "cloud" in the diagram, with AAI values reflecting the evolutionary closeness of the UniProt strains and the query strain. The horizontal axis has icons for the species for which only individual proteins have been sequenced. The icons are colored according to genus (bacteria) or order (eukaryotes). Eukaryotic species are marked with rhombuses; bacteria, with circles; archaea, with crosses; and everything else (viruses, metagenomes, and unclassified samples), with squares. The vertical dashed lines in Figure 3 correspond to the AAI cutoff values for strain demarcation on the basis of genus (AAI > 0.65) and species (AAI > 0.9) under the conditions presented in the ANI/AAI-Matrix resource, on the website[12], and in Rodriguez-R and Konstantinidis (2014).

The results (Fig. 3) show that the query proteome corresponds to a set of 11 *K. rosea* strains with average coverage and AAI values of 0.989 and 99.8 %, respectively. The closest to this set among the classified ones is that of two *K. polaris* strains (including the type strain CMS 76or), the icon of which is located in the area of the diagram with AAI values > 90 % (to the right of the vertical dashed line with an abscissa of 0.9). In the AAI test, this means that all these 13 strains and some other *Kocuria* members (not shown here) belong to the same species (Luo et al., 2014).

As an example, the Figure 3 diagram includes *K. turfanensis* strain NBRC 107627[T], the icon of which is located within the
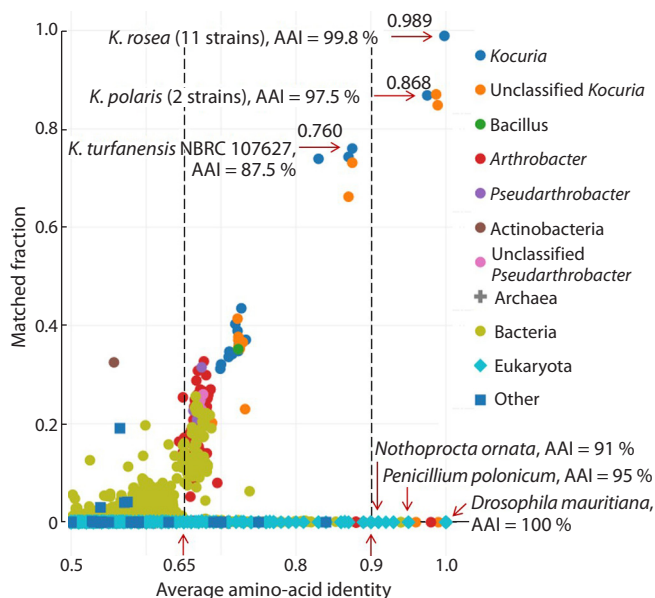
<hr/>

[12] Understanding Results. https://help.microbial-genomes.org/ understanding-results#distance. Accessed 09/06/2023.

Fig. 3. AAI distribution diagram for *K. rosea* ATCC 186[T], as found in the AAI-profiler output data.

Explanations are in the text.

range of 0.65 < ANI < 0.9, where most species belonging to the same genus in the AAI test (Luo et al., 2014) are concentrated. These include *Kocuria* of the following species: *coralli*, *flava*, *indica*, *marina*, *palustris*, *rhizophila*, *sediminis*, *soli*, *subflava*, *turfanensis*, *tytonicola*, *tytonis*, and *varians*. However, the same area in the diagram with coverage in the range 0.12–0.74 also includes members of the genus *Arthrobacter* (the most widely represented genus in the Fig. 3 diagram) and of the genera *Pseudarthrobacter*, *Micrococcus*, *Microbacterium*, and some other actinomycetes. The placement of these entries (and other bacterial species/strains marked with pink, green, and light green dots) into the main cluster of the genus *Kocuria*, which groups species related to the query strain *K. rosea* ATCC 186[T] (to the right of the dashed line with abscissa AAI = 0.65), could be interpreted as a signal for their probable taxonomic reclassification (Medlar et al., 2018) within the paradigm with vertically inherited phylogenetic markers (Koonin, 2012). However, a decision on this should be made after additional genotypic and phenotypic features are considered within a polyphasic approach (Oren, Garrity, 2014). These strains are listed together with their detailed characteristics in the AAI-profiler output.

Six icons corresponding to eukaryotes are located on the abscissa axis at the intraspecies level with AAI values > 0.9. The relatively small (symbolically zero) coverage values mean that only individual proteins have been sequenced for them (Medlar et al., 2018). The protein system from the fruit fly *Drosophila mauritiana* proved closest to that of *K. rosea* ATCC 186[T] on the basis of AAI = 100 %. Next in descending order of AAI values in the intraspecies interval 90 % < AAI < 100 % (Luo et al., 2014) are *Penicillium polonicum* (imperfect fungus), *Poeciliopsis prolifica* (small freshwater fish), *Drosophila sechellia* (another fruit fly species), *Nothoprocta ornata* (flightless bird), and *Hirsutella minnesotensis* (asexual propagating fungus). For clarity, some of these are marked

С.Ю. Щеголев, Г.Л. Бурыгин
Л.А. Дыкман, Л.Ю. Матора

Филогенетический и пангеномный анализ
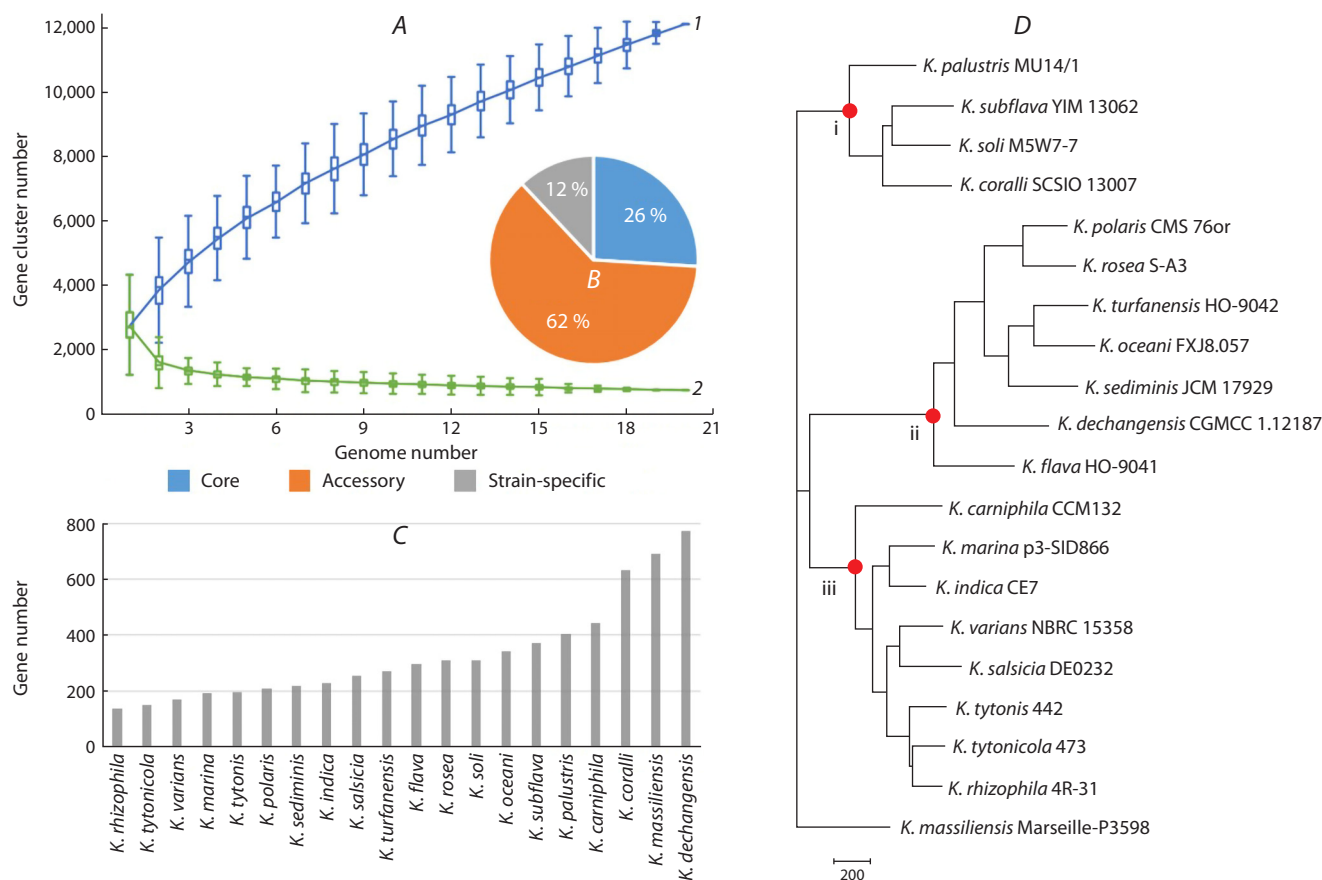представителей семейства *Micrococcaceae*

2024
28•3

Fig. 4. Results of pangenomic analysis of *Kocuria* strains with the reference genomes, listed in Supplementary Material 3. *A*, size of the pangenome (*1*) and core genome (*2*) versus the number of genomes being considered. *B*, pie chart of pangenome contents in Fig. *A* with core, accessory, and strain-specific genes. *C*, histogram of the number of strain-specific genes in the strains corresponding to the pangenome in Figs. *A* and *B*. *D*, phylogram for the set of strains from the pangenome in Figs. *A–C*.

The phylogram was obtained by the NJ method on the basis of the gene acquisition/loss matrix. Red dots and Roman numerals indicate nodes of monophyletic groups.

with arrows in Figure 3. In the same ANI > 90 % interval, the icons for members of the prokaryotes and the genera *Kocuria*, *Arthrobacter*, *Micrococcus* (all actinomycetes), and *Nitrosomonas* (β-proteobacteria) are shown on the abscissa axis.

For prokaryotes, this is explained by HGT (Treangen, Rocha, 2011; Medlar et al., 2018). For eukaryotes, which in our case include members of the animal and fungal kingdoms, the high homology between their protein systems and those of the genus *Kocuria* (class *Actinomycetia*) may be associated with symbiogenesis as a very probable mechanism of the origin of eukaryotes with the participation of prokaryotes (Dey et al., 2016; Provorov et al., 2018). Bioinformatic studies of this phenomenon have been reported, for example, in Markov and Kulikov (2005) and in Nikitin (2016). They provide data to show that although archaea [from which a considerable part of the eukaryotic genome originates (Stairs, Ettema, 2020)], α-proteobacteria (precursors to mitochondria), and cyanobacteria (precursors to plastids) play fundamental parts in symbiogenesis, substantial contributions to these processes are made by various bacteria, not limited to the above two taxa.

In the context of the above-mentioned AAI-profiler results, of interest is the information on the general genomic structure of *Kocuria* and *Arthrobacter* members. This information was obtained by pangenomic analysis with the PGAweb software

package described in Chen et al. (2018). In the database of the results of whole-genome DNA sequencing of prokaryotic strains[13], as of autumn 2023, we found a fairly representative set of genomes for 20 mostly type strains of *Kocuria* species, having the status of reference genomes (Supplementary Material 3) and used by us for pangenomic analysis (Fig. 4).

The Figure 4 results show the overall conservatism of the genomes being considered (26 % of the core genes) and the pronounced openness of the pangenome (Fig. 4*A*, *B*, curves *1* and *2*). The changes in the accessory (62 % of the total number of genes) and strain-specific (Fig. 4*C*) components of the pangenome, which reflect the species diversity of *Kocuria*, can also be attributed to the great diversity of habitats of these strains – from animal and plant organs and tissues to food, soil, air, and marine environments, including Antarctic cyanobacterial mats.

Figure 4*D* shows the phylogram presented in the final PGAweb results. It was obtained by the neighbor-joining (NJ) method on the basis of the gene acquisition/loss matrix for the pangenome as a whole (Chen et al., 2018). This tree shows a clear distribution of strains over three monophyletic groups (Fig. 4*D*, red dots). There is no sufficiently pronounced de-

---

[13] Genome. https://www.ncbi.nlm.nih.gov/datasets/genome. Accessed 09/06/2023.
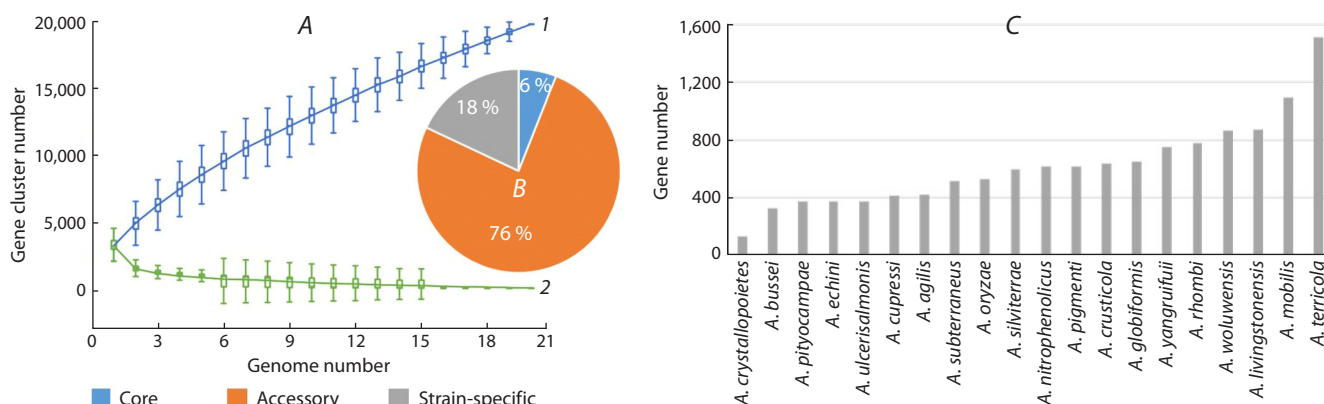
Fig. 5. Results of pangenomic analysis of *Arthrobacter* strains with reference genomes, listed in Supplementary Material 4. *A*, size of the pangenome (*1*) and core genome (*2*) versus the number of genomes being considered. *B*, pie chart of pangenome contents in Fig. *A* with core, accessory, and strain-specific genomes. *C*, histogram of the number of strain-specific genes in the strains corresponding to the pangenome in Figs. *A* and *B*.

pendence on geographical or environmental factors. It can only be noted that strains of animal origin are concentrated in Figure 4*D* in group iii.

For comparison and with account taken of the wide representation of *Arthrobacter* species in the AAI-profiler output (Fig. 3), we also performed a pangenomic analysis for the species of this genus corresponding to these members. The obtained set of predominantly type strains and genomes in the "Reference genomes" category, which we found in the Genome database, is presented in Supplementary Material 4.

Figure 5*A*, *B* shows the pangenome characteristics for the group of *Arthrobacter* species listed in Supplementary Material 4. Note that the relative number of core genes for the *Arthrobacter* group is much smaller than that for the *Kocuria* group (Supplementary Material 3, Fig. 4). The greater number of accessory and strain-specific genes in *Arthrobacter* (94 %), as compared with *Kocuria* (74 %), indicates higher overall genomic heterogeneity of the *Arthrobacter* group under consideration and greater openness of its pangenome. The number of strain-specific genes in each of the strains forming part of the *Arthrobacter* pangenome group varies in a wide range, from 130 (*A. crystallopoietes*) to 1,517 (*A. terricola*) (Fig. 5*C*).

Conclusions

We have shown that, according to the ANI test, the strains *K. salina* CV6 and *K. polaris* CMS 76or[T], together with *K. rosea* DSM 20447, *K. rosea* AF099C18, and *K. rosea* ATCC 186[T], formally within a phylogeny with vertically inherited markers, should be assigned to the same species (ANI > 95 %) with appropriate species verification of the strains. Because all five strains have been isolated from strongly contrasting ecological and geographical habitats, this fact could not but affect their genotypes and phenotypes and should be taken into account in the analysis of their systematic position.

We have clarified this contradiction by pangenomic analysis of a clade of 13 *Kocuria* strains closely related in the 16S rRNA and ITS1 tests to the *K. rosea* strain of interest, RCAM04488, isolated from surface-sterilized potato roots. The clade includes the above-mentioned *Kocuria* strains. The

analysis has shown the pangenome to be of the open type and has revealed large differences between the above strains in the content of accessory and strain-specific genes, which determine their individuality and possibly potential for adaptation to different ecological niches with the corresponding phenotypic traits. The largest number of unique genes, which are listed in the output of the PGAP program, was observed in the endophytic strain *K. salina* CV6 (480). This strain is followed by *K. rosea* AF099C18 (287), which is most closely related to *K. rosea* RCAM04488 in the 16S rRNA and ITS1 tests. These observations seem important for evaluating the possible gene content of *K. rosea* RCAM04488 in terms of its abilities as a PGPR. This will be the subject of our further work, which will use the results of whole-genome DNA sequencing of this strain.

Using AAI-profiler, we obtained similar results in a full-scale AAI test against the UniProt database (approximately 250 million records). In particular, these results confirm the need to assign *K. rosea* and *K. polaris* members and several other members of the genus *Kocuria* to the same species (AAI > 90 %). In the phylogenetic aspect, our most substantial finding is the established association of *Kocuria*, *Arthrobacter* (the genus most widely represented in these results), *Pseudarthrobacter*, *Micrococcus*, *Microbacterium*, and several other genera as members of the same genus according to the AAI > 65 % criterion. Within a paradigm with vertically inherited phylogenetic markers, this could be regarded as a signal for the following taxonomic reclassification of these entries. In this respect, it may help to comparatively evaluate their gene content and taxonomic relationships on the basis of pangenomic studies. However, to make this responsible decision, one should consider additional genotypic and phenotypic characteristics of the strains under study within a polyphasic approach.

References

Andreolli M., Lampis S., Zapparoli G., Angelini E., Vallini G. Diversity of bacterial endophytes in 3 and 15 year-old grapevines of *Vitis vinifera* cv. Corvina and their potential for plant growth promotion and phytopathogen control. *Microbiol. Res.* 2016;183:42-52. DOI 10.1016/j.micres.2015.11.009

С.Ю. Щеголев, Г.Л. Бурыгин
Л.А. Дыкман, Л.Ю. Матора

Филогенетический и пангеномный анализ
представителей семейства *Micrococcaceae*

2024
28 • 3

Candra R.T., Prasasty V.D., Karmawan L.U. Biochemical analysis of banana plants in interaction between endophytic bacteria *Kocuria rhizophila* and the fungal pathogen *Fusarium oxysporum* f. sp. *cubense* tropical race (*Foc* TR4). *Biol. Life Sci. Forum.* 2022;11(1):84. DOI 10.3390/IECPS2021-11990

Chen X., Zhang Y., Zhang Z., Zhao Y., Sun C., Yang M., Wang J., Liu Q., Zhang B., Chen M., Yu J., Wu J., Jin Z., Xiao J. PGAweb: a web server for bacterial pan-genome analysis. *Front. Microbiol.* 2018;9:1910. DOI 10.3389/fmicb.2018.01910

Cho K.M., Hong S.Y., Lee S.M., Kim Y.H., Kahng G.G., Lim Y.P., Kim H., Yun H.D. Endophytic bacterial communities in ginseng and their antifungal activity against pathogens. *Microb. Ecol.* 2007; 54(2):341-351. DOI 10.1007/s00248-007-9208-3

Church D.L., Cerutti L., Gürtler A., Griener T., Zelazny A., Emler S. Performance and application of 16S rRNA gene cycle sequencing for routine identification of bacteria in the clinical microbiology laboratory. *Clin. Microbiol. Rev.* 2020;33(4):e00053-19. DOI 10.1128/CMR.00053-19

Dey G., Thattai M., Baum B. On the archaeal origins of eukaryotes and the challenges of inferring phenotype from genotype. *Trends Cell Biol.* 2016;26(7):476-485. DOI 10.1016/j.tcb.2016.03.009

Gálvez A., Abriouel H., Benomar N., Lucas R. Microbial antagonists to food-borne pathogens and biocontrol. *Curr. Opin. Biotechnol.* 2010; 21(2):142-148. DOI 10.1016/j.copbio.2010.01.005

González-Benítez N., Martín-Rodríguez I., Cuesta I., Arrayás M., White J.F., Molina M.C. Endophytic microbes are tools to increase tolerance in *Jasione* plants against arsenic stress. *Front. Microbiol.* 2021;12:664271. DOI 10.3389/fmicb.2021.664271

Goris J., Konstantinidis K.T., Klappenbach J.A., Coenye T., Vandamme P., Tiedje J.M. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 2007;57(1):81-91. DOI 10.1099/ijs.0.64483-0

Gundlapally S.R., Ara S., Sisinthy S. Draft genome of *Kocuria polaris* CMS 76or^T isolated from cyanobacterial mats, McMurdo Dry Valley, Antarctica: an insight into CspA family of proteins from *Kocuria polaris* CMS 76or^T. *Arch. Microbiol.* 2015;197(8):1019-1026. DOI 10.1007/s00203-015-1138-8

Hansda A., Kumar V., Anshumali. Cu-resistant *Kocuria* sp. CRB15: a potential PGPR isolated from the dry tailing of Rakha copper mine. *3 Biotech.* 2017;7(2):132. DOI 10.1007/s13205-017-0757-y

Hu J., Maamar S.B., Glawe A.J., Gottel N., Gilbert J.A., Hartmann E.M. Impacts of indoor surface finishes on bacterial viability. *Indoor Air.* 2019;29(4):551-562. DOI 10.1111/ina.12558

Jain C., Rodriguez-R L.M., Phillippy A.M., Konstantinidis K.T., Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 2018;9(1):5114. DOI 10.1038/s41467-018-07641-9

Kandi V., Palange P., Vaish R., Bhatti A.B., Kale V., Kandi M.R., Bhoomagiri M.R. Emerging bacterial infection: identification and clinical significance of *Kocuria* species. *Cureus.* 2016;8(8):e731. DOI 10.7759/cureus.731

Kargapolova K.Yu., Tkachenko O.V., Burygin G.L. Evaluation of rhizospheric bacterial isolates for their ability to promote potato growth *in vitro* and *ex vitro*. In: Collection of articles from the Int. Sci. and Pract. Conf. dedicated to the 130th anniversary of the birth of Academician N.I. Vavilov "Vavilov Readings – 2017". Saratov: Saratov State Vavilov Agrarian University, 2017;127-128 (in Russian)

Kaur H., Dolma K., Kaur N., Malhotra A., Kumar N., Dixit P., Sharma D., Mayilraj S., Choudhury A.R. Marine microbe as nano-factories for copper biomineralization. *Biotechnol. Bioprocess Eng.* 2015; 20:51-57. DOI 10.1007/s12257-014-0432-7

Koonin E.V. The Logic of Chance: The Nature and Origin of Biological Evolution. New Jersey: Pearson Education, Inc., 2012

Kumar C.G., Sujitha P. Green synthesis of Kocuran-functionalized silver glyconanoparticles for use as antibiofilm coatings on silicone urethral catheters. *Nanotechnology.* 2014;25(32):325101. DOI 10.1088/0957-4484/25/32/325101

Lee I., Kim Y.O., Park S.-C., Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* 2016;66(2):1100-1103. DOI 10.1099/ijsem.0.000760

Li X., Sun P., Zhang Y., Jin C., Guan C. A novel PGPR strain *Kocuria rhizophila* Y1 enhances salt stress tolerance in maize by regulating phytohormone levels, nutrient acquisition, redox potential, ion homeostasis, photosynthetic capacity and stress-responsive genes expression. *Environ. Exp. Bot.* 2020;174:104023. DOI 10.1016/j.envexpbot.2020.104023

Luo C., Rodriguez-R L.M., Konstantinidis K.T. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Res.* 2014;42(8):e73. DOI 10.1093/nar/gku169

Markov A.V., Kulikov A.M. Origin of Eukaryota: conclusions based on the analysis of protein homologies in the three superkingdoms. *Paleontol. J.* 2005;39(4):345-357

Mawang C.-I., Azman A.-S., Fuad A.-S.M., Ahamad M. Actinobacteria: an eco-friendly and promising technology for the bioaugmentation of contaminants. *Biotechnol. Rep.* 2021;32:e00679. DOI 10.1016/j.btre.2021.e00679

Medlar A.J., Törönen P., Holm L. AAI-profiler: fast proteome-wide exploratory analysis reveals taxonomic identity, misclassification and contamination. *Nucleic Acids Res.* 2018;46(W1):W479-W485. DOI 10.1093/nar/gky359

Mukherjee G., Saha C., Naskar N., Mukherjee A., Mukherjee A., Lahiri S., Majumder A.L., Seal A. An endophytic bacterial consortium modulates multiple strategies to improve arsenic phytoremediation efficacy in *Solanum nigrum*. *Sci. Rep.* 2018;8(1):6979. DOI 10.1038/s41598-018-25306-x

Nikitin M. The Origin of Life: From Nebula to Cell. Moscow, 2016 (in Russian)

Odeberg G., Bläckberg A., Sunnerhagen T. Infection or contamination with *Rothia*, *Kocuria*, *Arthrobacter* and *Pseudoglutamicibacter* – a retrospective observational study of non-*Micrococcus Micrococcaceae* in the clinic. *J. Clin. Microbiol.* 2023;61(4):e0148422. DOI 10.1128/jcm.01484-22

Oren A., Garrity G.M. Then and now: a systematic review of the systematics of prokaryotes in the last 80 years. *Antonie Van Leeuwenhoek.* 2014;106(1):43-56. DOI 10.1007/s10482-013-0084-1

Parks D.H., Chuvochina M., Waite D.W., Rinke C., Skarshewski A., Chaumeil P.-A., Hugenholtz P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 2018;36(10):996-1004. DOI 10.1038/nbt.4229

Passari A.K., Mishra V.K., Singh G., Singh P., Kumar B., Gupta V.K., Sarma R.K., Saikia R., O'Donovan A., Singh B.P. Insights into the functionality of endophytic actinobacteria with a focus on their biosynthetic potential and secondary metabolites production. *Sci. Rep.* 2017;7(1):11809. DOI 10.1038/s41598-017-12235-4

Potanina P.A., Safronova V.I., Belimov A.A., Kargapolova K.Yu., Tkachenko O.V., Burygin G.L. Species identification of plant-growth-promoting bacteria isolated from potato roots in Saratov Oblast. In: Collection of articles from the Int. Sci. and Pract. Conf. dedicated to the 130th anniversary of the birth of Academician N.I. Vavilov "Vavilov Readings – 2017". Saratov: Saratov State Vavilov Agrarian University, 2017;149-154 (in Russian)

Provorov N.A., Tikhonovich I.A., Vorob'ev N.I. Symbiosis and Symbiogenesis. St. Petersburg, 2018 (in Russian)

Rao H.C.Y., Rakshith D., Satish S. Antimicrobial properties of endophytic actinomycetes isolated from *Combretum latifolium* Blume, a medicinal shrub from Western Ghats of India. *Front. Biol.* 2015; 10:528-536. DOI 10.1007/s11515-015-1377-8

Reimer L.C., Carbasse J.S., Koblitz J., Ebeling C., Podstawka A., Overmann J. BacDive in 2022: the knowledge base for standardized bacterial and archaeal data. *Nucleic Acids Res.* 2022;50(D1):D741-D746. DOI 10.1093/nar/gkab961

Rodriguez-R L.M., Konstantinidis K.T. Bypassing cultivation to identify bacterial species. *Microbe.* 2014;9(3):111-118. DOI 10.1128/microbe.9.111.1

Román-Ponce B., Wang D., Soledad Vásquez-Murrieta M., Chen W.F., Estrada-de Los Santos P., Sui X.H., Wang E.T. *Kocuria arsenatis* sp. nov., an arsenic-resistant endophytic actinobacterium associated with *Prosopis laegivata* grown on high-arsenic-polluted mine tailing. *Int. J. Syst. Evol. Microbiol.* 2016;66(2):1027-1033. DOI 10.1099/ijsem.0.000830

Shchyogolev S.Yu. A review of the textbook "Biodiversity and Systematics of Microorganisms", by I.B. Ivshina, A.V. Krivoruchko, M.S. Kuyukina. *Microbiology.* 2021;90(2):247-249. DOI 10.31857/S0026365621010079 (in Russian)

Stackebrandt E., Schumann P. *Kocuria*. In: Bergey's Manual of Systematics of Archaea and Bacteria. John Wiley & Sons, 2015. DOI 10.1002/9781118960608.gbm00120

Stairs C.W., Ettema T.J.G. The archaeal roots of the eukaryotic dynamic actin cytoskeleton. *Curr. Biol.* 2020;30(10):R521-R526. DOI 10.1016/j.cub.2020.02.074

Tavarideh F., Pourahmad F., Nemati M. Diversity and antibacterial activity of endophytic bacteria associated with medicinal plant, *Scrophularia striata*. *Vet. Res. Forum.* 2022;13(3):409-415. DOI 10.30466/vrf.2021.529714.3174

Tedsree N., Likhitwitayawuid K., Sritularak B., Tanasupawat S. Diversity and antimicrobial activity of plant growth promoting endophytic actinomycetes isolated from Thai orchids. *Environ. Nat. Resour. J.* 2022;20(4):379-392. DOI 10.32526/ennrj/20/202200039

Tettelin H., Medini D. (Eds.) The Pangenome. Diversity, Dynamics and Evolution of Genomes. Cham: Springer, 2020. DOI 10.1007/978-3-030-38281-0

Trachtenberg A.M., Goen A.E., MacLea K.S. Genome sequences for three strains of *Kocuria rosea*, including the type strain. *Genome Announc.* 2018;6(25):e00594-18. DOI 10.1128/genomeA.00594-18

Treangen T.J., Rocha E.P.C. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 2011;7(1):e1001284. DOI 10.1371/journal.pgen.1001284

Vital T.Z., Román-Ponce B., Orduña F.N.R., Estrada-de Los Santos P., Vásquez-Murrieta M.S., Deng Y., Yuan H.L., Wang E.T. An endophytic *Kocuria palustris* strain harboring multiple arsenate reductase genes. *Arch. Microbiol.* 2019;201(9):1285-1293. DOI 10.1007/s00203-019-01692-2

Yastrebova O.V., Plotnikova E.G. Phylogenetic diversity of bacteria of the family *Micrococcaceae* isolated from biotopes with different anthropogenic impact. *Bulletin of Perm University. Biology.* 2020;4:321-333. DOI 10.17072/1994-9952-2020-4-321-333 (in Russian)

Lake Baikal amphipods and their genomes, great and small

P.B. Drozdova [1, 2] ✉, E.V. Madyarova [1], A.N. Gurkov [1, 2], A.E. Saranchina [1], E.V. Romanova [3],
J.V. Petunina [3], T.E. Peretolchina [3], D.Y. Sherbakov [1, 3, 4], M.A. Timofeyev [1] ✉

[1] Irkutsk State University, Irkutsk, Russia
[2] Baikal Research Centre, Irkutsk, Russia
[3] Limnological Institute of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia
[4] Novosibirsk State University, Novosibirsk, Russia

✉ drozdovapb@gmail.com; m.a.timofeyev@gmail.com

Abstract. Endemic amphipods (Crustacea: Amphipoda) of Lake Baikal represent an outstanding example of large species flocks occupying a wide range of ecological niches and originating from a handful of ancestor species. Their development took place at a restricted territory and is thus open for comprehensive research. Such examples provide unique opportunities for studying behavioral, anatomic, or physiological adaptations in multiple combinations of environmental conditions and thus attract considerable attention. The existing taxonomies of this group list over 350 species and subspecies, which, according to the molecular phylogenetic studies of marker genes, full transcriptomes and mitochondrial genomes, originated from at least two introductions into the lake. The studies of allozymes and marker genes have revealed a significant cryptic diversity in Baikal amphipods, as well as a large variance in genetic diversity within some morphological species. Crossing experiments conducted so far for two morphological species suggest that the differences in the mitochondrial marker (cytochrome c oxidase subunit I gene) can potentially be applied for making predictions about reproductive isolation. For about one-tenth of the Baikal amphipod species, nuclear genome sizes and chromosome numbers are known. While genome sizes vary within one order of magnitude, the karyotypes are relatively stable ($2n = 52$ for most species studied). Moreover, analysis of the diversity of repeated sequences in nuclear genomes showed significant between-species differences. Studies of mitochondrial genomes revealed some unusual features, such as variation in length and gene order, as well as duplications of tRNA genes, some of which also underwent remolding (change in anticodon specificity due to point mutations). The next important steps should be (i) the assembly of whole genomes for different species of Baikal amphipods, which is at the moment hampered by complicated genome structures with high repeat content, and (ii) updating species taxonomy taking into account all the data.

Key words: Lake Baikal; amphipods; species flocks; speciation; population genetics; genomics.

Байкальские амфиподы и их геномы, большие и малые

П.Б. Дроздова [1, 2] ✉, Е.В. Мадьярова [1], А.Н. Гурков [1, 2], А.Е. Саранчина [1], Е.В. Романова [3],
Ж.В. Петунина [3], Т.Е. Перетолчина [3], Д.Ю. Щербаков [1, 3, 4], М.А. Тимофеев [1] ✉

[1] Иркутский государственный университет, Иркутск, Россия
[2] АНО «Байкальский исследовательский центр», Иркутск, Россия
[3] Лимнологический институт Сибирского отделения Российской академии наук, Иркутск, Россия
[4] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ drozdovapb@gmail.com; m.a.timofeyev@gmail.com

Аннотация. Эндемичные амфиподы (Crustacea: Amphipoda) озера Байкал – это один из наиболее ярких примеров возникновения большого количества видов (так называемых букетов видов), занимающих разнообразные экологические ниши, от небольшого числа исходных видов, которое происходило на ограниченной территории и потому доступно для всестороннего исследования. Подобные примеры предоставляют уникальные возможности изучения поведенческих, анатомических и физиологических адаптаций во множестве комбинаций условий среды и потому привлекают большое внимание. Существующие варианты таксономической классификации этой группы насчитывают более 350 морфологических видов и подвидов, которые,

согласно молекулярно-филогенетическим исследованиям маркерных генов, полных транскриптомов и митохондриальных геномов, произошли в результате не менее двух вселений в озеро. Исследования изо-ферментов и маркерных генов выявили существенное криптическое разнообразие байкальских амфипод, а также существенный разброс по уровню генетического разнообразия внутри некоторых морфологических видов. Экспериментальная проверка, проведенная на данный момент только для двух морфологических видов, показывает возможную применимость митохондриального маркера, гена первой субъединицы цитохром c-оксидазы, для предсказания репродуктивной изоляции. Приблизительно у десятой части видов байкальских амфипод был изучен размер ядерного генома и хромосомные числа, что позволило выявить почти десятикратную вариабельность размера генома при стабильных (2n = 52 для большинства изученных видов) кариотипах. При этом анализ разнообразия повторов в ядерных геномах показал существенные меж-видовые различия. Кроме того, выявлены необычные особенности некоторых митохондриальных геномов, такие как вариабельность по длине и по порядку генов, а также дупликации генов тРНК, часть из которых подверглась ремолдингу (изменению специфичности антикодона за счет точечных мутаций). Следующими важными шагами должны стать сборка полных геномов для разных видов байкальских амфипод, чему на данном этапе препятствует сложная структура этих геномов с большим содержанием повторов, и обновление таксономической классификации видов с учетом комплекса полученных данных.

Ключевые слова: озеро Байкал; бокоплавы; букеты видов; видообразование; генетика популяций; геномика.

Introduction

Ancient lakes are known speciation hotspots. However, even against this background, the biodiversity of Lake Baikal, the age of which is estimated as 25–30 or even 70 million years, stands out (Cristescu et al., 2010; Mats et al., 2011). The representatives of the order Amphipoda (Crustacea) constitute one of the largest groups of closely related species found in Baikal.

The diversity of amphipods in Baikal may be partially attributed to the broad range of habitats and ecological niches they occupy, as the species within this group differ in habitat depth (0–1,642 meters), feeding habits, and reproductive periods (Takhteev, 2000a, b). However, many species share the same habitat, being at the same time similar in size, feeding spectra and reproductive periods (Takhteev, 2000a, b), which raises the question of the evolutionary forces that drove their speciation. Earlier reviews have already presented global conclusions about the origin of Baikal endemic fauna based on molecular data from multiple studies (Sherbakov, 1999; Sherbakov et al., 2017). However, the recent years have seen the accumulation of a lot of new data, especially high-throughput sequencing, which have uncovered new details on speciation and genome evolution in Baikal amphipods.

How many amphipod species are there in Baikal?

Morphological classification

Currently, the formal identification of Baikal amphipod species is based on the morphological criterium, i. e. the presence of a unique set of morphological traits in all studied individuals of a particular species. The number of morphological species and subspecies in Baikal exceeds 350 (Takhteev, 2000a; Kamaltynov, 2001; Takhteev et al., 2015). In the case of Baikal amphipods, subspecies were mostly derived from morphological varieties that differed less than species would (Bazikalova, 1945; Takhteev, 2000a). All these species belong to the phylum Arthropoda, subphylum Crustacea, class Malacostraca, order Amphipoda, and superfamily Gammaroidea (Sket et al., 2019). The numbers of subspecies, species, genera and families differ according to different authors (Takhteev, 2019), but the most evident discrepancies are attributed to differing taxonomic levels (subspecies/species, congeneric species/ different genera etc.).

Multiple classifications complicate studies in Baikal amphipods. From a practical point of view, the most important discrepancies for researchers are different generic names for the same species. The correspondence between the names suggested by different authors can be easily checked using the World Amphipoda Database (WAD; https://www.marine species.org/amphipoda/) (Horton et al., 2023). It is worth noting that the systematics accepted by WAD (Kamaltynov, 2001, 2009) does not have an associated identification key, and thus many manuscripts use the species names indicated in the existing keys. The most comprehensive key for Baikal amphipods is still (Bazikalova, 1945), although some groups are covered in more detail in later sources (Bazikalova, 1962; Takhteev, 2000a). The only available English identification key for the genera of Baikal amphipods is provided by (Sket et al., 2019). An English language checklist of all known species according to the same classification is compiled in (Takhteev et al., 2015). However, none of the sources include the species described after 2000: *Eulimnogammarus messerschmidtii* Bedulina et Takhteev, 2014 (Bedulina et al., 2014), *Eulimnogammarus etingovae* and *Eulimnogammarus tchernykhi* Moskalenko, Neretina & Yampolsky, 2020 (Moskalenko et al., 2020).

Molecular genetics approaches to classification

Molecular phylogenetic studies in Baikal amphipods revealed three important conclusions. First, all studied species cluster within the freshwater radiation of the morphological genus *Gammarus* Fabricius, 1775 at the phylogenetic tree, which provides evidence of their descent from *Gammarus*-like freshwater ancestors (Macdonald III et al., 2005; Hou et al., 2014). Second, studies utilizing phylogenetic marker genes have shown that Baikal amphipods fall into two clades (Sherbakov, 1999; Macdonald III et al., 2005), indicating that their ancestors invaded the lake at least twice. This conclusion is supported by the phylogeny based on single-copy orthologs in transcriptomes (Naumenko et al., 2017) and whole

П.Б. Дроздова, Е.В. Мадьярова, А.Н. Гурков …
Т.Е. Перетолчина, Д.Ю. Щербаков, М.А. Тимофеев

Байкальские амфиподы
и их геномы, большие и малые

2024
28•3
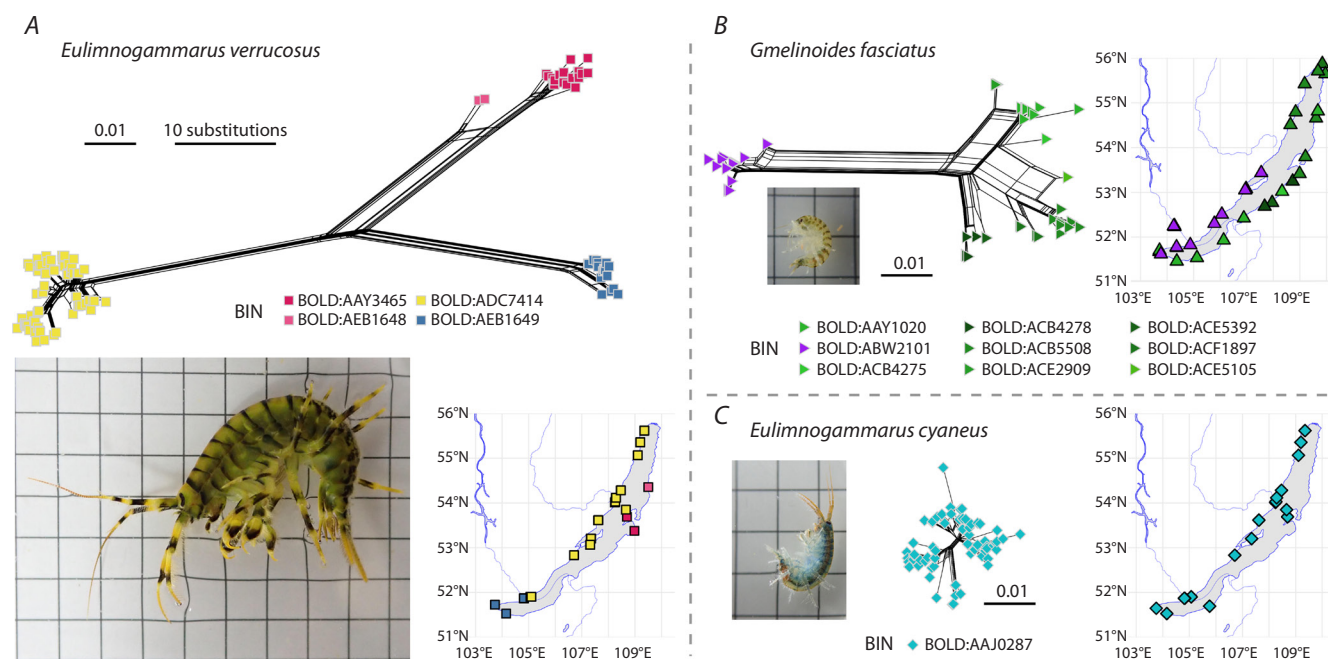
Fig. 1. Comparison of the levels of population genetic diversity of the *COI* fragment within the best studied morphological species *E. verrucosus* (*A*), *Gm. fasciatus* (*B*) and *E. cyaneus* (*C*).

Shown are representative photographs of each species at the same scale (grid size is 5 mm), along with split phylogenetic networks at the same scale (scale bar is 1 % substitutions, i.e. 5.1 substitutions in the 510-bp alignment), and corresponding sampling points. Sequence data were obtained from the BOLD database (Ratnasingham, Hebert, 2007). Sampling coordinates were added or corrected based on the original publications (Fazalova et al., 2010; Petunina, 2015; Romanova et al., 2016a; Gurkov et al., 2019). Different colors on networks and maps correspond to different barcode index numbers (BINs) automatically determined by BOLD (Ratnasingham, Hebert, 2013). For detailed methodology, please refer to https://github.com/drozdovapb/Baikal-amphipods-review-post-chr2023.

mitochondrial genomes (Romanova et al., 2016a). The first invasion gave rise to a much smaller number of recent species than the second invasion (Bazikalova, 1945; Naumenko et al., 2017). Third, several species of Baikal amphipods were found to exhibit cryptic diversity, i.e. the presence of genetically distinct groups that are morphologically indistinguishable or hard to distinguish.

Studies of allozyme spectra showed significant (in many cases species-level) differences within morphological species and led to suggestions to elevate some subspecies to species rank (Yampolsky et al., 1994; Väinölä, Kamaltynov, 1999) or, *vice versa*, synonymize (Daneliya et al., 2009). The differences in allozyme frequencies may indicate the presence of isolated populations, but they are difficult to directly translate into species boundaries. This issue also affects the outcomes of phylogenetic marker sequencing, albeit to a lesser degree. In this case, species delimitation may rely on calculated threshold values of patristic distances (Lefébure et al., 2006) or other techniques that take into account genetic distances, phylogenetic tree topology or shared alleles (Fišer et al., 2018). However, the obtained sample clusters could not be safely assigned to biological species. Therefore, they are termed molecular operational taxonomic units (MOTUs) (Blaxter, 2004).

Folmer fragment of the cytochrome c oxidase subunit I gene (*COI* or *cox1*) is the most well-known and frequently used marker sequence for amphipods and many other invertebrates (Folmer et al., 1994; Hebert et al., 2003). It is important to note that mitochondrial and nuclear-based phylogenies often

produce conflicting results, which is known as mito-nuclear discordance (Toews, Brelsford, 2012). In order to draw reliable conclusions about separated genetic lineages, which would indicate reproductively isolated species, it is recommended to also employ nuclear markers. Popular nuclear markers include rRNA gene clusters as well as whole-genome markers such as ultraconserved elements (UCEs), restriction site-associated DNA (RADs), and single-copy orthologs (SCOs) (Eberle et al., 2020). From this list, SCOs have already been utilized to study Baikal amphipods (Naumenko et al., 2017; Drozdova et al., 2021); for other amphipods, RADs have also been used (Jordan et al., 2020; Weston et al., 2022; Eme et al., 2023).

Population genetic diversity

In total, intraspecies diversity has been studied using different methods and with varying geographical coverage for over 20 morphological species of Baikal amphipods (Supplementary Material 1)[1]. Some of these species showed substantial intraspecific diversity (Gomanenko et al., 2005; Daneliya et al., 2011; Gurkov et al., 2019). It is noteworthy that even species with comparable distribution and ecological characteristics can exhibit dramatic differences in the level of intraspecific diversity (Fig. 1). For example, it was found that the species *Eulimnogammarus verrucosus* (Gerstfeldt, 1858), common in the littoral zone, is actually composed of at least three genetic lineages, inhabiting the western (up to the source of the Angara

---

[1] Supplementary Materials 1–5 are available at:
https://vavilovj-icg.ru/download/pict-2024-28/appx13.xlsx

river), southern and eastern parts of the Baikal shore (W, S, and E), respectively. Intraspecific pairwise differences in *COI* sequences reached 13 %, which is similar to the distances between morphological species (Gurkov et al., 2019). The most recent common ancestor of these lineages, according to a molecular clock-based estimate, existed around 4.5 million years ago (Drozdova et al., 2022). A nuclear marker, 18S rRNA gene fragment, fully corroborated this division (Gurkov et al., 2019).

Gmelinoides fasciatus (Stebbing, 1899) is another species common in the shallow water. It is also divided into genetic lineages correlated with geography, but here the differences are less pronounced, reaching about 8 % (Gomanenko et al., 2005), and the last common ancestor existed around 2 million years ago (Bukin et al., 2018). A nuclear marker, intron of the ATP synthase β subunit gene, showed a lower genetic diversity but also supported intraspecific differentiation (Kovalenkova, 2018). In contrast, preliminary data on the only pelagic planktonic species of Baikal amphipods, *Macrohectopus branickii* (Dybowsky, 1874), based on the fragments of the mitochondrial genes *COI* and NADH dehydrogenase fifth subunit (*ND5* or *nad5*) (Petunina et al., 2023; Zaidykov et al., 2023) did not reveal geographically separated genetic lineages.

Finally, *Eulimnogammarus cyaneus* (Dybowsky, 1874), another widely distributed species inhabiting a significant part of the Lake Baikal littoral, exhibits very weak genetic differentiation based on the *COI* fragment (Gurkov et al., 2019) but much more pronounced differentiation according to allozyme data (Mashiko et al., 2000). Furthermore, it is important to note that the borders between genetic lineages of *E. verrucosus*, such as the Angara river outflow, do not hold for *Gm. fasciatus* (Fig. 1, *A*, *B*); the geographic barriers for *Gm. fasciatus* are unclear. The source of the Angara river started to form at most 120,000 years ago (Arzhannikov et al., 2018), thus being much younger that the last common ancestor of *E. verrucosus* populations dwelling at different sides of the outflow (3.81 million years ago) (Drozdova et al., 2022). The current cryptic diversity within *E. verrucosus* and *Gm. fasciatus* appears to reflect past distribution barriers, such as dwelling in refugia during non-favorable climatic conditions (Bukin et al., 2018).

Reproductive barriers and cryptic species
Reproductive isolation is crucial for biologically sensible species delimitation. However, this issue has just recently started to be explored for Baikal amphipods. To date, experimental checks for reproductive incompatibility have only been carried out for two widely distributed littoral species, *E. verrucosus* and *E. cyaneus*. Crossing experiments were conducted with representatives of populations from Listvyanka (W) and Port Baikal (S) for both species (these populations were chosen due to the closest geographic proximity of different genetic lineages), and also from Ust-Bargusin (E) for *E. verrucosus*. In the case of *E. verrucosus*, both prezygotic and postzygotic reproductive barriers were found. Although these barriers are not absolute, their combination can ensure reproductive isolation when different lineages are mixed. In the case of *E. cyaneus*, the analysis of representatives of the populations

separated by the Angara river outflow did not show any prezygotic or postzygotic barriers. Mate choice was random, and upon crossing, at least the first generation hybrids developed normally (Drozdova et al., 2022, 2023). Therefore, in the case of *E. verrucosus* and *E. cyaneus*, differences in *COI* sequences indeed correlate with the presence of reproductive barriers. However, it would be premature to establish a general rule for Baikal amphipods based solely on these findings. It is necessary to conduct such experiments for other genera to draw comprehensive conclusions. Further research on reproductive barriers, as well as genomes and gene expression, may aid in comprehending the factors that contribute to reproductive incompatibility and thus serve as the genetic basis of speciation.

The next steps that need to be undertaken are renewal of the Baikal amphipod taxonomy and species redescription taking into account biological reality and possible competition between cryptic species. This necessity is not unique to Baikal, as cryptic species complexes without formal species descriptions are also characteristic of many other amphipods, including popular ecotoxicological models *Gammarus fossarum* and *Hyalella azteca* (Jourdan et al., 2023). However, it underlines the critical importance of always specifying the particular sampling place for Baikal amphipods in every publication and identifying the genetic lineage whenever possible.

What is known about genomes of Lake Baikal amphipods?
The genetics of Baikal amphipods is a relatively understudied area, with most of the research focusing on individual genetic markers. Nuclear genome sizes have been estimated using cytogenetic methods such as Feulgen image analysis densitometry (FIAD) and flow cytometry (FCM) for 36 morphological species (Jeffery et al., 2017; Drozdova et al., 2022). Karyotypes have been studied for 35 morphological species (Salemaa, Kamaltynov, 1994; Kamaltynov, 2001; Natyaganova, Sitnikova, 2012; Barabanova et al., 2019) (Supplementary Material 2). Transcriptome sequencing data are available for over 60 morphological species (Naumenko et al., 2017; Drozdova et al., 2022), enabling the extraction of most protein-coding gene sequences, as well as partial or complete mitochondrial genomes. These transcriptome assemblies are particularly valuable for proteomic studies (Bedulina et al., 2021; Zolotovskaya et al., 2021). Genome DNA sequencing data are available for seven species, which enabled the assembly of mitochondrial genomes and can be used to evaluate the diversity of repeated sequences in nuclear genomes (Rivarola-Duarte et al., 2014; Romanova et al., 2016a, 2021; Rivarola-Duarte, 2021; Yuxiang et al., 2023) (Supplementary Material 3).

Genome size variation and its possible causes
The genome sizes of the studied Baikal amphipods varied from 2 to 17 pg (1 pg is approximately 1 Gb) (Jeffery et al., 2017) (Fig. 2), which falls within the known range of amphipod genome sizes (Hultgren et al., 2018). For up-to-date information on accumulated data one can refer to the Animal

П.Б. Дроздова, Е.В. Мадьярова, А.Н. Гурков …
Т.Е. Перетолчина, Д.Ю. Щербаков, М.А. Тимофеев

Байкальские амфиподы
и их геномы, большие и малые

2024
28 • 3

Baikalogammarus pullus
Size: ≤8 mm; depths: 0.5–25 m
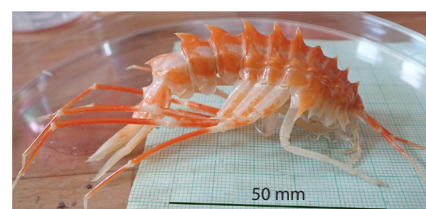
Brachyuropus grewingkii
Size: ≤130 mm; depths: 140–1,300 m
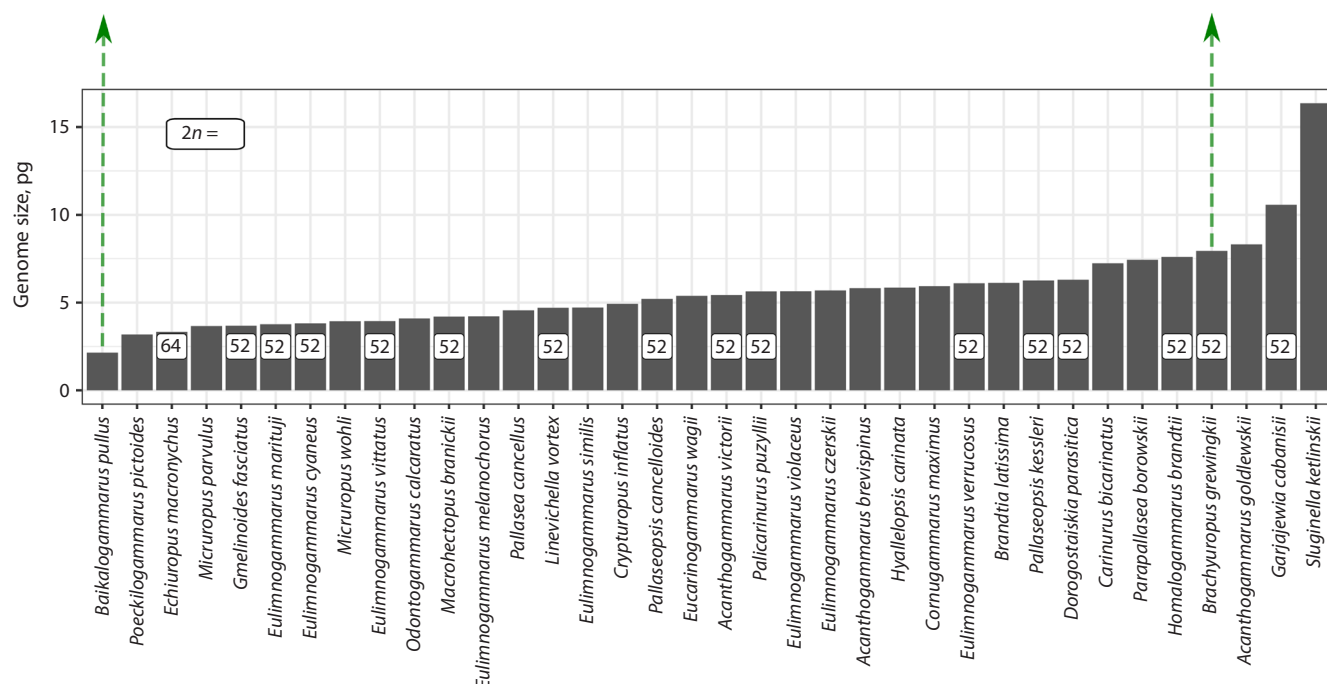
Fig. 2. Nuclear genome sizes of Baikal amphipods, as estimated with FIAD (Jeffery et al., 2017), and their chromosome numbers (Salemaa, Kamaltynov, 1994; Kamaltynov, 2001; Natyaganova, Sitnikova, 2012).

Please refer to Supplementary Material 2 for full data set. Species names are given according to (Jeffery et al., 2017). The photographs show *Baikalogammarus pullus* (Dybowsky, 1874), which has the smallest genome and small body length, and dwells in the littoral and sublittoral zones, and *Brachyuropus grewingkii* (Dybowsky, 1874), which is a deep-water species and one of the largest. The ecological characteristics of these species are given according to (Kamaltynov, 2001). The photo of *B. grewingkii* was generously provided by Ekaterina Shchapova.

Genome Size Database (http://www.genomesize.com/) (Gregory et al., 2007).

When comparing data obtained using different methods, it is worth keeping in mind that crustacean genome size estimates obtained with FIAD are typically slightly lower than those obtained with FCM (Wyngaard et al., 2022). Notably, genome size differences accumulate quite rapidly, as evidenced by the differing genome sizes of *E. verrucosus* lineages (6.1 pg for the E, 6.9 pg for the W, and 8.0 pg for the S lineage) (Drozdova et al., 2022). The analysis of genome sizes in different species showed a weak positive correlation with both maximal body length and habitat depths, which corresponds to the known ecological trends (Jeffery et al., 2017). However, chromosome numbers were found to be identical ($2n = 52$) for 33 out of 35 studied species (Salemaa, Kamaltynov, 1994; Kamaltynov, 2001; Natyaganova, Sitnikova, 2012) (Fig. 2), which corresponds to the modal chromosome number for gammaroid amphipods (Coleman, 1994). The lack of correlation between chromosome numbers and genome sizes suggests that repeated

sequences significantly contribute to this variation. Analysis of the diversity of repeated sequences revealed significant differences between species of Baikal amphipods (Yuxiang et al., 2023). In all studied species, the proportion of reads included in repeat clusters exceeded 50 % (Rivarola-Duarte et al., 2014; Yuxiang et al., 2023).

Mitochondrial genomes

The mitochondrial genome is the most extensively studied part of the genome in Baikal amphipods. It is a small, high-copy DNA molecule, and its sequence is generally easy to assemble from low-coverage genome-wide sequencing (Smith, 2016). Animal mitochondrial genomes are typically circular with a length of about 16 kb and contain 13 protein-coding genes, 2 rRNA genes and 22 tRNA genes. However, significant differences in genome architecture, size, and composition are known (Lavrov, Pett, 2016).

At the moment, eight complete and six partial mitochondrial genomes have been published for Baikal amphipods

(Rivarola-Duarte et al., 2014; Romanova et al., 2016a–c, 2021; Mamos et al., 2021) (Supplementary Material 4). Most of these assemblies are within 15–18 kb in length, but the mitochondrial genome of *M. branickii* is over 42 kb-long, making it one of the largest known animal mitochondrial genomes (Romanova et al., 2021). Furthermore, mitochondrial genomes of some Baikal amphipods exhibit gene order rearrangements, gene duplications and the phenomenon of tRNA gene remolding, i. e. changes in tRNA specificity due to a mutation in the anticodon sequence. Remolding is not unique for Baikal amphipods but occurs with higher frequency than in other amphipods (Romanova et al., 2020).

Perspectives in whole-genome studies

The next important step in the development of genome-wide studies of Baikal amphipods should be the assembly of whole nuclear genomes for a number of species. For the world amphipod fauna, seven genome assemblies are mentioned in the literature (Supplementary Material 5). Four of them (*H. azteca*, *Trinorchestia longiramus*, *Platorchestia hallaensis*, and *Parhyale hawaensis*) belong to the infraorder Talitrida (Kao et al., 2016; Poynton et al., 2018; Patra et al., 2020, 2021). Three species belong to the infraorder Gammarida (*Gammarus lacustris*, *G. roeselii*, and *E. verrucosus*). One of these species, *E. verrucosus*, inhabits Baikal (Jin et al., 2019; Cormier et al., 2021; Rivarola-Duarte, 2021). The genomes of gammarids are the largest within this list. Not surprisingly, creation of a high-quality assembly of these genomes is complicated and currently at the draft stage, with N50 of all assemblies being below 5 kb, and only the genome of *G. roeselii* being publicly available.

The development of third-generation genome sequencing techniques provides hope that technical difficulties in assembly of complex gammarid genomes can be overcome. For example, the assembly of the Antarctic krill, *Euphausia superba*, genome, which with 48 Gb is the largest assembled animal genome to date, demonstrates the potential of this technology (Shao et al., 2023). High-quality genome assemblies will greatly enhance the research on the adaptation mechanisms of endemic amphipods to various conditions in Lake Baikal and tracing their evolutionary history. This will be due to a wider range of possibilities for retrieving full gene sets (which is impossible with the current transcriptomic data) and regulatory elements, as well as new data on population history (Bourgeois, Warren, 2021) and higher resolution for phylogenetic analysis.

References

Arzhannikov S.G., Ivanov A.V., Arzhannikova A.V., Demonterova E.I., Jansen J.D., Preusser F., Kamenetsky V.S., Kamenetsky M.B. Catastrophic events in the Quaternary outflow history of Lake Baikal. *Earth-Sci. Rev.* 2018;177:76-113. DOI 10.1016/j.earscirev.2017.11.011

Barabanova L., Galkina S., Mikhailova E. Cytogenetic study on the invasive species *Gmelinoides fasciatus* in the ecosystem of the Gulf of Finland. *J. Mar. Biol. Assoc. UK.* 2019;99(3):611-618. DOI 10.1017/S0025315417001357

Bazikalova A.Y. Amphipods of Lake Baikal. *Trudy Baykal'skoy Limnologicheskoy Stantsii = Proceedings of the Baikal Limnological Station.* 1945;11:1-440 (in Russian)

Bazikalova A.Y. Taxonomy, ecology, and distribution of the genera *Micruropus* Stebbing and *Pseudomicruropus* nov. gen. (Amphipoda, Gammaridea). Systematics and ecology of crustaceans of Baikal. *Trudy Limnologicheskogo Instituta = Proceedings of the Limnological Institute.* 1962;2(22):3-140 (in Russian)

Bedulina D.S., Takhteev V.V., Pogrebnyak S.G., Govorukhina E.B., Madyarova E.V., Lubyaga Y.A., Vereshchagina K.P., Timofeyev M.A., Luckenbach T. On *Eulimnogammarus messerschmidtii*, sp. n. (Amphipoda: Gammaridea) from Lake Baikal, Siberia, with redescription of *E. cyanoides* (Sowinsky) and remarks on taxonomy of the genus *Eulimnogammarus*. *Zootaxa*. 2014;3838(5):518-544. DOI 10.11646/zootaxa.3838.5.2

Bedulina D., Drozdova P., Gurkov A., von Bergen M., Stadler P.F., Luckenbach T., Timofeyev M., Kalkhof S. Proteomics reveals sex-specific heat shock response of Baikal amphipod *Eulimnogammarus cyaneus*. *Sci. Total Environ.* 2021;763:143008. DOI 10.1016/j.scitotenv.2020.143008

Blaxter M.L. The promise of a DNA taxonomy. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2004;359(1444):669-679. DOI 10.1098/rstb.2003.1447

Bourgeois Y.X.C., Warren B.H. An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. *Mol. Ecol.* 2021;30(23):6036-6071. DOI 10.1111/mec.15989

Bukin Yu.S., Petunina J.V., Sherbakov D.Yu. The mechanisms for genetic diversity of Baikal endemic amphipod *Gmelinoides fasciatus*: relationships between the population processes and paleoclimatic history of the Lake. *Russ. J. Genet.* 2018;54(9):1059-1068. DOI 10.1134/S1022795418090053

Coleman Ch.O. Karyological studies in Amphipoda (Crustacea). *Ophelia*. 1994;39(2):93-105. DOI 10.1080/00785326.1994.10429537

Cormier A., Chebbi M.A., Giraud I., Wattier R., Teixeira M., Gilbert C., Rigaud T., Cordaux R. Comparative genomics of strictly vertically transmitted, feminizing Microsporidia endosymbionts of amphipod crustaceans. *Genome Biol. Evol.* 2021;13(1):evaa245. DOI 10.1093/gbe/evaa245

Cristescu M.E., Adamowicz S.J., Vaillant J.J., Haffner D.G. Ancient lakes revisited: from the ecology to the genetics of speciation. *Mol. Ecol.* 2010;19(22):4837-4851. DOI 10.1111/j.1365-294X.2010.04832.x

Daneliya M.E., Kamaltynov R.M., Kontula T., Väinölä R. Systematics of the Baikalian *Babr* (Crustacea: Amphipoda: Pallaseidae). *Zootaxa*. 2009;2276(1):49-68. DOI 10.11646/zootaxa.2276.1.3

Daneliya M.E., Kamaltynov R.M., Väinölä R. Phylogeography and systematics of *Acanthogammarus* s. str., giant amphipod crustaceans from Lake Baikal. *Zool. Scr.* 2011;40(6):623-637. DOI 10.1111/j.1463-6409.2011.00490.x

Drozdova P., Kizenko A., Saranchina A., Gurkov A., Firulyova M., Govorukhina E., Timofeyev M. The diversity of opsins in Lake Baikal amphipods (Amphipoda: Gammaridae). *BMC Ecol. Evol.* 2021;21(1):81. DOI 10.1186/s12862-021-01806-9

Drozdova P., Saranchina A., Madyarova E., Gurkov A., Timofeyev M. Experimental crossing confirms reproductive isolation between cryptic species within *Eulimnogammarus verrucosus* (Crustacea: Amphipoda) from Lake Baikal. *Int. J. Mol. Sci.* 2022;23(18):10858. DOI 10.3390/ijms231810858

Drozdova P.B., Saranchina A.E., Mutin A.D., Rzhechitskiy Ya.A., Gurkov A.N., Lipaeva P.V., Shatilina Zh.M., Timofeyev M.A. Geographic barriers and reproductive isolation in the formation of cripic species within the abundant representatives of Baikal endemic amphipods of the genus *Eulimnogammarus*. In: Proceedings of the IV All-Russia Conference "Development of Life on Earth in Abiotic

П.Б. Дроздова, Е.В. Мадьярова, А.Н. Гурков …
Т.Е. Перетолчина, Д.Ю. Щербаков, М.А. Тимофеев

Байкальские амфиподы
и их геномы, большие и малые

2024
28•3

Change Processes", 25–29 Sept. 2023, Listvyanka. Irkutsk, 2023; 70-73. DOI 10.24412/cl-34446-2023-4-70-73 (in Russian)

Eberle J., Ahrens D., Mayer C., Niehuis O., Misof B. A plea for standardized nuclear markers in Metazoan DNA taxonomy. *Trends Ecol. Evol.* 2020;35(4):336-345. DOI 10.1016/j.tree.2019.12.003

Eme D., Westfall K.M., Matthíasardóttir B., Kristjánsson B.K., Pálsson S. Contrasting phylogeographic patterns of mitochondrial and genome-wide variation in the groundwater amphipod *Crangonyx islandicus* that survived the Ice Age in Iceland. *Diversity*. 2023; 15(1):88. DOI 10.3390/d15010088

Fazalova V., Nevado B., Peretolchina T., Petunina J., Sherbakov D. When environmental changes do not cause geographic separation of fauna: differential responses of Baikalian invertebrates. *BMC Evol. Biol.* 2010;10(1):320. DOI 10.1186/1471-2148-10-320

Fišer C., Robinson C.T., Malard F. Cryptic species as a window into the paradigm shift of the species concept. *Mol. Ecol.* 2018;27(3):613-635. DOI 10.1111/mec.14486

Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 1994;3(5):294-299.

Gomanenko G.V., Kamaltynov R.M., Kuzmenkova Zh.V., Berenos K., Sherbakov D.Yu. Population structure of the Baikalian amphipod *Gmelinoides fasciatus* (Stebbing). *Russ. J. Genet.* 2005;41(8):907-912. DOI 10.1007/s11177-005-0179-5

Gregory T.R., Nicol J.A., Tamm H., Kullman B., Kullman K., Leitch I.J., Murray B.G., Kapraun D.F., Greilhuber J., Bennett M.D. Eukaryotic genome size databases. *Nucleic Acids Res.* 2007;35(Suppl.1):D332-D338. DOI 10.1093/nar/gkl828

Gurkov A., Rivarola-Duarte L., Bedulina D., Fernández Casas I., Michael H., Drozdova P., Nazarova A., Govorukhina E., Timofeyev M., Stadler P.F., Luckenbach T. Indication of ongoing amphipod speciation in Lake Baikal by genetic structures within endemic species. *BMC Evol. Biol.* 2019;19(1):138. DOI 10.1186/s12862-019-1470-8

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol. Sci.* 2003;270(1512):313-321. DOI 10.1098/rspb.2002.2218

Horton T., De Broyer C., Bellan-Santini D., Coleman C.O., Copilaș-Ciocianu D., Corbari L., Daneliya M.E., Dauvin J.-C., Decock W., Fanini L., Fišer C., Gasca R., Grabowski M., Guerra-García J.M., Hendrycks E.A., Hughes L.E., Jaume D., Kim Y.-H., King R.A., Lo Brutto S., Lörz A.-N., Mamos T., Serejo C.S., Senna A.R., Souza-Filho J.F., Tandberg A.H.S., Thurston M.H., Vader W., Väinölä R., Valls Domedel G., Vandepitte L., Vanhoorne B., Vonk R., White K.N., Zeidler W. The World Amphipoda Database: history and progress. *Rec. Aust. Mus.* 2023;75(4):329-342. DOI 10.3853/j.2201-4349.75.2023.1875

Hou Z., Sket B., Li S. Phylogenetic analyses of Gammaridae crustacean reveal different diversification patterns among sister lineages in the Tethyan region. *Cladistics*. 2014;30(4):352-365. DOI 10.1111/cla.12055

Hultgren K.M., Jeffery N.W., Moran A., Gregory T.R. Latitudinal variation in genome size in crustaceans. *Biol. J. Linn. Soc.* 2018;123(2): 348-359. DOI 10.1093/biolinnean/blx153

Jeffery N.W., Yampolsky L., Gregory T.R. Nuclear DNA content correlates with depth, body size, and diversification rate in amphipod crustaceans from ancient Lake Baikal, Russia. *Genome*. 2017;60(4): 303-309. DOI 10.1139/gen-2016-0128

Jin S., Bian C., Jiang S., Sun S., Xu L., Xiong Y., Qiao H., Zhang W., You X., Li J., Gong Y., Ma B., Shi Q., Fu H. Identification of candidate genes for the plateau adaptation of a Tibetan amphipod, *Gammarus lacustris*, through integration of genome and transcriptome sequencing. *Front. Genet.* 2019;10:53. DOI 10.3389/fgene.2019.00053

Jordan S., Hand B.K., Hotaling S., Delvecchia A.G., Malison R., Nissley C., Luikart G., Stanford J.A. Genomic data reveal similar genetic differentiation in aquifer species with different dispersal capabilities and life histories. *Biol. J. Linn. Soc.* 2020;129(2):315-322. DOI 10.1093/biolinnean/blz173

Jourdan J., Bundschuh M., Copilaș-Ciocianu D., Fišer C., Grabowski M., Hupało K., Jemec Kokalj A., Kabus J., Römbke J., Soose L.J., Oehlmann J. Cryptic species in ecotoxicology. *Environ. Toxicol. Chem.* 2023;42(9):1889-1914. DOI 10.1002/etc.5696

Kamaltynov R.M. Amphipods (Amphipoda: Gammaroidea). In: Index of Animal Species Inhabiting Lake Baikal and its Catchment Area. Novosibirsk, 2001;I(1):572-831 (in Russian)

Kamaltynov R.M. Higher crustaceans (Amphipoda: Gammaroidea) of Angara and Yenisey. In: Index of Animal Species Inhabiting Lake Baikal and its Catchment Area. Novosibirsk, 2009;II(1):297-329 (in Russian)]

Kao D., Lai A.G., Stamataki E., Rosic S., Konstantinides N., Jarvis E., Di Donfrancesco A., Pouchkina-Stancheva N., Sémon M., Grillo M., Bruce H., Kumar S., Siwanowicz I., Le A., Lemire A., Eisen M.B., Extavour C., Browne W.E., Wolff C., Averof M., Patel N.H., Sarkies P., Pavlopoulos A., Aboobaker A. The genome of the crustacean *Parhyale hawaiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *eLife*. 2016;5:e20062. DOI 10.7554/eLife.20062

Kovalenkova M.V. Analysis of the Evolution of Species-rich Groups of Baikal Invertebrates Based on Intron Sequences of ATP Synthase α- and β-subunit Genes. PhD Thesis. Irkutsk, 2018 (in Russian)

Lavrov D.V., Pett W. Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. *Genome Biol. Evol.* 2016;8(9):2896-2913. DOI 10.1093/gbe/evw195

Lefébure T., Douady C.J., Gouy M., Gibert J. Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Mol. Phylogenet. Evol.* 2006;40(2):435-447. DOI 10.1016/j.ympev.2006.03.014

Macdonald K.S. III, Yampolsky L., Duffy J.E. Molecular and morphological evolution of the amphipod radiation of Lake Baikal. *Mol. Phylogenet. Evol.* 2005;35(2):323-343. DOI 10.1016/j.ympev.2005.01.013

Mamos T., Grabowski M., Rewicz T., Bojko J., Strapagiel D., Burzyński A. Mitochondrial genomes, phylogenetic associations, and SNP recovery for the key invasive Ponto-Caspian amphipods in Europe. *Int. J. Mol. Sci.* 2021;22(19):10300. DOI 10.3390/ijms221910300

Mashiko K., Kamaltynov R., Morino H., Sherbakov D.Y. Genetic differentiation among gammarid (*Eulimnogammarus cyaneus*) populations in Lake Baikal, East Siberia. *Arch. Hydrobiol.* 2000;148(2): 249-261. DOI 10.1127/archiv-hydrobiol/148/2000/249

Mats V.D., Shcherbakov D.Y., Efimova I.M. Late Cretaceous–Cenozoic history of the Lake Baikal depression and formation of its unique biodiversity. *Stratigr. Geol. Correl.* 2011;19(4):404-423. DOI 10.1134/S0869593811040058

Moskalenko V.N., Neretina T.V., Yampolsky L.Y. To the origin of Lake Baikal endemic gammarid radiations, with description of two new *Eulimnogammarus* spp. *Zootaxa*. 2020;4766(3):457-471. DOI 10.11646/zootaxa.4766.3.5

Natyaganova A.V., Sitnikova T.Y. Karyotype of the Baikal amphipod *Polyacanthisca calceolata* Bazikalova, 1937, (Crustacea, Amphipoda). *Chromosome Sci.* 2012;15(1-2):43-48. DOI 10.11352/scr.15.43

Naumenko S.A., Logacheva M.D., Popova N.V., Klepikova A.V., Penin A.A., Bazykin G.A., Etingova A.E., Mugue N.S., Kondrashov A.S., Yampolsky L.Y. Transcriptome-based phylogeny of endemic Lake Baikal amphipod species flock: fast speciation accom-

panied by frequent episodes of positive selection. *Mol. Ecol.* 2017; 26(2):536-553. DOI 10.1111/mec.13927

Patra A.K., Chung O., Yoo J.Y., Kim M.S., Yoon M.G., Choi J.-H., Yang Y. First draft genome for the sand-hopper *Trinorchestia longiramus*. *Sci. Data*. 2020;7(1):85. DOI 10.1038/s41597-020-0424-8

Patra A.K., Chung O., Yoo J.Y., Baek S.H., Jung T.W., Kim M.S., Yoon M.G., Yang Y., Choi J.-H. The draft genome sequence of a new land-hopper *Platorchestia hallaensis*. *Front. Genet.* 2021;11: 621301. DOI 10.3389/fgene.2020.621301

Petunina Z.V. Comparative Ecological and Genetic Analysis of Microsporidia and Their Host, the Baikal Amphipod *Gmelinoides fasciatus*. PhD Thesis. Irkutsk, 2015 (in Russian)

Petunina J.V., Vavrishchuk N.V., Romanova E.V. Variability of morphological and genetic traits of *Macrohectopus branickii*. In: Development of Physical and Chemical Biology, Bioengineering and Bioinformatics at the Present Stage: Abstracts of reports of the IV All-Russian sci. and pract. conf. with int. participation, dedicated to the 45th anniversary of the Department of Physical and Chemical Biology, Bioengineering and Bioinformatics of ISU. Irkutsk, October 25–27, 2023. Irkutsk: Irkutsk State University Publ., 2023;111-113 (in Russian)

Poynton H.C., Hasenbein S., Benoit J.B., Sepulveda M.S., Poelchau M.F., Hughes D.S.T., Murali S.C., Chen S., Glastad K.M., Goodisman M.A.D., … Dinh H., Han Y., Doddapaneni H., Worley K.C., Muzny D.M., Gibbs R.A., Richards S. The toxicogenome of *Hyalella azteca*: a model for sediment ecotoxicology and evolutionary toxicology. *Environ. Sci. Technol.* 2018;52(10):6009-6022. DOI 10.1021/acs.est.8b00837

Ratnasingham S., Hebert P.D.N. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol. Ecol. Notes*. 2007;7(3): 355-364. DOI 10.1111/j.1471-8286.2007.01678.x

Ratnasingham S., Hebert P.D.N. A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. *PLoS One*. 2013;8(7):e66213. DOI 10.1371/journal.pone.0066213

Rivarola-Duarte L. Unraveling the genetic secrets of ancient Baikal amphipods. PhD Thesis. Leipzig: Universität Leipzig, 2021

Rivarola-Duarte L., Otto C., Jühling F., Schreiber S., Bedulina D., Jakob L., Gurkov A., Axenov-Gribanov D., Sahyoun A.H., Lucassen M., Hackermüller J., Hoffmann S., Sartoris F., Pörtner H.-O., Timofeyev M., Luckenbach T., Stadler P.F. A first glimpse at the genome of the Baikalian amphipod *Eulimnogammarus verrucosus*. *J. Exp. Zoolog. B Mol. Dev. Evol.* 2014;322(3):177-189. DOI 10.1002/jez.b.22560

Romanova E.V., Aleoshin V.V., Kamaltynov R.M., Mikhailov K.V., Logacheva M.D., Sirotinina E.A., Gornov A.Yu., Anikin A.S., Sherbakov D.Yu. Evolution of mitochondrial genomes in Baikalian amphipods. *BMC Genomics*. 2016a;17(14):1016. DOI 10.1186/s12864-016-3357-z

Romanova E.V., Mikhailov K.V., Logacheva M.D., Kamaltynov R.M., Aleoshin V.V., Sherbakov D.Y. The complete mitochondrial genome of Baikalian amphipoda *Eulimnogammarus vittatus* (Dybowsky, 1874). *Mitochondrial DNA Part A*. 2016b;27(3):1795-1797. DOI 10.3109/19401736.2014.963817

Romanova E.V., Mikhailov K.V., Logacheva M.D., Kamaltynov R.M., Aleoshin V.V., Sherbakov D.Yu. The complete mitochondrial genome of a deep-water Baikalian amphipoda *Brachyuropus grewingkii* (Dybowsky, 1874). *Mitochondrial DNA Part A*. 2016c;27(6): 4158-4159. DOI 10.3109/19401736.2014.1003891

Romanova E.V., Bukin Y.S., Mikhailov K.V., Logacheva M.D., Aleoshin V.V., Sherbakov D.Yu. Hidden cases of tRNA gene duplication and remolding in mitochondrial genomes of amphipods. *Mol. Phylogenet. Evol.* 2020;144:106710. DOI 10.1016/j.ympev.2019.106710

Romanova E.V., Bukin Y.S., Mikhailov K.V., Logacheva M.D., Aleoshin V.V., Sherbakov D.Y. The mitochondrial genome of a freshwater pelagic amphipod *Macrohectopus branickii* is among the longest in Metazoa. *Genes*. 2021;12(12):2030. DOI 10.3390/genes12122030

Salemaa H., Kamaltynov R.M. The chromosome numbers of endemic Amphipoda and Isopoda – an evolutionary paradox in the ancient lakes Ohrid and Baikal. In: Martens K., Goddeeris B., Coulter G. (Eds.) Speciation in Ancient Lakes. *Advances in Limnology*. Vol. 44. Stuttgart (Germany): Schweizerbart Science Publ.,1994;247-256

Shao C., Sun S., Liu K., Wang J., Li S., Liu Q., Deagle B.E., Seim I., Biscontin A., Wang Q., … Zhang G., Yang H., Xu X., Wang J., Zhao X., Meyer B., Fan G. The enormous repetitive Antarctic krill genome reveals environmental adaptations and population insights. *Cell*. 2023;186(6):1279-1294.e19. DOI 10.1016/j.cell.2023.02.005

Sherbakov D.Y. Molecular phylogenetic studies on the origin of biodiversity in Lake Baikal. *Trends Ecol. Evol.* 1999;14(3):92-95. DOI 10.1016/S0169-5347(98)01543-2

Sherbakov D.Yu., Kovalenkova M.V., Maikova O.O. Some results of molecular phylogenetic studies of Baikal endemic invertebrates. *Russ. J. Genet. Appl. Res.* 2017;7(4):345-349. DOI 10.1134/S2079059717040104

Sket B., Morino H., Takhteev V., Rogers D.C. Chapter 16.6 – Phylum Arthropoda: Malacostraca. In: Thorp and Covich's Freshwater Invertebrates. Vol. 4: Keys to Palaearctic Fauna. Boston: Acad. Press, 2019;789-889. DOI 10.1016/B978-0-12-385024-9.00022-8

Smith D.R. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Brief. Funct. Genomics*. 2016;15(1):47-54. DOI 10.1093/bfgp/elv027

Takhteev V.V. Essays on the Amphipods of Lake Baikal (Systematics, comparative ecology, evolution). Irkutsk, 2000a (in Russian)

Takhteev V.V. Trends in the evolution of Baikal amphipods and evolutionary parallels with some marine malacostracan faunas. In: Advances in Ecological Research. Vol. 31: Ancient Lakes: Biodiversity, Ecology and Evolution. Acad. Press, 2000b;197-220. DOI 10.1016/S0065-2504(00)31013-3

Takhteev V. On the current state of taxonomy of the Baikal Lake amphipods (Crustacea, Amphipoda) and the typological ways of constructing their system. *Arthropoda Sel.* 2019;28(1):374-402. DOI 10.15298/arthsel.28.3.03

Takhteev V.V., Berezina N.A., Sidorov D.A. Checklist of the Amphipoda (Crustacea) from continental waters of Russia, with data on alien species. *Arthropoda Sel.* 2015;24(3):335-370. DOI 10.15298/arthsel.24.3.09

Toews D.P.L., Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* 2012;21(16):3907-3930. DOI 10.1111/j.1365-294X.2012.05664.x

Väinölä R., Kamaltynov R.M. Species diversity and speciation in the endemic amphipods of Lake Baikal: molecular evidence. *Crustaceana*. 1999;72(8):945-956

Weston J.N.J., Jensen E.L., Hasoon M.S.R., Kitson J.J.N., Stewart H.A., Jamieson A.J. Barriers to gene flow in the deepest ocean ecosystems: evidence from global population genomics of a cosmopolitan amphipod. *Sci. Adv.* 2022;8(43):eabo6672. DOI 10.1126/sciadv.abo6672

Wyngaard G.A., Skern-Mauritzen R., Malde K., Prendergast R., Peruzzi S. The salmon louse genome may be much larger than sequencing suggests. *Sci. Rep.* 2022;12(1):6616. DOI 10.1038/s41598-022-10585-2

Yampolsky L.Yu., Kamaltynov R.M., Ebert D., Filatov D.A., Chernykh V.I. Variation of allozyme loci in endemic gammarids of Lake Baikal. *Biol. J. Linn. Soc.* 1994;53(4):309-323. DOI 10.1111/j.1095-8312.1994.tb01015.x

П.Б. Дроздова, Е.В. Мадьярова, А.Н. Гурков …
Т.Е. Перетолчина, Д.Ю. Щербаков, М.А. Тимофеев

Байкальские амфиподы
и их геномы, большие и малые

2024
28•3

Yuxiang W., Peretolchina T.E., Romanova E.V., Sherbakov D.Y. Comparison of the evolutionary patterns of DNA repeats in ancient and young invertebrate species flocks of Lake Baikal. *Vavilov J. Genet. Breed.* 2023;27(4):349-356. DOI 10.18699/VJGB-23-42

Zaidykov I.Y., Naumova E.Y., Sukhanova L.V. MtDNA polymorphism of *Macrohectopus branickii* Dybowsky, 1974 (Amphipoda) – an endemic pelagic key species of Lake Baikal. In: Chaplina T. (Ed.) Complex Investigation of the World Ocean (CIWO-2023). Springer Nature Switzerland, 2023;223-229. DOI 10.1007/978-3-031-47851-2_26

Zolotovskaya E., Nazarova A., Saranchina A., Mutin A., Drozdova P., Lubyaga Y., Timofeyev M. Hemocyte proteome of the Lake Baikal endemic *Eulimnogammarus verrucosus* (Crustacea: Amphipoda) sheds light on immune-related proteins. *Biol. Commun.* 2021;66(4): 290-301. DOI 10.21638/spbu03.2021.402

Whole exome sequencing enables the correct diagnosis of Frank–Ter Haar syndrome in a Saudi family

Y.N. Khan[1] ✉, M. Imad A.M. Mahmud[1], N. Othman[1], H.M. Radzuan[1], S. Basit[2, 3]

[1] Department of Basic Medical Sciences, Faculty of Medicine, International Islamic University Malaysia
[2] Department of Biochemistry and Molecular Medicine, College of Medicine, Taibah University Al Madinah Al Munawara, Saudi Arabia
[3] Center for Genetics and Inherited Diseases, Taibah University Al Madinah Al Munawara, Saudi Arabia

✉ ynkhanlughmani@gmail.com

Abstract. Frank–Ter Haar syndrome (FTHS) is a rare genetic hereditary autosomal recessive disorder characterized by defective malformation of cardiovascular, craniofacial, and skeletal system. Mutations in the *SH3PXD2B* gene are a common cause in the development of FTHS. We recruited a family with two affected individuals (3-year-old female and 2-month-old male infant) having bilateral clubfoot. Family pedigree shows an autosomal recessive mode of inheritance. DNA was extracted from the blood samples of six members of the family. Whole exome sequencing was done for the two affected individuals and the variant was validated in the whole family by using Sanger sequencing approach. Whole exome sequencing (WES) data analysis identified a rare homozygous variant (c.280C>G; p.R94G) in the *SH3PXD2B* gene, and Sanger sequencing showed that the same variant perfectly segregates with the phenotype in the pedigree. Moreover, the variant is predicted to be damaging and deleterious by several computation tools. Revisiting the family members for detailed clinical analysis, we diagnosed the patients as having the typical phenotype of FTHS. This study enabled us to correctly diagnose the cases of FTHS in a family initially recruited for having bilateral clubfoot by using WES. Moreover, this study identified a novel homozygous missense variant (c.280C>G; p.R94G) in (NM_001308175.2) the *SH3PXD2B* gene as a causative variant for autosomal recessive FTHS. This finding supports the evidence that homozygous mutations in the *SH3PXD2B* gene are the main cause in the development of FTHS.

Key words: exome sequencing; mutation; *SH3PXD2B* gene; Frank–Ter Haar syndrome.

Полное секвенирование экзома позволило безошибочно диагностировать синдром Франка–Тер Хаара в одной из саудоаравийских семей

Я.Н. Хан[1] ✉, М. Имад А.М. Махмуд[1], Н. Отман[1], Х.М. Рандзуан[1], С. Басит[2, 3]

[1] Кафедра фундаментальных медицинских наук, медицинский факультет, Международный исламский университет Малайзии
[2] Кафедра биохимии и молекулярной медицины, медицинский колледж, Университет Тайба, Медина, Саудовская Аравия
[3] Центр генетики и наследственных заболеваний, Университет Тайба, Медина, Саудовская Аравия

✉ ynkhanlughmani@gmail.com

Аннотация. Синдром Франка–Тер Хаара (Frank–Ter Haar syndrome, FTHS) – редкое генетическое заболевание с аутосомно-рецессивным типом наследования, характеризующееся аномалиями развития сердечно-сосудистой системы, костей лицевого черепа и скелета. Наиболее распространенной причиной развития данного синдрома являются мутации в гене *SH3PXD2B*. Для исследования была выбрана семья, в которой двое детей (трехлетняя девочка и двухмесячный мальчик) страдали двусторонней косолапостью. В семейной родословной указывался аутосомно-рецессивный тип наследования. Из крови шести членов семьи мы выделили образцы ДНК. Для упомянутых двоих детей было проведено полное секвенирование экзома, а секвенированием по Сэнгеру подтверждено наличие мутантного варианта у всех членов семьи. По результатам анализа данных

Я.Н. Хан, М. Имад А.М. Махмуд
Н. Отман, Х.М. Рандзуан, С. Басит

Полное секвенирование экзома позволило диагностировать
синдром Франка–Тер Хаара в саудоаравийской семье

2024
28·3

полноэкзомного секвенирования (WES) была выявлена редкая гомозиготная мутация (c.280C>G; p.R94G) в гене *SH3PXD2B*. Секвенирование по Сэнгеру показало, что эта мутация идеально сегрегирует с указанным фенотипом в родословной. Более того, при использовании ряда инструментальных средств получены данные, предсказывающие вредность и опасность этой мутации. При повторном посещении членов семьи с целью проведения развернутого клинического анализа было установлено, что фенотип двоих детей, страдавших двусторонней косолапостью, характерен для больных с синдромом FTHS. Таким образом, исследование позволило безошибочно диагностировать синдром FTHS в семье, первоначально выбранной в связи с двусторонней косолапостью у ее членов, с помощью WES. Более того, наше исследование показало, что причиной развития синдрома FTHS с аутосомно-рецессивным типом наследования была вновь выявленная гомозиготная миссенс-мутация (c.280C>G; p.R94G) в гене (NM_001308175.2) *SH3PXD2B*. Это служит дополнительным подтверждением существующих данных о том, что гомозиготные мутации в гене *SH3PXD2B* являются основной причиной развития синдрома FTHS.

Ключевые слова: секвенирование экзома; мутация; ген *SH3PXD2B*; синдром Франка–Тер Хаара.

Introduction

Frank–Ter Haar syndrome (FTHS) is a rare genetic hereditary autosomal recessive disorder characterized by cranial deformities like wide fontanelle and enlarged forehead, facial deformities such as small chin and full cheeks, ocular anomalies, namely exophthalmos, enlarged cornea with or without glaucoma and hypertelorism, protruded ear auricles, cardiovascular and skeletal deformities including a long coccyx bone with an overlying skin fold (Mass et al., 2004). Clinical features and genetic relations of the syndrome were first described by Frank et al. in a Dutch family in 1973 (Frank et al., 1973). Nine years later, Ter Haar et al. confirmed that the phenotype is inherited in an autosomal recessive manner (ter Har et al., 1982). Hence the name of the phenotype – Frank–Ter Haar syndrome.

Genetic studies suggested that mutation in the *SH3PXD2B* gene is a common cause in the development of FTHS. A study on 13 homozygously affected families mapped out and revealed gour different intronic mutations with two complete deletions in the *SH3PXD2B* gene (Iqbal et al., 2010; Massadeh et al., 2022). A knock out study showed that a deficient protein TKS4 encoded by the *SH3PXD2B* gene presents similar morphological features such as craniofacial, musculoskeletal, cardiovascular, and ocular anomalies (Iqbal et al., 2010). A literature review by Durand B. et al. in 2020 showed that 40 patients manifesting clinical features similar to FTHS have been reported worldwide, half of them were carrying mutations in *SH3PXD2B* (Durand et al., 2020).

Whole exome sequencing (WES) has revolutionized the modern era of clinical diagnosis, especially the diseases with variable phenotypic presentations and of multiorgan involvement. Whole exome sequencing allows the diagnosis of monogenic diseases and is recommended by the American College of Medical Genetics and Genomics (ACMG) as a first-line testing option to detect mutations causing genetic disorders presenting one or more congenital abnormalities and development delays, also ascertaining potential risks in individuals prior to disease manifestation, thereby avoiding unnecessary diagnostic tests (Manickam et al., 2021). One study accurately established the clinical diagnosis of Cohen syndrome when genomic analysis on DNA samples of affected and unaffected individuals was performed; otherwise, the diagnosis would have been impossible to make because of the different clinical presentations of the same disease in the affected family members (Hashmi et al., 2020). García-Aznar et al. reported a female patient having features suggestive of Soto syndrome

and initial genetic analysis did not reveal a mutation in the pathogenic gene but whole exome sequencing of all the genes showed a frameshift variant in the *AMER1* gene causing the phenotype of osteopathia striata with cranial sclerosis, which was later confirmed upon doing retrospective clinical and instrumental examination (García-Aznar et al., 2021). Hence, the role of the whole exome sequencing is crucially important in diseases with non-specific clinical presentations. Furthermore, exome sequencing carries a positive impact on management of the affected individuals and genetic counseling of their family members. A case report of a patient with severe transfusion-dependent anemia that was clinically diagnosed as Diamond–Blackfan anemia (DBA), but WES analysis finally revealed the condition as a variant of hereditary hemolytic anemia. Thus, the child was successfully managed with splenectomy, which ultimately reduced his blood transfusion dependency (Khurana et al., 2018).

Here we report a family of 6 members, where two children having bilateral clubfoot were studied to identify the genetic defects underlying the clubfoot phenotype. WES identified a pathogenic variant in the *SH3PXD2B* gene. Clinical re-examination revealed additional morphological features in the patients, establishing the diagnosis as FTHS.

Methods

A single four-generation family with 2 affected individuals was phenotypically and genetically analyzed. The family pedigree shown in the Figure was drawn to assess the pattern of inheritance of this disorder. Ethical review committee date 20-09-2020 Study ID: 036-1441 of the Taibah University, Medina, Kingdom of Saudi Arabia approved the research study. Parents of the affected individuals signed the written informed consent after understanding the aims of the study, which were explained in their local (Arabic) language.

Genomic study (DNA extraction). Blood samples were collected from the parents (III:1 and III:2), two unaffected healthy sibs (IV:1 and IV:2) and two affected individuals (IV:3 and IV:4) (see the Figure). Genomic DNA was extracted by using the QIAmp DNA micro kit (Hilden, Germany). DNA quantity and quality was assessed by using a Nano Drop TM spectrophotometer.

Next Generation Sequencing (NGS) methods. After confirming the standard DNA quality and quantity, whole exome sequencing was performed on the affected individuals (IV:3 and IV:4) using the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA). The SureSelect Target Enrichment
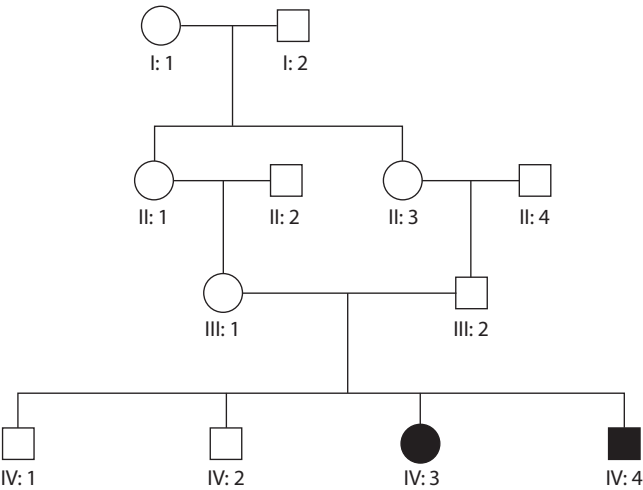
Family pedigree shows consanguinity, carriers, and affected individuals. The pedigree depicts an autosomal recessive mode of inheritance for this variant mutation. The female and male individuals are represented with circle and square symbols respectively. Filled symbols signify homozygous individuals for the missense variant (c.280C>G) in *SH3PXD2B*.

Kit v6 was used to prepare the libraries as elaborated in earlier studies (Rafiullah et al., 2022; Ullah et al., 2022). Sequencing data coverage was 30x and sequencing data depth was 100x. Standard filtration steps were followed to analyze VCF (variant calling files) of the two affected individuals, which were uploaded by using the online Illumina Base Space analysis tool (https://basespace.illumina.com). As shown in the family pedigree (see the Figure), due to an autosomal recessive pattern of inheritance with consanguineous marriage in the family, only two affected individuals having homozygous and heterozygous variants were filtered for the analysis.

Sanger sequencing for validation and segregation analysis. Variant-specific primers were designed for the prioritized variant after exome filtration. Ensembl genome browser (https://m.ensembl.org) was used to download the exonic sequence for the specific gene. Primer 3 software (http://primer3.ut.ee) was used to design the specific primers for identified variants with 30x sequencing data coverage and 100x sequencing data depth. Purification of PCR-amplified DNA was achieved using the Marligen Biosciences kits (Ijamsville, MD, USA). Sanger sequencing was performed using the BigDye sequencing kit (Applied Biosystems, USA) as described earlier (Alluqmani, Basit, 2022; Ijaz et al., 2022). Alignment of the Sanger sequencing reads with reference sequences were obtained using BIOEDIT to confirm variant identity.

***In silico* tools were used to calculate pathogenicity scores.** Various *in silico* tools were used to calculate the pathogenicity scores including meta scores as well as individual scores of the variant by using BayesDel addAF (https://fengbj-laboratory.org/BayesDel/BayesDel.html), MetaLR (https://www.ensembl.org/info/genome/variation/prediction/protein_function.html), MetaSVM (http://cancergenome.nih.gov), and REVEL (https://blog.goldenhelix.com/annotate-your-varseq-projects-with-revel/engines). Moreover CADD, (https://asia.ensembl.org/info/genome/variation/prediction/protein_function.html#CADD), DANN, FATHMM, LRT,

Mutation assessor (http://fathmm.biocompute.org.uk/), MutationTaster (https://www.mutationtaster.org/), MutPred (http://mutpred.mutdb.org/), PolyPhen2 (http://genetics.bwh.Haarvard.edu/pph2/), PROVEAN (https://www.jcvi.org/research/provean), and SIFT (https://www.merriam-webster.com/dictionary/sift) engines were also used to calculate individual pathogenicity scores.

Results
Both affected individuals were referred to specialists in multiple disciplines such as pediatrician, cardiologist, ophthalmologist, orthopedic surgeon, pediatric neurologist and finally referred to a specialist in clinical genetics at the Maternity and Children Hospital, Al Madinah Al Munawara for further evaluation and care. Details of the clinical presentation of both cases (IV:3, IV:4) as documented by the specialists of different clinical departments at the Maternity and Children Hospital Al Madina Al Munawara are mentioned in Table 1.

Potentially pathogenic missense mutation in *SH3PXD2B* in both patients
Sequencing reads were aligned to the reference genome and variants were annotated and prioritized based on the phenotype of the patients (IV:3 and IV:4). WES data failed to identify any pathogenic variant in the genes associated with clubfoot. All variants in the WES data were annotated, filtered, and prioritized for rare (minor allele frequency less than 0.001), homozygous or heterozygous, shared (common to both affected individuals) and potentially pathogenic variants (based on SIFT and PolyPhen2 scores). Variants in *OBSL1* (NM_015311.3; c.4989+5G>A), *SH3PXD2B* (NM_001308175.2; c.280C>G; p.R94G), and *MAN2B1* (NM_000528.4; c.2402dupG; p.S802fs*129) were initially prioritized.

Sanger sequencing validated and confirmed the autosomal recessive inheritance of the *SH3PXD2B* variant in the family
Primers were designed for all three variants that were amplified by polymerase chain reaction (PCR) in all available members III:1, III:2, IV:1, IV:2, IV:3, IV:4 of the family. Variants in *OBSL1* (c.4989+5G>A) and *MAN2B1* (c.2402dupG) were found not to segregate in the family, therefore, they were not considered for further analysis. A variant in *SH3PXD2B* (c.280C>G) perfectly segregates with the phenotype in the pedigree. Both parents and unaffected individuals are found to be heterozygous for the variant and both affected individuals are homozygous for it. Therefore, a rare (0 % gnomAD frequency) homozygous missense variant (c.280C>G; p.R94G) in the *SH3PXD2B* (NM_001308175.2) gene was considered as the most plausible candidate variant for the disease phenotype in this family. The variant is present in the exome data of both affected individuals (IV:3 and IV:4).

In silico analysis predicted the variant (c.280C>G) in *SH3PXD2B* to be potentially pathogenic
Most of the *in silico* engines including CADD, DANN, FATHMM, LRT, Mutation assessor, MutationTaster, MutPred, PolyPhen2, PROVEAN, and SIFT predicted the variant to be disease causing, damaging or pathogenic. Table 2 shows the score and prediction obtained after analyzing the variant with

Я.Н. Хан, М. Имад А.М. Махмуд
Н. Отман, Х.М. Рандзуан, С. Басит

Полное секвенирование экзома позволило диагностировать
синдром Франка–Тер Хаара в саудоаравийской семье

2024
28·3

Table 1. Comparison of the clinical manifestations of Frank–Ter Haar syndrome
in family studies by Iqbal et al., 2010, and by Durand et al., 2020

Clinical manifestations of confirmed cases of Frank–Ter Haar syndrome having *SH3PXD2B* mutations	Iqbal et al., 2010	Durand et al., 2020	Present study, 2023	
Family	13 families	21 families	1 family	
			Case IV:3	Case IV:4
Gender	3F, 10 M	8F, 13M	F	M
Consanguinity	12/13	05/14	Yes	Yes
Cognitive disabilities				
Vision, adaptation, learning	NA	4/8	Yes	Yes
Hearing, communication, learning difficulty	NA	4/8	Yes	Yes
Motor developmental abnormality	NA	9/18	Yes	Yes
Craniofacial				
Large open anterior fontanelle	12/13	17/18	Yes	Yes
Protruding forehead	13/13	21/21	Yes	Yes
Increased intracranial pressure	12/13	NA	Yes	Yes
Bilateral coronal craniosynostosis	NA	8/18	NA	NA
Bilateral sagittal craniosynostosis	12/13	NA	Yes	Yes
Orbital hypertelorism	12/12	21/21	Yes	Yes
Unturned nostrils	6/9	14/20	Yes	Yes
Puffy cheeks	13/13	21/21	Yes	Yes
Long philtrum	NA	13/17	Yes	Yes
Thin upper lip	NA	8/15	Yes	Yes
Macro stomia	13/13	18/18	Yes	Yes
Microgenia	10/13	16/19	Yes	Yes
Gingival hyperplasia	NA	11/11	NA	NA
Micrognathia	10/13	16/19	Yes	Yes
Otapostasis	8/10	13/15	Yes	Yes
Broad alveolar ridges	6/11	7/9	Yes	Yes
Ophthalmic				
Eyes protrusion	NA	18/18	Yes	Yes
Macro cornea	9/12	14/18	Yes	Yes
Bilateral buphthalmias	NA	8/14	NA	NA
Congenital raised intraocular pressure/Glaucoma	NA	6/18	Yes	Yes
Cardiology				
Cardiomegaly	NA	NA	Yes	Yes
Arterial septal defect	NA	1/18	Yes	Yes
Patent ductus arteriosus	2/3	NA	Yes	Yes
Ventricular septal defect	5/10	8/18	Yes	Yes
Aortic regurgitation/prolapse	3/9	2/17	Yes	Yes
Double right outlet (Pulmonary trunk)	NA	NA	Yes	Yes
Mitral valves prolapse/regurgitation	6/9	7/17	Yes	Yes
Tricuspid valves prolapse/regurgitation	1/9	2/18	NA	NA
Musculo-skeletal				
Talipes Equiano Varus (clubfoot)	7/11	11/19	Yes	Yes
Feet size discrepancy	11/13	NA	Yes	Yes
Congenital hand deformities	13/13	20/20	Yes	Yes
I) contractures flexion/extension deformity	4/13	14/21	Yes	Yes
II) brachydactyl, shorthand, digits deformity	13/13	14/21	Yes	Yes
Bowing of the long bones	7/10	11/15	Yes	Yes
Kyphosis	6/11	13/18	Yes	Yes
Prominent coccyx	9/13	14/18	Yes	Yes
Pectus excavatum	NA	5/18	Yes	Yes
Subcutaneous nodules	12/13	NA	NA	NA

N o t e. –/–, Number of families positive for mentioned clinical features/total number of families studied. "Yes" is for patients having the mentioned clinical features and "NA" indicates the not available or absence of the clinical features.

Table 2. *In silico* analytical prediction
of the potential pathogenicity
of the missense variant (c.280C>G) in *SH3PXD2B*

Tools	Prediction	Score
REVEL	Damaging	0.877
MetaLR	Pathogenic	0.8334
SIFT	Damaging	0.00
CADD	Deleterious	
PolyPhen2	Probably damaging	0.010
Conservation GERP	Highly conserved	5.480
GenoCanyon	Deleterious	1.000
fitCons	Deleterious	0.730
Aggregated	Deleterious	0.870
Mutation assessor	Pathogenic	3.700
PhastCons100way	Highly conserved	1.000
PhyloP100way	Conserved	6.369
MutPred	Pathogenic	0.786
MutationTaster	Disease-causing	0.9999
PROVEAN	Damaging	–5.820

various *in silico* software. A very low frequency in gnomAD (PM2) and support from multiple lines of computational evidence (PolyPhen2, SIFT, CADD) (PP3), as well as segregation of the variant with the disease phenotype in the family support the hypothesis that this variant is an underlying cause of the phenotype in our case.

Discussion

Congenital inherited disorders such as FTHS have broad overlapping clinical presentations that often make them difficult and unlikely to be diagnosed. Biochemical laboratory tests do not even show any evidential clues for these disorders and the genes are only investigated for research purposes. Next-generation technologies such as whole exome sequencing are considerably affordable, and a preferable testing platform in situations where two or more than two affected individuals are found in a consanguineous marriage family (Alluqmani, Basit, 2022).

In this study, a consanguineous marriage family from Saudi Arabia having two affected individuals was investigated both clinically and genetically. The family was referred to the Center for Genetics and Inherited Diseases, Taibah University for the genetic diagnosis of clubfoot. Family members were registered, and WES was performed. Initially, genes associated with clubfoot (*PITX1*, *TBX4*, *HOXA9*, *HOXD10*, *HOXD12*, *HOXD13*, *HOXA9*, *TPM1*, *TPM2*, *COL9A1*, *FLNB*, *CASP8*, *CASP10*, *UTX*, *CHD1*, *RIPPLY2*, *CAND2*, *WNT7*) were screened for potential variants. However, WES data analysis failed to detect any potential pathogenic variant in clubfoot-associated genes. Therefore, an unbiased and hypothesis-free approach was used to analyze WES data to filter and prioritize variants of interest. A potentially pathogenic variant in the *SH3PXD2B* gene was identified. Patients were recalled by the physician, and they were thoroughly re-examined. Clinical review of the affected individuals showed additional features of

musculoskeletal deformity, cardiac, ophthalmic, craniofacial disorders, and cognitive disabilities. These clinical features helped us to classify our cases as FTHS (Iqbal et al., 2010). In this family, the affected individuals were also found to have cardiomegaly and a double pulmonary trunk, which were not reported previously. While gingival hyperplasia, buphthalmia, and subcutaneous nodules are the features commonly reported in such cases in the literature, these are not seen in our cases (Durand et al., 2020).

FTHS is primarily caused by mutation in the *SH3PXD2B* gene. This gene, located on 5q35.1 chromosome, encodes a 911-amino-acid protein, which has a phox homology (PX) domain, known as Tks4 (tyrosine kinase substrate with four SH3 domains) (Iqbal et al., 2010). This protein is involved in the formation of actin-rich membrane protrusions called podosomes, which coordinate pericellular proteolysis with cell migration and regulate proliferation, growth, and differentiation in the cells with extracellular matrix remodeling (Gimona et al., 2008). The gene mutation leads to the absence of Tks4 and thus embryonic fibroblasts decrease the formation of mature and functional podosomes; hence, they fail to degrade the extracellular matrix (Saeed et al., 2011). Filamin A protein is present in the podosome belt, and it needs to be cleaved by calpain for maintaining osteoclast motility during bone development (Marzia et al., 2006). Filamin A is also required for podosome rosette formation, proteolysis of the extracellular matrix mediated by podosomes in macrophages, and three-dimensional mesenchymal cells build up, so mutation in the genes encoding for actin-rich membrane structures causes serious congenital anomalies of the heart, skeleton, and craniofacial region (Cejudo-Martin, Courtneidge, 2011). Newly published knockout studies proved that TKS4, once lost, can adversely affect the differentiation of different cell lineages and maturation processes, thus leading to the development of FTHS (László et al., 2022).

Hence, the ambiguous clinical presentation of FTHS is commonly seen due to overlapping features as the defect occurs during the differentiation of primordial germ layer development, which influences multiple organs and systems of the organism. Therefore, clinical use of genetic testing like WES is essential when a clinician encounters a case showing unclear clinical and/or laboratory presentation (Sharma, Nalepa, 2016).

Whole exome sequencing has played an important role in diagnosis of other diseases as well. A consanguineous Saudi family having five individuals with steroid resistant Nephrotic syndrome were examined by WES which identified a homozygous novel insertion mutation (c.6272_6273insT) in the *PLCE1* gene (Hashmi et al., 2018). WES is also considered a useful time-saving practical diagnostic tool in the evaluation of patients with rare and complex hereditary disorders like episodic ataxia type 1. This diagnostic approach can hasten early therapeutic intervention strategies and directly affect patient care (Tacik et al., 2015).

Conclusion

This study provides us with further evidence for the importance of validation of genetic variants involved in the development of the FTHS with the use of WES. Here we reported that the homozygous missense variant (c.280C>G; p.R94G)

Я.Н. Хан, М. Имад А.М. Махмуд
Н. Отман, Х.М. Рандзуан, С. Басит

Полное секвенирование экзома позволило диагностировать
синдром Франка–Тер Хаара в саудоаравийской семье

2024
28·3

in the *SH3PXD2B* (NM_001308175.2) gene can be considered as the candidate variant resulting in autosomal recessive FTHS. This study covers the *SH3PXD2B* gene mutation spectrum, which might further reflect on the importance of properly correlating genotypes with phenotypes and provides support to the importance of genetic testing and analysis of the *SH3PXD2B* gene in the Kingdom of Saudi Arabia and probably certain other locations. This will also be beneficial in marriage counseling and planning of future pregnancies among FTHS carrier families.

References

Alluqmani M., Basit S. Association of *SORD* mutation with autosomal recessive asymmetric distal hereditary motor neuropathy. *BMC Med. Genomics*. 2022;15(1):88. DOI 10.1186/s12920-022-01238-4

Cejudo-Martin P., Courtneidge S.A. Podosomal proteins as causes of human syndromes: a role in craniofacial development? *Genesis*. 2011;49(4):209-221. DOI 10.1002/dvg.20732

Durand B., Stoetzel C., Schaefer E., Calmels N., Scheidecker S., Kempf N., De Melo C., Guilbert A.S., Timbolschi D., Donato L., Astruc D. A severe case of Frank-ter Haar syndrome and literature review: further delineation of the phenotypical spectrum. *Eur. J. Med. Genet*. 2020;63(4):103857. DOI 10.1016/j.ejmg.2020.103857

Frank Y., Ziprkowski M., Romano A., Stein R., Katznelson M.B., Cohen B., Goodman R.M. Megalocornea associated with multiple skeletal anomalies: a new genetic syndrome. *J. Genet. Hum.* 1973; 21(2):67-72

García-Aznar J.M., Ramírez N., De Uña D., Santiago E., Monserrat L. Whole exome sequencing provides the correct diagnosis in a case of osteopathia striata with cranial sclerosis: case report of a novel frameshift mutation in *AMER1*. *J. Pediatr. Genet*. 2021;10(2):139-146. DOI 10.1055/s-0040-1710058

Gimona M., Buccione R., Courtneidge S.A., Linder S. Assembly, and biological role of podosomes and invadopodia. *Curr. Opin. Cell Biol.* 2008;20(2):235-241. DOI 10.1016/j.ceb.2008.01.005

Hashmi J.A., Safar R.A., Afzal S., Albalawi A.M., Abdu-Samad F., Iqbal Z., Basit S. Whole exome sequencing identification of a novel insertion mutation in the phospholipase C ε-1 gene in a family with steroid resistant inherited nephrotic syndrome. *Mol. Med. Rep*. 2018;18(6):5095-5100. DOI 10.3892/mmr.2018.9528

Hashmi J.A., Fadhli F., Almatrafi A., Afzal S., Ramzan K., Thiele H., Nürnberg P., Basit S. Homozygosity mapping and whole exome sequencing provide exact diagnosis of Cohen syndrome in a Saudi family. *Brain Dev*. 2020;42(8):587-593. DOI 10.1016/j.braindev. 2020.04.010

Ijaz A., Alfadhli F., Alharbi A., Khan Y.N., Alhawas Y.K., Hashmi J.A., Wali A., Basit S. *NPHP3* splice acceptor site variant is associated with infantile nephronophthisis and asphyxiating thoracic dystrophy; a rare combination. *Eur. J. Med. Genet*. 2022;65(10):104578. DOI 10.1016/j.ejmg.2022.104578

Iqbal Z., Cejudo-Martin P., de Brouwer A., van der Zwaag B., Ruiz-Lozano P., Scimia M.C., Lindsey J.D., Weinreb R., Albrecht B., Megarbane A., Alanay Y. Disruption of the podosome adaptor protein TKS4 (*SH3PXD2B*) causes the skeletal dysplasia, eye, and cardiac abnormalities of Frank-Ter Haar syndrome. *Am. J. Hum. Genet*. 2010;86(2):254-261. DOI 10.1016/j.ajhg.2010.01.009

Khurana M., Edwards D., Rescorla F., Miller C., He Y., Potchanant E.S., Nalepa G. Whole-exome sequencing enables correct diagnosis and surgical management of rare inherited childhood anemia. *Cold Spring Harb. Mol. Case Stud.* 2018;4(5):a003152. DOI 10.1101/mcs.a003152

László L., Maczelka H., Takács T., Kurilla A., Tilajka Á., Buday L., Vas V., Apáti Á. A novel cell-based model for a rare disease: the Tks4-KO human embryonic stem cell line as a Frank-Ter Haar syndrome model system. *Int. J. Mol. Sci.* 2022;23(15):8803. DOI 10.3390/ijms23158803

Maas S.M., Kayserili H., Lam J., Apak M.Y., Hennekam R.C. Further delineation of Frank-Ter Haar syndrome. *Am. J. Med. Genet. A*. 2004;131(2):127-133. DOI 10.1002/ajmg.a.30244

Manickam K., McClain M.R., Demmer L.A., Biswas S., Kearney H.M., Malinowski J., Massingham L.J., Miller D., Yu T.W., Hisama F.M.; ACMG Board of Directors. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* 2021;23(11):2029-2037. DOI 10.1038/s41436-021-01242-6

Marzia M., Chiusaroli R., Neff L., Kim N.Y., Chishti A.H., Baron R., Horne W.C. Calpain is required for normal osteoclast function and is down-regulated by calcitonin. *J. Biol. Chem.* 2006;281(14):9745-9754. DOI 10.1074/jbc.M513516200

Massadeh S., Alhabshan F., AlSudairi H.N., Alkwai S., Alsuwailm M., Kabbani M.S., Chaikhouni F., Alaamery M. The role of the disrupted podosome adaptor protein (*SH3PXD2B*) in Frank-Ter Haar syndrome. *Genes*. 2022;13(2):236. DOI 10.3390/genes13020236

Rafiullah R., Albalawi A.M., Alaradi S.R., Alluqmani M., Mushtaq M., Wali A., Basit S. An expansion of phenotype: novel homozygous variant in the *MED17* identified in patients with progressive microcephaly and global developmental delay. *J. Neurogenet*. 2022;36(4):108-114. DOI 10.1080/01677063.2022.2149748

Saeed M., Shair Q.A., Saleem S.M. Frank-Ter Haar Syndrome. *J. Coll. Physicians Surg. Pak*. 2011;21(4):252-253

Sharma R., Nalepa G. Evaluation, and management of chronic pancytopenia. *Pediatr. Rev.* 2016;37(3):101-111. DOI 10.1542/pir.2014-0087

Tacik P., Guthrie K.J., Strongosky A.J., Broderick D.F., Riegert-Johnson D.L., Tang S., El-Khechen D., Parker A.S., Ross O.A., Wszolek Z.K. Whole-exome sequencing as a diagnostic tool in a family with episodic ataxia type 1. *Mayo Clin. Proc*. 2015;90(3):366-371. DOI 10.1016/j.mayocp.2015.01.001

ter Haar B., Hamel B., Hendriks J., de Jager J., Opitz J.M. Melnick–Needles syndrome: indication for an autosomal recessive form. *Am. J. Med. Genet*. 1982;13(4):469-477. DOI 10.1002/ajmg.1320130418

Ullah A., Shah A.A., Alluqmani M., Haider N., Aman H., Alfadhli F., Almatrafi A.M., Albalawi A.M., Krishin J., Ullah Khan F., Anjam B.A., Abdullah, Lozano E.P., Samad A., Ahmad W., Hansen T., Xia K., Basit S. Clinical and genetic characterization of patients segregating variants in *KPTN*, *MINPP1*, *NGLY1*, *AP4B1*, and *SON* underlying neurodevelopmental disorders: genetic and phenotypic expansion. *Int. J. Dev. Neurosci*. 2022;82(8):789-805. DOI 10.1002/jdn.10231

Computer modeling of the peculiarities in the interaction of IL-1 with its receptors in schizophrenia

N.Yu. Chasovskikh (ID) (✉), A.A. Bobrysheva, E.E. Chizhik (ID)

Siberian State Medical University of the Ministry of Healthcare of the Russian Federation, Tomsk, Russia

✉ nch03@mail.ru

Abstract. One of the primary theories regarding the development of schizophrenia revolves around genetics, indicating the involvement of hereditary factors in various processes, including inflammation. Research has demonstrated that inflammatory reactions occurring in microglia can impact the progression of the disease. It has also been established that genetically determined changes in IL-1 can contribute to schizophrenia, thereby confirming the role of the IL-1 gene cluster in disease susceptibility. The aim of this study is a computer-based assessment of the structural interactions of IL-1 proteins with their receptors in schizophrenia. The study utilized the DisGeNET database, enabling the assessment of the reliability of identified IL-1 polymorphisms. Polymorphisms were also sought using NCBI PubMed. The NCBI Protein service was employed to search for and analyze the position of the identified polymorphisms on the chromosome. Structures for modeling were extracted from the Protein Data Bank database. Protein modeling was conducted using the SWISS-MODEL server, and protein interaction modeling was performed using PRISM. Notably, this study represents the first prediction of the interactions of IL-1α, IL-1β, and IL-1RA proteins, taking into account the presence of single-nucleotide polymorphisms associated with schizophrenia in the sequence of the corresponding genes. The results indicate that the presence of SNP rs315952 in the IL-1RA protein gene, associated with schizophrenia, may lead to a weakening of the IL-1RA binding to receptors, potentially triggering the initiation of the IL-1 signaling pathway by disrupting or weakening the IL-1RA binding to receptors and facilitating the binding of IL-1 to them. Such alterations could potentially lead to a change in the immune response. The data obtained contribute theoretically to the development of ideas about the molecular mechanisms through which hereditary factors in schizophrenia influence the interactions of proteins of the IL-1 family, which play an important role in the processes of the immune system.

Key words: IL-1; schizophrenia; molecular modeling; SNP; single-nucleotide polymorphisms; PRISM.

Компьютерное моделирование особенностей взаимодействий IL-1 с его рецепторами при шизофрении

Н.Ю. Часовских (ID) (✉), А.А. Бобрышева, Е.Е. Чижик (ID)

Сибирский государственный медицинский университет Министерства здравоохранения Российской Федерации, Томск, Россия

✉ nch03@mail.ru

Аннотация. Одной из основных теорий развития шизофрении является генетическая, свидетельствующая о вовлечении наследственных факторов в различные процессы, в том числе воспаление. Показано, что воспалительные реакции, протекающие в микроглии, могут влиять на развитие заболевания. Также установлено, что генетически обусловленные изменения IL-1 могут способствовать шизофрении, подтверждая роль кластера генов IL-1 в восприимчивости к болезням. Целью работы была компьютерная оценка структурных взаимодействий белков IL-1 с их рецепторами при шизофрении. Использовалась база данных DisGeNET, позволяющая оценить достоверность выявленных полиморфизмов IL-1. Проведен поиск полиморфизмов с помощью NCBI PubMed. Сервис NCBI Protein использовался для поиска и анализа положения на хромосоме найденных полиморфизмов. Из базы данных Protein Data Bank были извлечены структуры для проведения моделирования. Моделирование белков выполнялось с помощью сервера SWISS-MODEL, а моделирование белковых взаимодействий – с помощью PRISM. В настоящем исследовании впервые проведено прогнозирование взаимодействий белков IL-1α, IL-1β и IL-1RA с учетом наличия в последовательности соответствующих генов однонуклеотидных полиморфизмов, ассоциированных с шизофренией. Показано, что наличие ассоциированного с шизофренией полиморфизма rs315952 гена белка IL-1RA может привести к ослаблению связи IL-1RA с рецепторами и, предположительно, к запуску сигнального пути IL-1 путем разрыва либо ослаб-

Н.Ю. Часовских, А.А. Бобрышева
Е.Е. Чижик

Компьютерное моделирование особенностей
взаимодействий IL-1 с его рецепторами при шизофрении

2024
28 · 3

ления связи IL-1RA с рецепторами и связыванием IL-1 с ними, что, возможно, вызовет изменение иммунного
ответа. Полученные данные вносят теоретический вклад в развитие представлений о молекулярных механизмах влияния наследственных факторов шизофрении на взаимодействия белков семейства IL-1, играющих важную роль в процессах иммунной системы.

Ключевые слова: IL-1; шизофрения; моделирование; SNP; однонуклеотидные полиморфизмы; PRISM.

Introduction

The investigation of the causes of multifactorial diseases, characterized by complex inheritance and associated with the action of multiple genes (Bochkov, 2011), is a current challenge in contemporary medical biological science. When studying such diseases, special attention is given to their potential associations with single nucleotide polymorphisms (SNPs), as well as the involvement of the corresponding genes in the implementation of molecular mechanisms underlying pathologies.

Currently, a pressing issue is the exploration of the mechanisms underlying the development of such a prevalent disorder as schizophrenia. This condition has several etiopathogenetic concepts, with one of the main theories being genetic. It suggests the involvement of genetic factors in various physiological processes of the body, including inflammatory processes. The activation of the inflammatory response system, associated with the pathophysiology of schizophrenia, has been demonstrated in numerous studies (Xu, He, 2010; Sommer et al., 2015; Kapelski et al., 2016; Miyaoka et al., 2017; Müller, 2019). Studies on animal models of schizophrenia, including mice and primates, indicate that inflammatory reactions during pregnancy may influence brain development and contribute to the etiology of this disorder (Frodl, Amico, 2014). It has been shown that microglial cells are activated in schizophrenia and play a crucial role in inflammatory processes (Müller, 2019). Additionally, the nonsteroidal anti-inflammatory drug Celecoxib has been found to exert therapeutic effects on patients with schizophrenia. Considering these findings, immunomodulation is currently widely discussed as a potential approach to the treatment of this disorder (Müller, 2019).

Clinical case descriptions of patients undergoing bone marrow transplantation demonstrate the inflammatory nature of schizophrenia. For instance, T. Miyaoka et al. (2017) presented the case of a 24-year-old man with treatment-resistant schizophrenia who underwent bone marrow transplantation for acute myeloid leukemia. After the procedure, he showed a significant reduction in schizophrenia symptoms without the use of neuroleptics. I.E. Sommer et al. (2015) described a reverse case where schizophrenia was transmitted from a sibling. At present, the mechanism of changes introduced by bone marrow transplantation from a healthy individual influencing the development of schizophrenia is not fully understood. However, it has been shown that this process normalizes microglial changes, which are significant for this disorder (Miyaoka et al., 2017). While the examination of individual cases cannot definitively confirm the immune pathogenesis of schizophrenia, the involvement of the immune system may be one of the key factors in the development of this disorder (Sommer et al., 2015).

It has been demonstrated that genetically determined changes in the regulation of IL-1 metabolism, one of the key components of the immune response, may contribute to schi-

zophrenia, thereby supporting the role of the IL-1 gene cluster in disease susceptibility (Zanardini et al., 2003). Pro-inflammatory cytokines can modify neurotransmitter metabolism, influencing the development of the nervous system. IL-1 participates in both acute and chronic neurodegeneration, suggesting that cytokines induced by the activation of the IL-1 signaling pathway may play a pivotal role both in the acute phase of the disease and during developmental stages of the brain that affect an individual's susceptibility to schizophrenia-related factors in later life (Katila et al., 1999).

Accumulated data to date provide an opportunity for a more detailed examination of the influence of individual cytokine genes, particularly IL-1, on the mechanisms underlying schizophrenia development. Bioinformatic methods enable the exploration of changes in gene sequences associated with this disorder and an assessment of the properties of the corresponding protein molecules. This includes their involvement in interleukin receptor interactions, impacting the realization of the pro-inflammatory effects of IL-1. This will expand theoretical knowledge and identify approaches for further investigations into potential mechanisms of the immune system's involvement in schizophrenia development.

The objective of this study is a computer-based assessment of the interactions between IL-1 proteins and their receptors in the context of schizophrenia.

Materials and methods

We investigated the genetic factors associated with schizophrenia using the DisGeNET platform renowned for hosting one of the largest publicly available collections of genes and variants linked to human diseases (Piñero et al., 2020). The search for SNPs and proteins related to the IL-1 family was conducted through the NCBI (National Center for Biotechnology Information) PubMed service and the Protein database (Sayers et al., 2021).

To ensure the reliability of the data obtained from the DisGeNET platform, we assessed the identified polymorphisms using the Evidence Index. An Evidence Index (EI) of 1 signifies unanimous support for Gene-Disease Associations (GDA) or Variant-Disease Associations (VDA) across all publications. A value of EI < 1 indicates the absence of a correlation between the gene/variants and the disease (Piñero et al., 2020).

Following the selection of polymorphisms in genes encoding proteins associated with the IL-1 family, we analyzed their chromosomal positions using the NCBI resource functionality (Sherry et al., 2001). It was imperative to locate the polymorphisms within the coding region for modeling the corresponding proteins.

The amino acid sequences for protein modeling were sourced from the NCBI Protein database (Sayers et al., 2021). Subsequently, we manually replaced the amino acids in the sequences based on the positions of the polymorphisms. Pro-

tein modeling using the obtained sequences was carried out using the SWISS-MODEL protein structure modeling server (Waterhouse et al., 2018).

We extracted the IL-1R1+IL-1RAP+IL-1β complex from the Protein Data Bank (PDB) database, which houses known spatial structures of proteins. Subsequently, IL-1β was removed from this structure as our analysis focused on the receptor complex without interleukin.

The receptor complexes obtained were imported into Protein Interactions by Structural Matching (PRISM) (Baspinar et al., 2014), where their surfaces underwent structural comparison with template interfaces – previously identified binding regions. An interface (binding region) is defined as a pair of sets of amino acid residues $\{(A_1, ..., A_N), (B_1, ..., B_M)\}$, where for any amino acid residue A_i from protein A, there is at least one amino acid residue B_i from protein B. This occurs in such a way that the distance between these residues does not exceed a specified threshold (typically ranging from 6 to 12 Å) (Hadarovich et al., 2020). Within the binding region, hot spots exist – amino acid residues contributing significantly to binding energy (Tuncbag et al., 2012).

PRISM operates as an algorithm for predicting and modeling protein interactions through structural matching, encompassing four key stages: extraction of the target protein surface; assessment of structural similarity with template interface partners; superimposition of protein surface areas resembling the template interface on the template; flexible refinement of the obtained complexes, and energy calculation (Aytuna et al., 2005; Tuncbag et al., 2011).

Through the modeling of molecular interactions, the PRISM service provides an interface for forecasting, the complex structure, and an energy indicator. The latter signifies binding energy, denoting the minimum work required to separate the system into its constituent particles. It characterizes system stability and consistently carries a negative value, with the system boasting the lowest binding energy considered the most stable (Acuner Ozbabacan et al., 2014).

An energy threshold value of –10 kJ/mol was employed to identify energetically favorable predictions. Interactions demonstrating conformational advantage, backed by experimental data and IS-assessment (interface similarity assessment), with an output energy less than –10 kJ/mol were deemed favorable (Gao, Skolnick, 2011; Kuzu et al., 2013). The IS score, a metric for evaluating protein-protein interaction predictions, takes into account not only geometric distances but also the preservation of interfacial contacts. For the PRISM algorithm, an IS score greater than 0.12 is considered acceptable (Gao, Skolnick, 2011).

To visualize the localization of amino acid substitutions and interactions of IL-1 with receptors in the obtained protein complexes, the YASARA program (Krieger, Vriend, 2014) was utilized.

Results

Identification of molecules
initiating the IL-1 signaling pathway
The structures of IL-1α, IL-1β, and IL-1RA proteins, crucial for initiating the IL-1 pathway, underwent examination (Dinarello, 1994). These proteins interact with specific recep-
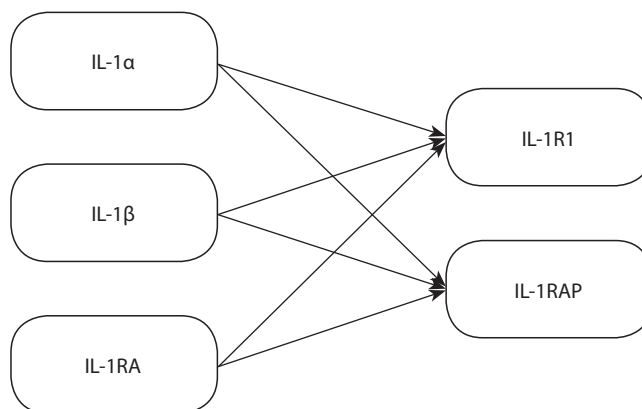

Fig. 1. IL-1 molecules and their interaction with cell receptors (based on Acuner Ozbabacan et al., 2014; Dinarello, 2018).

tors IL-1R1 and IL-1RAP (Acuner Ozbabacan et al., 2014). Subsequently, we evaluated the presence of polymorphisms in the genes of IL-1α, IL-1β, IL-1RA, IL-1R1, and IL-1RAP proteins, modeling their interactions according to the scheme presented in Figure 1.

Search for SNPs in IL-1 genes
associated with schizophrenia
A search in the DisGeNET catalog identified four single-nucleotide polymorphisms in genes initiating the IL-1 pathway associated with schizophrenia.

For IL-1α, SNPs rs113129609 and rs1800587 were found. While the EI for rs113129609 was 1, the corresponding article did not confirm its presence. For rs1800587, with an EI index less than 0.001, evidence was lacking. The rs16944 polymorphism in IL-1β, with an EI of 1, was supported by several studies (Shirts et al., 2006; Xu, He, 2010; Fatjó-Vilas et al., 2012), and the polymorphism was included in the list for further investigation. The rs1794068 polymorphism for IL-1RA had an EI less than 0.001, and further investigation was not pursued.

A PubMed search yielded 39 articles, and polymorphisms were extracted and listed in Table 1.

Analysis of the localization of SNPs
of genes initiating the IL-1 signaling pathway
Localization and information on amino acid substitution for each polymorphism were analyzed using the dbSNP resource (Sherry et al., 2001) (Table 1). The IL-1RA rs315952 polymorphism, involving the substitution of serine with arginine, was identified for further modeling.

Modeling of proteins initiating the IL-1 signaling pathway
Since the rs315952 polymorphism is located in the *IL-1RA* amino acid sequence, it was selected for modeling. The original sequence was extracted from the NCBI Protein sequence database: >NP_776214.1 interleukin-1 receptor antagonist protein isoform 1 precursor [Homo sapiens].

Three-dimensional structures of IL-1RA were modeled with and without the polymorphism using the SWISS-MODEL service. The obtained molecular models were saved as ".pdb" files.

Н.Ю. Часовских, А.А. Бобрышева
Е.Е. Чижик

Компьютерное моделирование особенностей
взаимодействий IL-1 с его рецепторами при шизофрении

2024
28·3

Table 1. Analysis of the localization and substitution of an amino acid in the sequence

| Polymorphism	Gene name	Publication	Type of amino acid substitution	Localization of the polymorphism
rs16944	*IL-1B*	Papiol et al., 2007; Xu, He, 2010	No information on amino acid substitution is available	The polymorphism is located outside the protein coding region
rs1794068	*IL-1RA*	Ben Nejma et al., 2013		Intron
rs1143627	*IL-1B*	Hudson, Miller, 2018		Intron
rs1143623	*IL-1B*	Kapelski et al., 2015		The polymorphism is located outside the protein coding region
rs4848306	*IL-1B*	Yoshida et al., 2012; Kapelski et al., 2015	No information	No information
rs4251961	*IL-1RA*	Kapelski et al., 2015	No information on amino acid substitution is available	The polymorphism is located outside the protein coding region
rs9005	*IL-1RA*	Kapelski et al., 2016		No information
rs1143633	*IL-1B*	Sasayama et al., 2011		Intron
rs11677416	*IL-1A*	McClay et al., 2011	No information	No information
rs315952	*IL-1RA*	Kapelski et al., 2016	Serine is replaced by arginine	Position 130
rs419598	*IL-1RA*	Kapelski et al., 2016	A synonymous variant, thymine to cytosine substitution (T>C), does not result in a change of the amino acid alanine	Position 57
rs1143634	*IL-1B*	Xu, He, 2010	The alteration does not lead to the replacement of the amino acid phenylalanine	Position 27
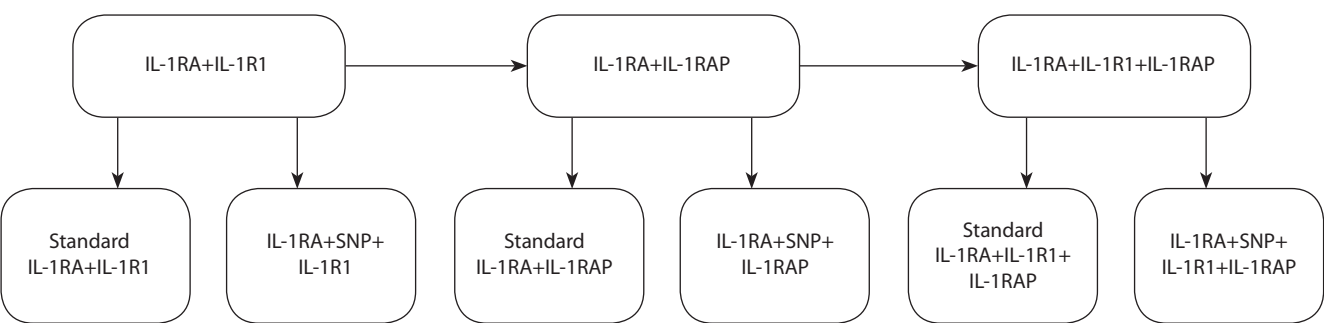
Fig. 2. Stages of protein interactions modeling.

Since IL-1RA interacts with IL-1R1, IL-1RAP, and the IL-1R1+IL-1RAP complex (Fig. 1), three-dimensional structures of the corresponding proteins were required for modeling and analysis. The PDB structure of the IL-1β signaling complex was obtained, including IL-1β (chains A, D), IL-1R1 (chains B, E), and IL-1RAP (chains C, F). The IL-1R1+IL-1RAP complex, IL-1RAP, and IL-1R1 were obtained using the PyMol program.

Modeling protein interactions initiating the IL-1 pathway
Modeling of interactions followed the scheme presented in Figure 2. Interactions of standard (non-polymorphic) IL-1RA with IL-1R1, IL-1RAP, and the IL-1R1+IL-1RAP receptor complex were modeled and analyzed sequentially. Subsequently, interactions with IL-1RA rs315952 were also modeled.

Modeling IL-1RA interactions with IL-1R1
Modeling interaction of standard IL-1RA with IL-1R1. In the obtained models, the minimum energy indicators demonstrated interaction according to the 1itbAB template, characterizing the most probable interaction where the structure is maximally stable (Table 2).

Considering the rs315952 polymorphism (Table 1) involving the replacement of serine by arginine at position 130, the interaction at this point was evaluated under normal and polymorphic conditions. According to the contact list of template residues, serine at position 130 of the IL-1RA molecule binds to leucine at position 237 of IL-1R1 (see Table 4).

Modeling interaction of IL-1RA with the rs315952 polymorphism with IL-1R1. The structures of IL-1RA molecules with the rs315952 polymorphism and IL-1R1 were utilized for modeling.

Table 2. Interfaces (binding regions) and interaction energies of standard IL-1RA with IL-1R1

Interface (binding region)	Decoding of the binding domain name	Energy, kJ/mol
1itbAB	Interaction of chains A (IL-1β) and B (IL-1β type 1 receptor) in the 1itb structure from the PDB database	–29.35
3kxyGH	Interaction of chains G (Protein C exoenzyme S synthesis) and H (Protein C exoenzyme S synthesis) in the 3kxy structure from the PDB database	–13.68
1iraXY	Interaction of chains X (IL-1 receptor antagonist) and Y (IL-1 receptor) in the 1ira structure from the PDB database	–5.03
1itbAB	Interaction of chains A (IL-1β) and B (IL-1β type 1 receptor) in the 1itb structure from the PDB database	–4.62

Table 3. Interfaces (binding regions) and interaction energies of IL-1RA with the rs315952 polymorphism with IL-1R1

Interface (binding region)	Decoding of the binding domain name	Energy, kJ/mol
1iraXY	Interaction of chains X (IL-1 receptor antagonist) and Y (IL-1 receptor) in the 1ira structure from the PDB database	–30.08
1itbAB	Interaction of chains A (IL-1β) and B (IL-1β type 1 receptor) in the 1itb structure from the PDB database	–28.18
1iraXY	Interaction of chains X (IL-1 receptor antagonist) and Y (IL-1 receptor) in the 1ira structure from the PDB database	–18.28
3kxyGH	Interaction of chains G (Protein C exoenzyme S synthesis) and H (Protein C exoenzyme S synthesis) in the 3kxy structure from the PDB database	–6.62
1itbAB	Interaction of chains A (IL-1β) and B (IL-1β type 1 receptor) in the 1itb structure from the PDB database	–6.38

In the results of modeling this interface, minimum energy values were found for the 1iraXY template (Table 3). However, as the interaction without the polymorphism in the IL-1RA structure (IL-1RA+IL-1R1) showed minimal interaction energy according to the 1itbAB template, the energy should be compared using the same template.

Comparison between Tables 2 and 3 suggests that the interaction of IL-1RA with the rs315952 polymorphism with IL-1R1 (–30.08 kJ/mol) is the most energetically advantageous. However, when comparing energies using the same 1itbAB template, this interaction becomes less energetically favorable (–29.35 kJ/mol and –28.18 kJ/mol, respectively). This suggests that in the presence of the rs315952 polymorphism in IL1-RA (serine substitution at position 130 for arginine (Table 4)), the interleukin-receptor interaction complex weakens, becoming less stable and more susceptible to decay.

Thus, based on the interactions of IL-1RA with the rs315952 polymorphism with IL-1R1, we cannot draw a definitive conclusion regarding the polymorphism's impact on its involvement in initiating the IL-1 signaling pathway. However, the modeled interactions indicate that the polymorphism participates in the formation of the protein-protein complex.

Modeling interactions of IL-1RA with IL-1RAP

Two investigations were conducted for modeling interactions: the interaction of standard IL-1RA with IL-1RAP and IL-1RA with the rs315952 polymorphism with IL-1RAP. In both cases, the algorithm did not create a model of protein interaction.

Modeling interactions of IL-1RA with the IL-1R1+IL-1RAP complex

Modeling interaction of standard IL-1RA with the IL-1R1+IL-1RAP complex. The interaction of IL-1RA with the receptor complex using the 1iraXY template showed a stable interaction (–34.27 kJ/mol) (Table 5). However, analyzing the interaction using the 1itbAB template revealed a very weak interaction (–2.67 kJ/mol).

According to the contact list of template residues, serine at position 130 is also a hotspot (Table 7).

The results in Table 5 indicate that the 1itbAB template is suitable for interaction with the added IL-1RAP protein, but its stability is almost minimal, implying the formed complex will quickly break down. Therefore, for further interaction analysis, we use the 1iraXY template.

Modeling interaction of IL-1RA with the rs315952 polymorphism with the IL-1R1+IL-1RAP complex. According to the contact list of template residues, arginine is also a hotspot. The simulation results presented in Table 6 show that the minimum energy of the complex is observed with the 1iraXY template at –25.27 kJ/mol.

Comparing interactions with the complex without polymorphism using the 1iraXY template, it is evident that the

Н.Ю. Часовских, А.А. Бобрышева
Е.Е. Чижик

Компьютерное моделирование особенностей
взаимодействий IL-1 с его рецепторами при шизофрении

2024
28•3

Table 4. Contacts of interface residues of 1itbAB IL-1RA and IL-1R1 and 1itbAB IL-1RA with SNP rs315952 and IL-1R1

1itbAB IL-1RA and IL-1R1		1itbAB IL-1RA with SNP rs315952 and IL-1R1	
IL-1RA	IL-1R1	IL-1RA with SNP	IL-1R1
SER_97	ASN_30	ARG_51	GLU_11
VAL_95	PRO_31	ASN_64	GLN_108
GLU_100	PRO_28	GLN_154	PRO_31
PRO_132	SER_263	GLN_61	TYR_127
LYS_121	ARG_163	ARG_117	ASP_251
LYS_121	ASP_162	ARG_117	GLU_252
TRP_144	ILE_13	ARG_117	ASP_253
GLY_131	LEU_237	**ARG_130**	**THR_300**
GLN_119	ARG_163	GLU_77	ILE_250
ASP_99	LEU_15	ARG_51	ILE_13
PRO_142	VAL_124	GLN_45	LEU_123
TRP_144	TYR_127	GLN_45	VAL_124
LYS_96	GLU_129	LYS_34	LEU_237
ASP_120	ARG_163	LEU_60	TYR_127
SER_97	CYS_27	LYS_34	SER_263
CYS_141	LEU_115	GLU_77	VAL_249
ASP_120	VAL_124	GLU_77	ILE_240
SER_97	PRO_28	TYR_59	GLN_113
SER_97	LEU_29	PRO_78	ILE_240
GLY_98	CYS_27	GLY_131	THR_300
GLY_98	PRO_26	GLN_61	LEU_15
GLU_100	ILE_13	GLN_61	ILE_13
GLY_98	LEU_29	PRO_63	ILE_13
GLY_98	PRO_28	GLY_62	ILE_13
PRO_142	GLN_113	PRO_63	ILE_14
SER_130	**LEU_237**	GLN_45	GLN_113
ASP_99	PRO_28		
ASN_160	ILE_13		
ASP_99	PRO_26		

minimum energy without polymorphism is –34.27 kJ/mol, while with polymorphism it is –25.27 kJ/mol. Thus, it can be hypothesized that the studied rs315952 polymorphism affects the formation of IL-1RA binding with the IL-1R1+IL-1RAP receptor complex, creating a less stable complex more prone to decay.

The study results allow us to make an assumption that the p.Ser130Arg mutation in the IL-1RA protein gene may lead to the formation of a weakened complex between IL-1RA and the associated receptors IL-1R1+IL-1RAP, which could impact schizophrenia mechanisms.

Discussion

The functions of IL-1 family molecules are primarily associated with innate immunity. While inflammation normally acts as a protective mechanism, it can cause damage to the body when it becomes uncontrollable (Dinarello, 2018). IL-1 has been implicated in neuronal cell damage (Allan et al., 2005), and excessive phagocytosis may contribute to pathologies in Alzheimer's disease, schizophrenia, and aging (Vilalta, Brown, 2018). IL-1 triggers phagocytosis in the brain by acting as a chemoattractant for neutrophils. Initiating the IL-1 signaling pathway also leads to the release of cytokines

Table 5. Interfaces (binding regions) and interaction energies of standard IL-1RA with the IL-1R1+IL-1RAP protein complex

Interface (binding region)	Decoding of the binding domain name	Energy, kJ/mol
1iraXY	Interaction of chains X (IL-1 receptor antagonist) and Y (IL-1 receptor) in the 1ira structure from the PDB database	–34.27
3fmpCD	Interaction of chains C (Nuclear pore complex protein Nup214) and D (ATP-dependent RNA helicase DDX19B) in the 3fmp structure from the PDB database	–15.92
1itbAB	Interaction of chains A (IL-1β) and B (IL-1β type 1 receptor) in the 1itb structure from the PDB database	–2.67

Table 6. Interface (binding region) and interaction energies of IL-1RA with the rs315952 polymorphism with the IL-1R1+IL-1RAP protein complex

Interface (binding region)	Decoding of the binding domain name	Energy, kJ/mol
1iraXY	Interaction of chains X (IL-1 receptor antagonist) and Y (IL-1 receptor) in the 1ira structure from the PDB database	–25.27
1iraXY	Interaction of chains X (IL-1 receptor antagonist) and Y (IL-1 receptor) in the 1ira structure from the PDB database	–19.03
1iraXY	Interaction of chains X (IL-1 receptor antagonist) and Y (IL-1 receptor) in the 1ira structure from the PDB database	–15.47
1itbAB	Interaction of chains A (IL-1β) and B (IL-1β type 1 receptor) in the 1itb structure from the PDB database	–12.36

TNFα and IFN-γ, which activate macrophages (Sasayama et al., 2011).

Studies confirm an increase in the level of IL-1 in the blood of individuals with schizophrenia (Chu et al., 2018; Zhou et al., 2019). The reporter system of genetic knockout in mice, used to track the reciprocal deletion or expression of the IL-1 receptor (IL-1R1) in endothelial cells, ventricles, peripheral myeloid cells, microglia, astrocytes, and neurons, revealed that endothelial IL-1R1 is necessary and sufficient for mediating pain behavior. It is also shown to stimulate the proliferation of leukocytes in the central nervous system (CNS) and attenuate neurogenesis. Ventricular IL-1R1 is critical for the proliferation of monocytes in the CNS. Although microglia does not express IL-1R1, stimulation of endothelial cells with IL-1 leads to the induction of IL-1 in microglia (Liu et al., 2019).

The IL-1RA protein, which is an antagonist of IL-1 receptors and has an anti-inflammatory function (Dinarello, 1994), has also been found to be associated with schizophrenia (Kim et al., 2004). Preliminary results suggest that the IL-1RA protein gene may contribute to the ventricular changes observed in patients with this disease (Papiol et al., 2005).

An association has been found between single nucleotide polymorphisms in proteins involved in the IL-1 pathway and the risk of developing schizophrenia (Xu, He, 2010). There is a tendency for the association of the GAGG haplotype (rs1143627, rs16944, rs1143623, rs4848306) of the *ILB* gene; TG haplotypes (rs315952, rs9005) and TT61 rs5254 (rs4) of *IL1RN*, and CT haplotype (rs4251961, rs419598) in *IL1RN* with the risk of schizophrenia. Statistically significant association is shown for rs1143634 (*IL1B* gene; T3953C). This suggests a connection between pro-inflammatory fac-

tors, specifically polymorphisms in genes initiating the IL1 pathway, and the development of this disorder (Xu, He, 2010; Kapelski et al., 2016).

IL-1RA, acting as an antagonist to the IL-1 receptor, exhibits anti-inflammatory properties. In turn, IL-1α and IL-1β, by binding to the IL-1 receptor, initiate the IL-1 signaling pathway, participating in the implementation of the inflammatory response. Elevated synthesis of IL-1RA blocks this pathway, leading to inhibition of the immune response and weakening of the inflammatory process.

In the analysis of the interaction of the studied proteins, no differences in energy outputs were observed between standard IL-1RA and IL-1RA with rs315952 interacting with IL-1R1. When standard IL-1RA interacts with the IL-1R1+IL-1RAP complex, a lower energy value is observed compared to the case with the polymorphism, presumably indicating a weakening of the interaction between IL-1RA and IL-1R1+IL-1RAP. Notably, IL-1RA does not interact separately with IL-1RAP.

IL-1RA protein, upon binding to IL-1R1 and IL-1R1+IL-1RAP, inhibits the binding of IL-1 and, consequently, the activation of the IL-1 signaling pathway (Weber et al., 2010). In schizophrenia, the appearance of a single nucleotide polymorphism in the *IL-1RA* gene (p.Ser130Arg) may lead to the formation of a weakened complex between IL-1RA and associated receptors IL-1R1+IL-1RAP. This, presumably, could subsequently trigger the IL-1 signaling pathway and, as a result, the development of an uncontrolled immune response.

The results of the study showed that the functions of interleukin-1, namely the interactions of IL-1 family proteins, may be associated with structural changes in the corresponding genes. The analysis of SNP associations of these genes with

Н.Ю. Часовских, А.А. Бобрышева
Е.Е. Чижик

Компьютерное моделирование особенностей
взаимодействий IL-1 с его рецепторами при шизофрении

2024
28•3

Table 7. Contacts of residues in the interface of 1itbAB standard IL-1RA with the complex of IL-1R1+IL-1RAP proteins
and 1itbAB IL-1RA with the rs315952 polymorphism with the complex of IL-1R1+IL-1RAP proteins

1itbAB standard IL-1RA with the complex of IL-1R1+IL-1RAP proteins		1itbAB IL-1RA with the rs315952 polymorphism with the complex of IL-1R1+IL-1RAP proteins	
IL-1RA	IL-1R1+IL-1RAP	IL-1RA with the rs315952 polymorphism	IL-1R1+IL-1RAP
pdb1_A_SER_130	pdb2_C_ILE_184	pdb1_A_ARG_130	pdb2_C_ILE_184
pdb1_A_ARG_51	pdb2_B_GLU_11	pdb1_A_ARG_51	pdb2_B_GLU_11
pdb1_A_MET_150	pdb2_C_ASN_168	pdb1_A_ARG_51	pdb2_B_ILE_13
pdb1_A_MET_150	pdb2_B_ARG_163	pdb1_A_ASN_64	pdb2_B_ALA_109
pdb1_A_ASN_116	pdb2_B_GLU_252	pdb1_A_ASN_64	pdb2_B_ALA_107
pdb1_A_GLY_165	pdb2_C_PHE_167	pdb1_A_ASN_44	pdb2_B_VAL_124
pdb1_A_MET_161	pdb2_C_TYR_162	pdb1_A_GLN_154	pdb2_B_PRO_31
pdb1_A_TYR_59	pdb2_B_TYR_127	pdb1_A_PRO_63	pdb2_B_ILE_110
pdb1_A_ARG_117	pdb2_B_ASP_251	pdb1_A_PRO_63	pdb2_B_PHE_111
pdb1_A_GLN_154	pdb2_B_PRO_31	pdb1_A_ARG_117	pdb2_B_ASP_251
pdb1_A_SER_130	**pdb2_B_THR_300**	pdb1_A_ARG_117	pdb2_B_GLU_252
pdb1_A_LYS_170	pdb2_C_SER_185	pdb1_A_GLU_77	pdb2_B_GLU_259
pdb1_A_ASP_129	pdb2_B_THR_300	pdb1_A_ARG_130	pdb2_B_THR_300
pdb1_A_ASP_163	pdb2_C_MET_159	pdb1_A_GLU_77	pdb2_B_ILE_250
pdb1_A_ASP_163	pdb2_C_SER_185	pdb1_A_LYS_170	pdb2_C_SER_185
pdb1_A_VAL_43	pdb2_B_LYS_114	pdb1_A_ASP_129	pdb2_B_THR_300
pdb1_A_GLU_164	pdb2_C_MET_159	pdb1_A_ASN_64	pdb2_B_GLN_108
pdb1_A_LYS_170	pdb2_C_LEU_183	pdb1_A_GLU_164	pdb2_C_TYR_162
pdb1_A_GLU_164	pdb2_C_TYR_162	pdb1_A_GLN_45	pdb2_B_VAL_124
pdb1_A_GLU_77	pdb2_B_VAL_249	pdb1_A_GLN_45	pdb2_B_PRO_126
pdb1_A_TYR_59	pdb2_B_PHE_111	pdb1_A_GLU_175	pdb2_B_LEU_237
pdb1_A_LEU_67	pdb2_B_ILE_13	**pdb1_A_ARG_130**	**pdb2_C_ARG_286**
pdb1_A_GLU_77	pdb2_B_TYR_242	pdb1_A_LYS_34	pdb2_B_LEU_237
pdb1_A_GLY_131	pdb2_B_THR_300	pdb1_A_ASN_64	pdb2_B_ILE_110
pdb1_A_GLY_165	pdb2_C_MET_159	pdb1_A_LYS_34	pdb2_B_SER_263
pdb1_A_GLN_61	pdb2_B_GLU_11	pdb1_A_VAL_43	pdb2_C_LEU_183
pdb1_A_GLN_61	pdb2_B_ILE_13	pdb1_A_GLU_77	pdb2_B_VAL_249
pdb1_A_PRO_63	pdb2_B_LYS_12	pdb1_A_ASP_153	pdb2_B_PRO_31
pdb1_A_PRO_63	pdb2_B_ILE_13	pdb1_A_GLU_77	pdb2_B_TYR_261
pdb1_A_ARG_30	pdb2_B_ASP_260	pdb1_A_GLU_77	pdb2_B_ILE_240
pdb1_A_GLY_62	pdb2_B_ILE_13	pdb1_A_PRO_78	pdb2_B_ILE_240
pdb1_A_LYS_170	pdb2_C_ASN_168	pdb1_A_GLY_131	pdb2_B_THR_300
pdb1_A_PRO_63	pdb2_B_ILE_14	pdb1_A_HIS_79	pdb2_B_ILE_240
pdb1_A_GLN_45	pdb2_B_PHE_111	pdb1_A_GLN_61	pdb2_B_ILE_13
pdb1_A_GLN_45	pdb2_B_LYS_112	pdb1_A_VAL_65	pdb2_B_LYS_112
pdb1_A_GLN_45	pdb2_B_GLN_113	pdb1_A_PRO_63	pdb2_B_ILE_13
		pdb1_A_GLY_62	pdb2_B_ILE_13
		pdb1_A_PRO_63	pdb2_B_ILE_14
		pdb1_A_ASN_44	pdb2_C_ASN_168
		pdb1_A_GLN_45	pdb2_B_GLN_113
		pdb1_A_GLY_165	pdb2_C_TYR_162

schizophrenia, together with information about the influence of inflammation on the mechanisms of its development, can serve as a theoretical basis for a more detailed and careful study of the mechanisms of the inflammatory response.

Conclusion

It is known that *in silico* mutagenesis and the comparison of changes in interaction energies between the standard and mutated variants shed light on the mechanisms underlying the development of several diseases. The results obtained in this study demonstrate that in schizophrenia, structural changes in genes may influence the functions of interleukin-1 (protein interactions within the IL-1 family). This, in turn, allows correlating existing data on the impact of inflammation on the development of schizophrenia with associations of SNPs in genes related to the IL-1 family. The conducted research makes a theoretical contribution to the understanding of the details of the mechanisms involved in the inflammatory response in schizophrenia, and the results may serve as a basis for further studies (both *in silico* and experimental) in this field.

References

Acuner Ozbabacan S.E., Gursoy A., Nussinov R., Keskin O. The structural pathway of interleukin 1 (IL-1) initiated signaling reveals mechanisms of oncogenic mutations and SNPs in inflammation and cancer. *PLoS Comput. Biol*. 2014;10(2):e1003470. DOI 10.1371/journal.pcbi.1003470

Allan S.M., Tyrrell P.J., Rothwell N.J. Interleukin-1 and neuronal injury. *Nat. Rev. Immunol*. 2005;5(8):629-640. DOI 10.1038/nri1664

Aytuna A.S., Gursoy A., Keskin O. Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. *Bioinformatics*. 2005;21(12):2850-2855. DOI 10.1093/bioinformatics/bti443

Baspinar A., Cukuroglu E., Nussinov R., Keskin O., Gursoy A. PRISM: a web server and repository for prediction of protein-protein interactions and modeling their 3D complexes. *Nucleic Acids Res*. 2014; 42(W1):W285-W289. DOI 10.1093/nar/gku397

Ben Nejma M., Zaabar I., Zaafrane F., Thabet S., Mechri A., Gaha L., Ben Salem K., Bel Hadj Jrad B. A gender-specific association of interleukin 1 receptor antagonist polymorphism with schizophrenia susceptibility. *Acta Neuropsychiatr*. 2013;25(6):349-355. DOI 10.1017/neu.2012.32

Bochkov N.P. Clinical Genetics. Moscow: GEOTAR-Media, 2011 (in Russian)

Chu C.S., Li D.J., Chu C.L., Wu C.C., Lu T. Decreased IL-1ra and NCAM-1/CD56 serum levels in unmedicated patients with schizophrenia before and after antipsychotic treatment. *Psychiatry Investig*. 2018;15(7):727-732. DOI 10.30773/pi.2017.11.10

Dinarello C.A. The interleukin-1 family: 10 years of discovery. *FASEB J*. 1994;8(15):1314-1325

Dinarello C.A. Overview of the IL-1 family in innate inflammation and acquired immunity. *Immunol. Rev*. 2018;281(1):8-27. DOI 10.1111/imr.12621

Fatjó-Vilas M., Pomarol-Clotet E., Salvador R., Monté G.C., Gomar J.J., Sarró S., Ortiz-Gil J., Aguirre C., Landín-Romero R., Guerrero-Pedraza A., Papiol S., Blanch J., McKenna P.J., Fañanás L. Effect of the interleukin-1β gene on dorsolateral prefrontal cortex function in schizophrenia: a genetic neuroimaging study. *Biol. Psychiatry*. 2012;72(9):758-765. DOI 10.1016/j.biopsych.2012.04.035

Frodl T., Amico F. Is there an association between peripheral immune markers and structural/functional neuroimaging findings? *Prog. Neuropsychopharmacol. Biol. Psychiatry*. 2014;48:295-303. DOI 10.1016/j.pnpbp.2012.12.013

Gao M., Skolnick J. New benchmark metrics for protein-protein docking methods. *Proteins*. 2011;79(5):1623-1634. DOI 10.1002/prot.22987

Hadarovich A.Y., Kalinouski A.A., Tuzikov A.V. Protein homodimers structure prediction based on deep neural network. *Informatika = Informatics*. 2020;17(2):44-53. DOI 10.37661/1816-0301-2020-17-2-44-53 (in Russian)

Hudson Z.D., Miller B.J. Meta-analysis of cytokine and chemokine genes in schizophrenia. *Clin. Schizophr. Relat. Psychoses*. 2018; 12(3):121-129B. DOI 10.3371/CSRP.HUMI.070516

Kapelski P., Skibinska M., Maciukiewicz M., Wilkosc M., Frydecka D., Groszewska A., Narozna B., Dmitrzak-Weglarz M., Czerski P., Pawlak J., Rajewska-Rager A., Leszczynska-Rodziewicz A., Slopien A., Zaremba D., Twarowska-Hauser J. Association study of functional polymorphisms in interleukins and interleukin receptors genes: IL1A, IL1B, IL1RN, IL6, IL6R, IL10, IL10RA and TGFB1 in schizophrenia in Polish population. *Schizophr. Res*. 2015;169(1-3):1-9. DOI 10.1016/j.schres.2015.10.008

Kapelski P., Skibinska M., Maciukiewicz M., Pawlak J., Dmitrzak-Weglarz M., Szczepankiewicz A., Zaremba D., Twarowska-Hauser J. An association between functional polymorphisms of the interleukin 1 gene complex and schizophrenia using transmission disequilibrium test. *Arch. Immunol. Ther. Exp.* (*Warsz.*). 2016;64(Suppl.1):161-168. DOI 10.1007/s00005-016-0434-6

Katila H., Hänninen K., Hurme M. Polymorphisms of the interleukin-1 gene complex in schizophrenia. *Mol. Psychiatry*. 1999;4(2):179-181. DOI 10.1038/sj.mp.4000483

Kim S.J., Lee H.J., Koo H.G., Kim J.W., Song J.Y., Kim M.K., Shin D.H., Jin S.Y., Hong M.S., Park H.J., Yoon S.H., Park H.K., Chung J.H. Impact of IL-1 receptor antagonist gene polymorphism on schizophrenia and bipolar disorder. *Psychiatr Genet*. 2004;14(3):165-167. DOI 10.1097/00041444-200409000-00009

Krieger E., Vriend G. YASARA View – molecular graphics for all devices – from smartphones to workstations. *Bioinformatics*. 2014; 30(20):2981-2982. DOI 10.1093/bioinformatics/btu426

Kuzu G., Gursoy A., Nussinov R., Keskin O. Exploiting conformational ensembles in modeling protein-protein interactions on the proteome scale. *J. Proteome Res*. 2013;12(6):2641-2653. DOI 10.1021/pr400006k

Liu X., Nemeth D.P., McKim D.B., Zhu L., DiSabato D.J., Berdysz O., Gorantla G., Oliver B., Witcher K.G., Wang Y., Negray C.E., Vegesna R.S., Sheridan J.F., Godbout J.P., Robson M.J., Blakely R.D., Popovich P.G., Bilbo S.D., Quan N. Cell-type-specific interleukin 1 receptor 1 signaling in the brain regulates distinct neuroimmune activities. *Immunity*. 2019;50(2):317-333.e6. DOI 10.1016/j.immuni.2018.12.012

McClay J.L., Adkins D.E., Aberg K., Bukszár J., Khachane A.N., Keefe R.S., Perkins D.O., McEvoy J.P., Stroup T.S., Vann R.E., Beardsley P.M., Lieberman J.A., Sullivan P.F., van den Oord E.J. Genome-wide pharmacogenomic study of neurocognition as an indicator of antipsychotic treatment response in schizophrenia. *Neuropsychopharmacology*. 2011;36(3):616-626. DOI 10.1038/npp.2010.193

Miyaoka T., Wake R., Hashioka S., Hayashida M., Oh-Nishi A., Azis I.A., Izuhara M., Tsuchie K., Araki T., Arauchi R., Abdullah R.A., Horiguchi J. Remission of psychosis in treatment-resistant schizophrenia following bone marrow transplantation: a case report. *Front. Psychiatry*. 2017;8:174. DOI 10.3389/fpsyt.2017.00174

Müller N. COX-2 inhibitors, aspirin, and other potential anti-inflammatory treatments for psychiatric disorders. *Front. Psychiatry*. 2019; 10:375. DOI 10.3389/fpsyt.2019.00375

Papiol S., Molina V., Desco M., Rosa A., Reig S., Gispert J.D., Sanz J., Palomo T., Fañanás L. Ventricular enlargement in schizophrenia is associated with a genetic polymorphism at the interleukin-1 receptor antagonist gene. *Neuroimage*. 2005;27(4):1002-1006. DOI 10.1016/j.neuroimage.2005.05.035

Papiol S., Molina V., Rosa A., Sanz J., Palomo T., Fañanás L. Effect of interleukin-1β gene functional polymorphism on dorsolateral prefrontal cortex activity in schizophrenic patients. *Am. J. Med. Genet. B Neuropsychiatr. Genet*. 2007;144B(8):1090-1093. DOI 10.1002/ajmg.b.30542

Н.Ю. Часовских, А.А. Бобрышева
Е.Е. Чижик

Компьютерное моделирование особенностей
взаимодействий IL-1 с его рецепторами при шизофрении

2024
28•3

Piñero J., Ramírez-Anguita J.M., Saüch-Pitarch J., Ronzano F., Centeno E., Sanz F., Furlong L.I. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res*. 2020;48(D1): D845-D855. DOI 10.1093/nar/gkz1021

Sasayama D., Hori H., Teraishi T., Hattori K., Ota M., Iijima Y., Tatsumi M., Higuchi T., Amano N., Kunugi H. Possible association between interleukin-1β gene and schizophrenia in a Japanese population. *Behav. Brain Funct*. 2011;7:35. DOI 10.1186/1744-9081-7-35

Sayers E.W., Beck J., Bolton E.E., Bourexis D., Brister J.R., Canese K., Comeau D.C., Funk K., Kim S., Klimke W., Marchler-Bauer A., Landrum M., Lathrop S., Lu Z., Madden T.L., O'Leary N., Phan L., Rangwala S.H., Schneider V.A., Skripchenko Y., Wang J., Ye J., Trawick B.W., Pruitt K.D., Sherry S.T. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2021;49(D1):D10-D17. DOI 10.1093/nar/gkaa892

Sherry S.T., Ward M.H., Kholodov M., Baker J., Phan L., Smigielski E.M., Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308-311. DOI 10.1093/nar/29.1.308

Shirts B.H., Wood J., Yolken R.H., Nimgaonkar V.L. Association study of IL10, IL1β, and IL1RN and schizophrenia using tag SNPs from a comprehensive database: suggestive association with rs16944 at IL1β. *Schizophr. Res*. 2006;88(1-3):235-244. DOI 10.1016/j.schres.2006.06.037

Sommer I.E., van Bekkum D.W., Klein H., Yolken R., de Witte L., Talamo G. Severe chronic psychosis after allogeneic SCT from a schizophrenic sibling. *Bone Marrow Transplant*. 2015;50(1):153-154. DOI 10.1038/bmt.2014.221

Tuncbag N., Gursoy A., Nussinov R., Keskin O. Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat. Protoc*. 2011;6(9):1341-1354. DOI 10.1038/nprot.2011.367

Tuncbag N., Keskin O., Nussinov R., Gursoy A. Fast and accurate modeling of protein-protein interactions by combining template-interface docking with flexible refinement. *Squirrels*. 2012;80(4):1239-1249. DOI 10.1002/prot.24022

Vilalta A., Brown G.C. Neurophagy, the phagocytosis of live neurons and synapses by glia, contributes to brain development and disease. *FEBS J*. 2018;285(19):3566-3575. DOI 10.1111/febs.14323

Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G., Gumienny R., Heer F.T., Beer T.A.P., Rempfer C., Bordoli L., Lepore R., Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303. DOI 10.1093/nar/gky427

Weber A., Wasiliew P., Kracht M. Interleukin-1 (IL-1) pathway. *Sci. Signal*. 2010;3(105):cm1. DOI 10.1126/scisignal.3105cm1

Xu M., He L. Convergent evidence shows a positive association of interleukin-1 gene complex locus with susceptibility to schizophrenia in the Caucasian population. *Schizophr. Res*. 2010;120(1-3):131-142. DOI 10.1016/j.schres.2010.02.1031

Yoshida M., Shiroiwa K., Mouri K., Ishiguro H., Supriyanto I., Ratta-Apha W., Eguchi N., Okazaki S., Sasada T., Fukutake M., Hashimoto T., Inada T., Arinami T., Shirakawa O., Hishimoto A. Haplotypes in the expression quantitative trait locus of interleukin-1β gene are associated with schizophrenia. *Schizophr. Res*. 2012;140(1-3):185-191. DOI 10.1016/j.schres.2012.06.031

Zanardini R., Bocchio-Chiavetto L., Scassellati C., Bonvicini C., Tura G.B., Rossi G., Perez J., Gennarelli M. Association between IL-1β-511C/T and IL-1RA (86bp)$_n$ repeats polymorphisms and schizophrenia. *J. Psychiatr. Res*. 2003;37(6):457-462. DOI 10.1016/s0022-3956(03)00072-4

Zhou Y., Peng W., Wang J., Zhou W., Zhou Y., Ying B. Plasma levels of IL-1Ra are associated with schizophrenia. *Psychiatry. Clin. Neurosci*. 2019;73(3):109-115. DOI 10.1111/pcn.12794

mRNA-lncRNA gene expression signature in HPV-associated neoplasia and cervical cancer

E.D. Kulaeva (ID) (✉), E.S. Muzlaeva (ID), E.V. Mashkina (ID)

Southern Federal University, Rostov-on-Don, Russia
✉ ked05685@gmail.com

Abstract. Cervical cancer is one of the most frequent cancers in women and is associated with human papillomavirus (HPV) in 70 % of cases. Cervical cancer occurs because of progression of low-differentiated cervical intraepithelial neoplasia through grade 2 and 3 lesions. Along with the protein-coding genes, long noncoding RNAs (lncRNAs) play an important role in the development of malignant cell transformation. Although human papillomavirus is widespread, there is currently no well-characterized transcriptomic signature to predict whether this tumor will develop in the presence of HPV-associated neoplastic changes in the cervical epithelium. Changes in gene activity in tumors reflect the biological diversity of cellular phenotype and physiological functions and can be an important diagnostic marker. We performed comparative transcriptome analysis using open RNA sequencing data to assess differentially expressed genes between normal tissue, neoplastic epithelium, and cervical cancer. Raw data were preprocessed using the Galaxy platform. Batch effect correction, identification of differentially expressed genes, and gene set enrichment analysis (GSEA) were performed using R programming language packages. Subcellular localization of lncRNA was analyzed using Locate-R and iLoc-LncRNA 2.0 web services. 1,572 differentially expressed genes (DEGs) were recorded in the "cancer *vs.* control" comparison, and 1,260 DEGs were recorded in the "cancer *vs.* neoplasia" comparison. Only two genes were observed to be differentially expressed in the "neoplasia *vs.* control" comparison. The search for common genes among the most strongly differentially expressed genes among all comparison groups resulted in the identification of an expression signature consisting of the *CCL20*, *CDKN2A*, *CTCFL*, *piR-55219*, *TRH*, *SLC27A6* and *EPHA5* genes. The transcription level of the *CCL20* and *CDKN2A* genes becomes increased at the stage of neoplastic epithelial changes and stays so in cervical cancer. Validation on an independent microarray dataset showed that the differential expression patterns of the *CDKN2A* and *SLC27A6* genes were conserved in the respective gene expression comparisons between groups.

Key words: human papillomavirus; neoplasia; cervical cancer; transcriptome analysis; lncRNA; *CDKN2A*; *CCL20*.

Профиль экспрессии мРНК-днРНК при неоплазии и цервикальном раке, ассоциированными с ВПЧ-инфекцией

Е.Д. Кулаева (ID) (✉), Е.С. Музлаева (ID), Е.В. Машкина (ID)

Южный федеральный университет, Ростов-на-Дону, Россия
✉ ked05685@gmail.com

Аннотация. Рак шейки матки является одним из наиболее частых онкологических заболеваний у женщин и в 70 % случаев связан с вирусом папилломы человека (ВПЧ). Рак шейки матки развивается в результате прогрессии цервикальной интраэпителиальной неоплазии через поражения второй и третьей степени. Помимо белок-кодирующих генов, важную роль в развитии злокачественной трансформации клеток играют длинные некодирующие РНК. Хотя вирус папилломы человека широко распространен, в настоящее время нет хорошо охарактеризованных транскриптомных признаков, позволяющих предсказать злокачественную трансформацию клеток эпителия при наличии связанной с ВПЧ неоплазии эпителия шейки матки. Изменения генной активности в опухолях отражают биологическое разнообразие клеточного фенотипа и физиологических функций и могут быть важным диагностическим маркером. Используя открытые данные секвенирования РНК, мы провели сравнительный анализ транскриптома для оценки дифференциально экспрессируемых генов в образцах нормальной ткани, эпителия с диспластическими изменениями и раком шейки матки. Первичные данные были предварительно обработаны с использованием платформы Galaxy. Коррекция пакетного эффекта, идентификация дифференциально экспрессируемых генов и анализ обогащения набора генов выполнены в пакетах языка программирования R. Субклеточная локализация днРНК была проанализирована с помощью веб-сервисов Locate-R и iLoc-LncRNA 2.0. В сравнении «рак *vs.* контроль» зарегистрировано 1572 дифференциально экспрессируемых гена, в сравнении «рак *vs.* неоплазия» – 1260. Только два дифференциально экспрессируемых гена выявлено при сравнении контро-

Е.Д. Кулаева, Е.С. Музлаева
Е.В. Машкина

Профиль экспрессии мРНК-днРНК при неоплазии
и цервикальном раке, ассоциированными с ВПЧ-инфекцией

2024
28•3

ля и неоплазии. Поиск общих среди наиболее сильно дифференциально экспрессируемых генов во всех группах сравнения привел к выявлению сигнатуры экспрессии, состоящей из генов *CCL20*, *CDKN2A*, *CTCFL*, *piR-55219*, *TRH*, *SLC27A6* и *EPHA5*. Повышенный уровень транскрипции генов *CCL20* и *CDKN2A* возникает на стадии неопластических изменений эпителия и сохраняется при раке шейки матки. Валидация на независимом наборе данных микрочипа показала, что паттерны дифференциальной экспрессии генов *CDKN2A* и *SLC27A6* сохраняются в соответствующих сравнениях экспрессии генов между группами.

Ключевые слова: вирус папилломы человека; неоплазия; рак шейки матки; транскриптомный анализ; lncRNA; *CDKN2A*; *CCL20*.

Introduction

Cervical cancer is the fourth most common cancer in women worldwide after breast cancer, colorectal cancer, and lung cancer. The World Health Organization (WHO) estimates that 604,127 new cases and 341,831 deaths from the disease worldwide were registered in 2020 (Sung et al., 2021; Gebrie, 2022). Cervical cancer occurs as a result of progression of low-differentiated cervical intraepithelial neoplasia (CIN1) through grade 2 and 3 lesions (CIN2 and CIN3). Inflammatory responses are rarely observed in persistent low-grade lesions and are thought to be due to the inflammation-suppressing activity of high-risk HPV oncoproteins (Walch-Ruckheim et al., 2015).

Although HPV is the most significant factor in cervical cancer, the development of cervical cancer is considered multifactorial. Common risk factors for cervical cancer also include smoking, a high number of sexual partners, low social and/or economic status and its consequences, and immune suppression caused by infection such as human immunodeficiency virus (HIV) or the use of immunosuppressants after organ transplantation (Walch-Ruckheim et al., 2015).

According to the International Human Papillomavirus Reference Center data (Eklund et al., 2020), only 12 out of 220 HPV strains have the greatest impact on cancer development (these strains include HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59). About 70 % of cervical cancer and precancerous lesions of the cervix cases are specifically associated with HPV types 16 and 18 (Okunade, 2020).

Eighty percent of sexually active women become infected with HPV during their lifetime, but the infection persists in only 5–10 % of those initially infected and leads to cervical cancer in only 3 % (Schubert et al., 2023). In the absence of a clearly persistent HPV infection, the risk of developing cervical cancer is extremely low. However, virus persistence may be associated with many factors. Host genetic factors are thought to play an important role in the response to HPV infection and further development of oncology.

Along with the protein-coding genes, long noncoding RNAs (lncRNAs) play an important role in the development of malignant cell transformation. Results of the TCGA project showed that approximately the same number of protein-coding genes and lncRNA genes carried mutations in more than 5,000 different tumor samples. However, at the same time, 60 % of lncRNAs showed tumor type specificity and are superior to protein-coding genes in terms of specificity to the type of cancer (Yan et al., 2015). Consequently, lncRNAs can be a good class of biomarkers for cancer prognosis and early diagnosis.

Both protein-coding and lncRNAs can be analyzed as efficiently as possible by high-throughput RNA sequencing (RNA-seq). Profiling the entire transcriptome can iden-

tify genes that are differentially expressed in related tissues. Changes in gene activity in tumors reflect the biological diversity of cellular phenotype and physiological functions and can be an important diagnostic marker (Martin, Wang, 2011; Bao et al., 2019). A significant change in the expression of both protein-coding and non-coding parts of the genome may be a consequence of local chromatin remodeling in the region of the virus integration site, which plays a somewhat spontaneous but often important role in oncogenesis (Karimzadeh et al., 2023).

The aim of this work was to perform bioinformatics analysis of RNA sequencing data from epitheliocytes of women with cervical epithelial neoplasia and cervical cancer based on open data from three different studies (Royse et al., 2014; Hu et al., 2015; Qi et al., 2022).

Materials and methods

Datasets. The study material was raw RNA sequencing data of cervical epithelial samples from three separate studies analyzing the transcriptome in cervical cancer, neoplasia, and normal tissue. Neoplasia grade data were also available. The main characteristics of the studies used are summarized in Table 1.

Data preprocessing. Raw RNA-seq data (fastq format) were processed using the Galaxy platform (https://usegalaxy.org/). Read quality was assessed with FastQC, adapter trimming was performed with TrimGalore, transcript alignment and mapping was performed with RNA STAR, and transcript counting was performed with featureCounts, respectively.

Data variability analysis and batch effect correction. Analysis of data variability and assessment of the batch effect (effect of the subsample/sequencing platform rather than biological variability) were performed using principal component analysis (PCA) with the plotPCA function of the DESeq2 v.1.42.0 package for R. Based on the results of the variability analysis, a conclusion was made about the inclusion/exclusion of samples in further analysis.

Differential gene expression analysis was performed using the DESeq2 package in R. Genes were filtered by $\log_2FC > 2$, $\log_2FC < (-2)$, and adjusted *p*-value < 0.05 (as visualized in the R package EnchancedVolcano). Genes encoding mRNAs and lncRNAs were categorized using Ensembl Ids. To identify differentially expressed genes (DEGs) between three biological states (neoplasia *vs.* control; cancer *vs.* neoplasia; cancer *vs.* control), comparisons were performed and the top 10 genes with statistically significant increased and decreased expression were identified. Heatmap for common DEGs was plotted with the pheatmap package in R.

Gene set enrichment analysis (GSEA) to estimate activated and repressed biological pathways in the comparison groups was performed using the clusterProfiler v.4.10.0 package for R.

Table 1. Main characteristics of the used studies

No.	Study ID	Samples count			Sample type	Reference
		Control	Neoplasia	Cervical cancer		
1	SRP048735	6	12	–	Cervical biopsy (FFPE)	Royse et al., 2014
2	SRA189004	–	–	7	Cervical smear; cell lines	Hu et al., 2015
3	GSE149763	3	3	3	Cervical biopsy (FF)	Qi et al., 2022

lncRNAs subcellular localization analysis. The web services Locate-R (Ahmad et al., 2020) and iLoc-LncRNA 2.0 (Su et al., 2018) were used to determine the subcellular localization of lncRNAs.

Validation on an independent dataset. An independent microarray dataset from the GEO database (GSE63514; 24 normal samples, 40 CIN3 samples, and 28 cancer samples) was used to validate the obtained results. Gene expression was obtained using GEO2R GUI available on the sample panel in GEO (all comparison groups were likened to those performed on the original dataset). Differential expression analysis was performed within GEO2R.

Results

The variability analysis of RNA sequencing data and the PCA assessment of the batch effect presented in Figure 1 showed that study No. 1 was significantly different from studies No. 2 and No. 3 (Fig. 1*a*), which may indicate the presence of a batch effect. After it was excluded from the analysis, the variability of samples from different studies decreased significantly (Fig. 1*b*). In further analysis, only data from studies 2 and 3 were used. Samples from these two datasets were combined for analysis into a single dataset without normalization due to the overall low batch effect.

The results of the differential gene expression analysis are shown in Figure 2. 1,572 DEGs were recorded in the "cancer *vs.* control" comparison, also 1,260 DEGs were recorded in the "cancer *vs.* neoplasia" comparison. It is important to note, that only 2 genes were observed to be differentially expressed in the "neoplasia *vs.* control" comparison.

The genes with the largest difference in the expression level for all comparisons are shown in Table 2. The top 10 genes
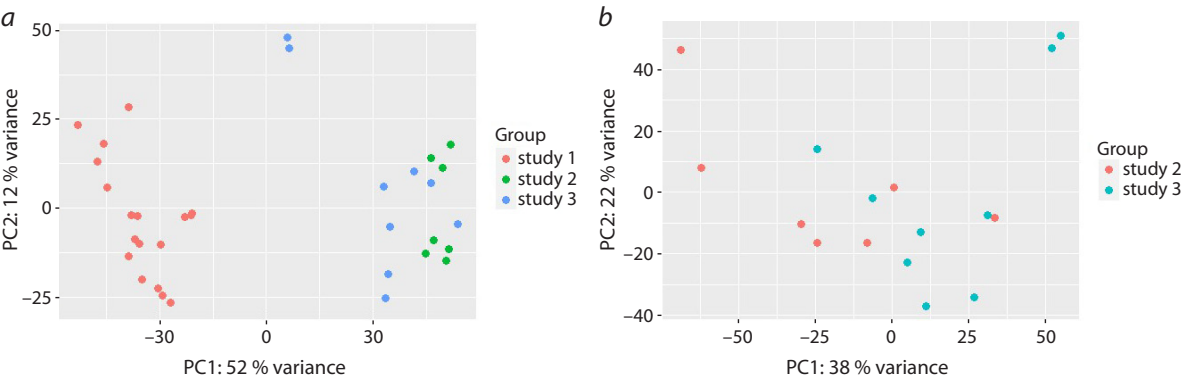

Fig. 1. RNA sequencing data variability analysis using PCA before the exclusion of study No. 1 (*a*) and after its exclusion (*b*).
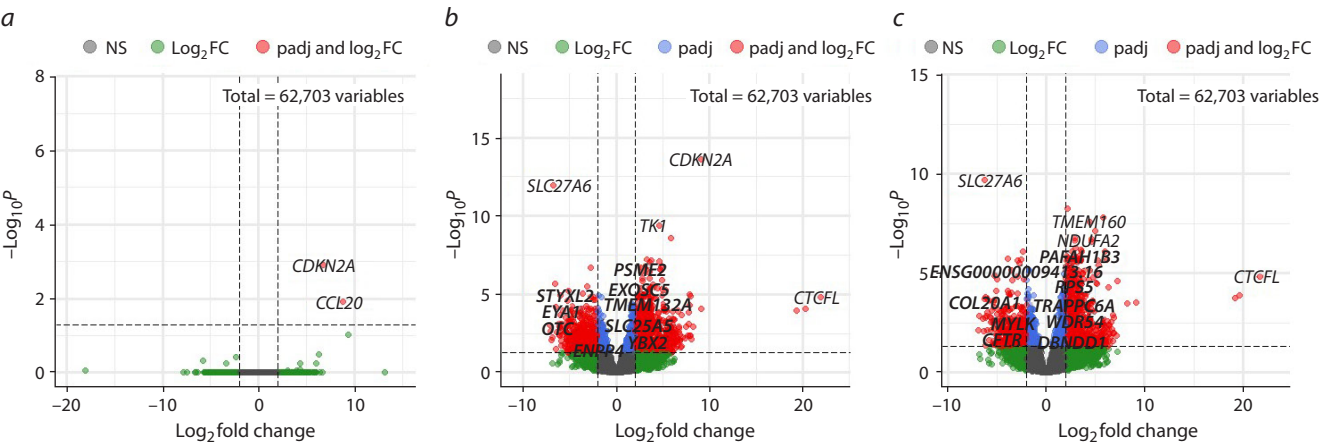

Fig. 2. Volcano plots for differentially expressed genes in the "neoplasia *vs.* control" (*a*), "cancer *vs.* control" (*b*), and "cancer *vs.* neoplasia" (*c*) comparison groups.

Е.Д. Кулаева, Е.С. Музлаева
Е.В. Машкина

Профиль экспрессии мРНК-днРНК при неоплазии
и цервикальном раке, ассоциированными с ВПЧ-инфекцией

2024
28 • 3

Table 2. Top differentially expressed genes in the "neoplasia *vs.* control", "cancer *vs.* control", and "cancer *vs.* neoplasia" comparisons

| Comparison	ENSEMBL ID	Gene name	Transcript type	log$_2$FC
Neoplasia *vs.* control, increased expression	ENSG00000115009.13	CCL20	Protein-coding	8.78
	ENSG00000147889.18	CDKN2A	Protein-coding	6.75
Cancer *vs.* control, increased expression	ENSG00000124092.13	CTCFL	Protein-coding	21.91
	ENSG00000290242.1	piR-55219	piRNA	20.25
	ENSG00000176165.13	FOXG1	Protein-coding	19.33
	ENSG00000147889.18	CDKN2A	Protein-coding	9.16
	ENSG00000019186.10	CYP24A1	Protein-coding	9.15
	ENSG00000149968.12	MMP3	Protein-coding	8.29
	ENSG00000119547.6	ONECUT2	Protein-coding	8.04
	ENSG00000196611.6	MMP1	Protein-coding	7.99
	ENSG00000118156.13	ZNF541	Protein-coding	7.98
	ENSG00000163064.7	EN1	Protein-coding	7.92
Cancer *vs.* control, decreased expression	ENSG00000253105.6	AP003548.1	lncRNA	−7.15
	ENSG00000124205.18	EDN3	Protein-coding	−7.06
	ENSG00000170893.4	TRH	Protein-coding	−6.91
	ENSG00000259458.1	MGC15885	lncRNA	−6.83
	ENSG00000113396.13	SLC27A6	Protein-coding	−6.72
	ENSG00000285336.1	LOC101928882	lncRNA	−6.58
	ENSG00000145242.14	EPHA5	Protein-coding	−6.50
	ENSG00000185069.2	KRT76	Protein-coding	−6.49
	ENSG00000279030.1	AC007336.3	Uncategorized transcript	−6.39
	ENSG00000280650.1	KCNIP4-IT1	lncRNA	−6.10
Cancer *vs.* neoplasia, increased expression	ENSG00000124092.13	CTCFL	Protein-coding	21.74
	ENSG00000282122.1	IGHV7-4-1	Protein-coding	19.65
	ENSG00000290242.1	piR-55219	piRNA	19.22
	ENSG00000127129.10	EDN2	Protein-coding	9.09
	ENSG00000213058.3	RPS14	Pseudogene	8.26
	ENSG00000107159.14	CA9	Protein-coding	7.22
	ENSG00000241749.4	RPSAP52	Pseudogene	7.12
	ENSG00000133328.4	PLAAT2	Protein-coding	6.85
	ENSG00000287929.1	lnc-LAMC1-1	lncRNA	6.84
	ENSG00000181617.6	FDCSP	Protein-coding	6.79
Cancer *vs.* neoplasia, decreased expression	ENSG00000170893.4	TRH	Protein-coding	−6.91
	ENSG00000289337.1	piR-52324-054	piRNA	−6.79
	ENSG00000145808.10	ADAMTS19	Protein-coding	−6.47
	ENSG00000248698.6	LINC01085	lncRNA	−6.34
	ENSG00000113396.13	SLC27A6	Protein-coding	−6.24
	ENSG00000178115.12	GOLGA8Q	Protein-coding	−6.21
	ENSG00000279622.2	lnc-ZDHHC7-3	lncRNA	−6.16
	ENSG00000145242.14	EPHA5	Protein-coding	−6.13
	ENSG00000143536.7	CRNN	Protein-coding	−6.13
	ENSG00000249421.2	ADAMTS19-AS1	lncRNA	−6.09
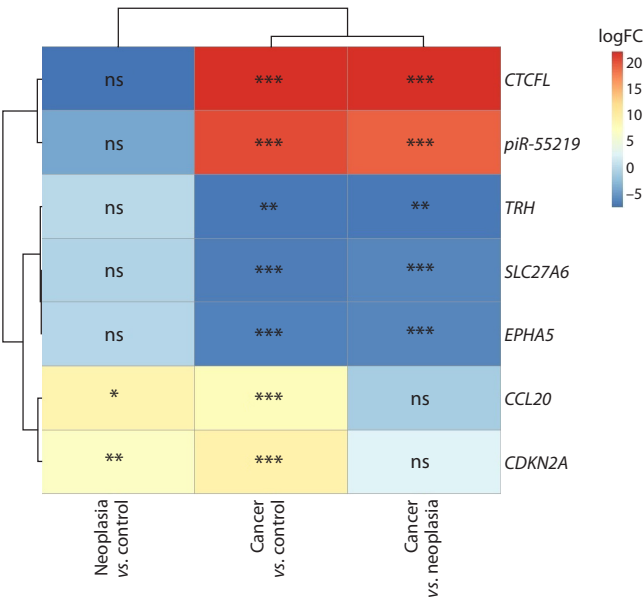
Fig. 3. Heatmap of the expression change patterns by logFC of selected genes.

The stars are intended to flag levels of significance for *p*-adjusted >0.05 (ns), <0.05 (*), <0.01 (**), <0.001 (***).

vs. neoplasia" comparison, 1 belongs to the lncRNA class, 1 belongs to the piwi-interacting RNA (piRNA) class, 2 belong to the pseudogene class, 6 belong to the protein-coding gene class; out of the genes with decreased expression, 3 genes belong to the lncRNA class, 1 gene belongs to the piRNA class, and the remaining 6 belong to the protein-coding gene class. In turn, out of the top 10 genes with increased expression for the "cancer *vs.* control" comparison, 1 belongs to the piRNA class, and 9 belong to the protein-coding gene class; out of the genes with decreased expression, 4 belong to the lncRNA class, 1 belongs to the uncategorized transcript class, and the remaining 5 belong to the protein-coding gene class.

Search for common genes among the most strongly differentially expressed genes among all comparison groups resulted in the identification of an expression signature consisting of the genes *CCL20*, *CDKN2A*, *CTCFL*, *piR-55219*, *TRH*, *SLC27A6* and *EPHA5*. The expression patterns of these genes are shown in Figure 3.

The gene enrichment analysis shown in Figure 4 demonstrated that in the "cancer *vs.* control" comparison, the molecular pathways associated with the cell cycle, DNA packaging, replication and translational mRNA base-pairing repression were activated, and the pathways associated with the membrane structure and cell-cell adhesion were repressed. Conversely, in the "cancer *vs.* neoplasia" comparison, molecular pathways related to the immunoglobulin production, antigen binding, respiratory chain and respirasome were activated, while pathways related to the translational mRNA base-pairing repression, post-transcriptional silencing, RISC and RNAi effector complexes were repressed.
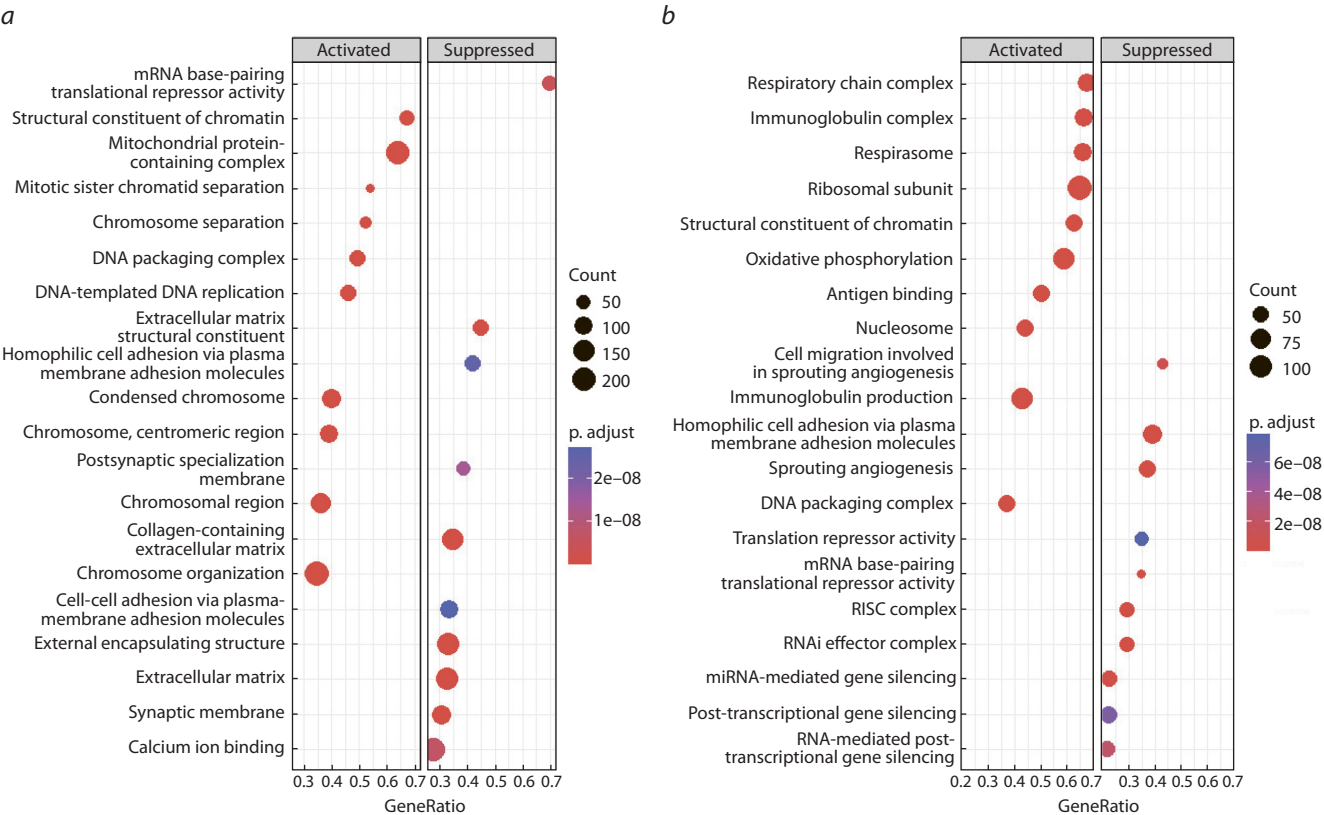
with increased and decreased expression for the "cancer *vs.* neoplasia" and "cancer *vs.* control" comparisons and 2 DEGs for the "neoplasia *vs.* control" comparison were presented. Out of the top 10 genes with increased expression for the "cancer

a

b

Fig. 4. Results of gene set enrichment analysis for the "cancer *vs.* control" (*a*) and "cancer *vs.* neoplasia" (*b*) comparison groups.

Е.Д. Кулаева, Е.С. Музлаева
Е.В. Машкина

Профиль экспрессии мРНК-днРНК при неоплазии
и цервикальном раке, ассоциированными с ВПЧ-инфекцией

2024
28 · 3

Table 3. Results of differentially expressed lncRNAs subcellular localization analysis using Locate-R and iLoc-LncRNA 2.0

lncRNA	Locate-R		iLoc-LncRNA 2.0	
	Localization	Score	Localization	Score
AP003548.1	Cytoplasm	1	Cytoplasm	0.84
MGC15885		1		0.73
LOC101928882		0.99		0.85
KCNIP4-IT1		1		0.83
lnc-LAMC1-1		0.94		0.86
LINC01085		1		0.69
lnc-ZDHHC7-3		0.92		0.87
ADAMTS19-AS1	Nucleus	0.97	Exosome	0.66

Table 4. Comparative analysis of the DEGs in datasets for study and validation

Comparison	Studied dataset (SRP048735+ SRA189004 + GSE149763, RNA-seq)	Validation dataset (GSE63514, microarray)	Overlap total	Overlap within the expression signature
Neoplasia *vs.* control	2	450	1	*CDKN2A*
Cancer *vs.* control	1,572	961	171	*CDKN2A, SLC27A6*
Cancer *vs.* neoplasia	1,260	215	15	–

Analysis of the subcellular localization of differentially expressed lncRNAs using two different web resources (Table 3) showed that subcellular localization is identified ambiguously for ADAMTS19-AS1 (nucleus and exosome), which may be related to differences in the computational approaches by which Locate-R (Local Deep SVM approach) and iLoc-LncRNA 2.0 (SVM approach) are implemented, despite the fact that both models are based on an analysis of the RNALocate lncRNA localization database. The results clearly indicated cytoplasmic localization for most of the transcripts (Table 3).

Validation on an independent dataset
Differential gene expression analysis on an independent dataset (Table 4) demonstrated that the number of DEGs observed in the "neoplasia *vs.* control" comparison was much higher than in the same comparison in our study, whereas for the other two comparisons, the number of DEGs was lower in the independent dataset than in our study. In particular, the *CDKN2A* and *SLC27A6* genes confirmed their expression change in the same pattern after validation.

Discussion
The division of HPV-infected cervical epithelial cells leads to neoplastic tissue changes or cervical intraepithelial neoplasia (CIN). The changes detected at the levels of cell morphology and tissue structure are the consequence of alterations at the molecular level. Neoplastic changes of epithelial cells in HPV infection are characterized by an increased level of transcription of the *CCL20* and *CDKN2A* genes, which lasts through the progression of the malignant process.

Tumor development assumes long-term persistence of HPV and the formation of a high viral load. Moreover, the virus can use the replicative apparatus of human cells and avoid the action of immune system factors. The main mechanisms of evasion from the immune system include modulation of antigen presentation, inhibition of cytokines and chemoattractants, modulation of cell adhesion molecule synthesis and inhibition of antigen-presenting cell migration.

The *CCL20* gene, differentially expressed in neoplastic changes of infected cells, may be a direct participant in these processes. *CCL20* belongs to the subfamily of small cytokine CC genes, it is located on chromosome 2q and contains 4 exons and 3 introns. This gene encodes macrophage inflammatory protein (MIP)-3α, predominantly expressed in liver, colon, prostate, cervix, and skin. It has been reported that endothelial cells, neutrophils, T helper 17 (Th17) cells, B cells, natural killer cells, dendritic cells (DC) and macrophages secrete CCL20 (Yamazaki et al., 2008; Nandi et al., 2014). CCL20 as a chemoattractant is involved in recruiting lymphocytes and dendritic cells to epithelial cells. It is believed that CCL20 may play an important role in the regulation of Langerhans cells, which are the main antigen-presenting cells for HPV presentation, causing an immune response.

From this perspective, it is reasonable to assume that active expression of CCL20 would be triggered in response to the appearance of human papillomavirus in the body. However, many studies indicate that HPV oncoproteins E6 and E7 may reduce the production of the chemokine CCL20 in keratinocytes by inhibiting its transcription. And thus, HPV, in an attempt to avoid an immune response, may negatively modulate the expression of this chemokine in the epithelium, thereby blocking the migration of inflammatory cells, such as Langerhans cells, to the lesion site (Guess, McCance, 2005; Wang et al., 2010; Jiang, Xue, 2015; Fernandes et al., 2021).

However, in the later stages of cervical carcinogenesis, the landscape changes and CIN3 lesions often contain myeloid cells such as macrophages and dendritic cells (Mazibrada et

al., 2008) and, as a number of studies have shown, CCL20 levels in cervical cancer tissues are significantly higher than in non-tumor and normal control tissues (Yu et al., 2015). Cervical cancer cells have been found to instruct cervical fibroblasts to produce CCL20 (Walch-Ruckheim et al., 2015). The rationale is that although normal immune cells attack and suppress tumor cells, some immune cells that infiltrate cancer tissue lose their anti-tumor function and play a role in promoting tumor progression (Beatty, Gladney, 2015; Binnewies et al., 2018).

Alteration of the cell cycle of infected epitheliocytes is possible due to changes in the transcription level of the *CDKN2A* gene. *CDKN2A* is a cyclin-dependent kinase 2a inhibitor gene that, through the use of alternative reading frames, produces two major proteins: p16 (INK4), an inhibitor of cyclin-dependent kinase 2 which arrests the G1-S transition in the cell cycle, and p14 (ARF), which binds the p53-stabilizing protein MDM2 (Robertson, Jones, 1999). It is important to note that *CDKN2A* is overexpressed in various cancers, and often its expression level correlates with the number of mutations, microsatellite instability in the tumor genome, and immune infiltration in the tumor microenvironment (Chen Z. et al., 2021). However, a study of *CDKN2A* expression in cervical cancer cell lines performed by real-time PCR and western blotting showed that it was reduced; moreover, the authors concluded that CDKN2A inhibits cell proliferation and invasion in cervical cancer through the AKT-mTOR lactate dehydrogenase mediated pathway (Luan et al., 2021). Several bioinformatics studies analyzing RNA sequencing data of cervical cancer samples have found that *CDKN2A* is a kind of "nodal gene" of tumorigenesis through interactions with various transcription factors, signaling molecules and microRNAs (e. g. miR-424-5p and miR-9-5p) and is overexpressed in cervical carcinoma in the TCGA project (Zhao et al., 2018; Chen Z. et al., 2021). In our study, *CDKN2A* expression was upregulated in patients with both HPV-associated neoplasia and HPV-associated cervical cancer, which draws attention to the importance of a more thorough study of the expression pattern of this gene and the features of the above pathologic conditions.

At the same time, the pattern of *CDKN2A* methylation in cervical cancer is relatively well known; several meta-analyses have shown that *CDKN2A* hypermethylation (relative to control samples) can be an indicator of early disease progression (Li J. et al., 2016). *CDKN2A* methylation was found to gradually increase with disease progression from stage 1 neoplasia to cervical cancer (Wijetunga et al., 2016), which can also be used as a comparative marker of disease severity. We would like to emphasize the need for a study linking the expression and methylation status of *CDKN2A* in HPV-associated neoplasia and cervical cancer to expand the understanding of the functional role of *CDKN2A* regulation in these conditions.

The most significant reduction of expression level in cancer cells relative to both control and neoplasia was found for five transcripts: *SLC27A6*, *EPHA5*, *TRH*, *CTCFL*, and *piR-55219*.

The *SLC27A6* gene encodes a fatty acid transfer protein through the cell membrane. Long-chain fatty acids are essential for various physiological processes. The function of SLC27A6 in cervical cancer has not, to our knowledge, been clarified. However, it is reported that *SLC27A6* expression was decreased in esophageal squamous cell carcinoma and breast cancer cells as well as nasopharyngeal carcinoma cells compared to normal cells (Xu C.Q. et al., 2015; Yen et al., 2019). It was also observed that the methylation ratio of the *SLC27A6* promoter was higher in nasopharyngeal carcinoma than in nonmalignant tissues (Xu C.Q. et al., 2015). On the contrary, *SLC27A6* gene expression was increased in papillary thyroid carcinoma (Dai et al., 2020).

The function of the ephrin A5 receptor encoded by the *EPHA5* gene in cervical cancer is also unclear. However, suppression of *EPHA5* expression by methylation has been established for breast cancer (Fu et al., 2010), prostate cancer (Li S. et al., 2015), and colorectal cancer (Kober et al., 2011). The loss of *EPHA5* expression was associated with the degree of serous ovarian carcinoma – the expression of this gene in cancer was reduced by 45 % in relation to neoplasia (Chen X. et al., 2016).

The *TRH* gene encodes a member of the thyrotropin-releasing hormone family involved in the hypothalamus–pituitary–thyroid axis which exhibits feedback of thyroid hormone, thereby regulating metabolic and immunological homeostasis. *TRH* has been well investigated in the type of cancer such as acute myeloid leukemia, and a correlation between risk groups and *TRH* expression was found, and it was discovered that patients with higher *TRH* expression were more sensitive to chemotherapy (Gao et al., 2022). Regarding CIN and cervical cancer, site-specific assessment of *TRH* gene methylation (cg01009664) was investigated for the detection of CIN2+ and demonstrated high sensitivity and specificity with clinician-collected samples, but not with the self-collected ones (Chaiwongkot et al., 2023). A similar analysis was also performed using screening of *TRH* cg01009664 methylation for prediction of oral squamous cell carcinoma and oropharyngeal squamous cell carcinoma (Puttipanyalears et al., 2018).

Conversely, the *CTCFL* and *piR-55219* genes are significantly upregulated in their expression in both "cancer *vs.* neoplasia" and "cancer *vs.* control" comparisons.

The *CTCFL* gene, which is sometimes also called *BORIS*, is a paralog of the widely known CTCF transcription factor and is normally expressed in pre-meiotic male germ cells together with ubiquitously expressed *CTCF* being involved in the regulation of the testis-specific genes (Soltanian, Dehghani, 2018; Debaugny, Skok, 2020). Unlike *CTCF*, *CTCFL* is more frequently amplified or transcriptionally activated, rather than mutated in cancers, and in cervical cancer the aberrant expression of *CTCFL* is linked with the re-initiating promoter hypomethylation of this gene (Debaugny, Skok, 2020). Moreover, a study performed on the cervical cancer stem-like cells (CSCs)/cancer-initiating cells (CICs) claimed that *BORIS* sf6 (isoform from subfamily 6) is specifically expressed in cervical CSCs/CICs and has a role in the maintenance of CSCs/CICs and proposed a peptide isoform BORIS C34_24(9) as a promising candidate for cervical CSC/CIC-targeting immunotherapy (Asano et al., 2016). Clinically, in cases of epithelial ovarian cancer and cervical cancer, high levels of *BORIS* expression were associated with poorer prognosis/less median survival times of patients and advanced cancer stages (Soltanian, Dehghani, 2018).

Way less is known about piwi-interacting RNA *piR-55219*. In general, piwi-interacting RNAs (piRNAs), which are

Е.Д. Кулаева, Е.С. Музлаева
Е.В. Машкина

Профиль экспрессии мРНК-днРНК при неоплазии
и цервикальном раке, ассоциированными с ВПЧ-инфекцией

2024
28·3

25–31 nucleotides in length, have been found to cluster at transposon loci in the genome and are thought to be critical for silencing these mobile genetic elements, via DNA methylation, to maintain genomic integrity in germline stem cells. Although they have only recently been identified in cancers, it is possible that the piRNAs that mediate transposon silencing during normal germline differentiation are hijacked in cancer cells to silence other parts of the genome, resulting in a tumorigenic state (Siddiqi, Matushansky, 2012; Suzuki et al., 2012). Unfortunately, no specific information on the involvement of *piR-55219* in cancer processes has been shown, which emphasizes the need for a more detailed investigation to establish a functional relationship between piRNAs and other cancer-specific genes.

Conclusion

We identified a predominantly cytoplasmic localization for the majority of differentially expressed lncRNAs. These lncRNAs can be involved in post-transcriptional regulation through their influence on the stability of mRNAs, act as translation regulators while forming mRNA-lncRNA complexes and can release miRNAs from their target genes as miRNA "sponges" (Xu Y. et al., 2023). All these processes may be impaired in cervical cancer, so it is important to further investigate the molecular mechanisms of function of lncRNAs selected in this study.

The results of our study differ significantly between the discovery and validation cohorts, which may be related to sample preparation protocols (FF+FFPE *vs.* cryosectioning) and expression assessment method (RNA-seq *vs.* microarray), which once again confirms the need to generate large protocol-uniformed datasets for studying neoplasia and cervical cancer at the same time.

Therefore, the analysis of differential gene expression in HPV-infected neoplasia and cervical cancer revealed a pattern of 7 genes with altered transcription levels. The increased transcription level of the *CCL20* and *CDKN2A* genes occurs at the stage of neoplastic epithelial changes and persists in cervical cancer. The *CDKN2A* and *SLC27A6* genes confirmed their expression change in the same patterns after validation on the independent microarray dataset.

References

Ahmad A., Lin H., Shatabda S. Locate-R: subcellular localization of long non-coding RNAs using nucleotide compositions. *Genomics.* 2020;112(3):2583-2589. DOI 10.1016/j.ygeno.2020.02.011

Asano T., Hirohashi Y., Torigoe T., Mariya T., Horibe R., Kuroda T., Tabuchi Y., Saijo H., Yasuda K., Mizuuchi M., Takahashi A., Asanuma H., Hasegawa T., Saito T., Sato N. Brother of the regulator of the imprinted site (BORIS) variant subfamily 6 is involved in cervical cancer stemness and can be a target of immunotherapy. *Oncotarget.* 2016;7(10):11223-11237. DOI 10.18632/oncotarget.7165

Bao Y., Wang L., Shi L., Yun F., Liu X., Chen Y., Chen C., Ren Y., Jia Y. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. *Cell. Mol. Biol. Lett.* 2019;24:38. DOI 10.1186/s11658-019-0162-0

Beatty G.L., Gladney W.L. Immune escape mechanisms as a guide for cancer immunotherapy. *Clin. Cancer Res.* 2015;21(4):687-692. DOI 10.1158/1078-0432.CCR-14-1860

Binnewies M., Roberts E.W., Kersten K., Chan V., Fearon D., Merad M., Coussens L., Gabrilovich D., Ostrand-Rosenberg S., Hedrick C., Vonderheide R., Pittet M., Jain R., Zou W., Howcroft T., Woodhouse E., Weinberg R., Krummel M. Understanding the tumor

immune microenvironment (TIME) for effective therapy. *Nat. Med.* 2018;24(5):541-550. DOI 10.1038/s41591-018-0014-x

Chaiwongkot A., Buranapraditkun S., Oranratanaphan S., Chuen-Im T., Kitkumthorn N. Efficiency of CIN2+ detection by thyrotropin-releasing hormone (TRH) site-specific methylation. *Viruse.* 2023; 15(9):1802. DOI 10.3390/v15091802

Chen X., Wang X., Wei X., Wang J. EphA5 protein, a potential marker for distinguishing histological grade and prognosis in ovarian serous carcinoma. *J. Ovarian Res.* 2016;9(1):83. DOI 10.1186/s13048-016-0292-1

Chen Z., Guo Y., Zhao D., Zou Q., Yu F., Zhang L., Xu L. Comprehensive analysis revealed that *CDKN2A* is a biomarker for immune infiltrates in multiple cancers. *Front. Cell Dev. Biol.* 2021;9:808208. DOI 10.3389/fcell.2021.808208

Dai J., Yu X., Han Y., Chai L., Liao Y., Zhong P., Xie R., Sun X., Huang Q., Wang J., Yin Z., Zhang Y., Lv Z., Jia C. TMT-labeling proteomics of papillary thyroid carcinoma reveal invasive biomarkers. *J. Cancer.* 2020;11(20):6122-6132. DOI 10.7150/jca.47290

Debaugny R., Skok J. CTCF and CTCFL in cancer. *Curr. Opin. Genet. Dev.* 2020;61:44-52. DOI 10.1016/j.gde.2020.02.021

Eklund C., Lagheden C., Robertsson K.D., Forslund O., Dillner J. Technical Report on the Global HPV LabNet DNA Genotyping Proficiency Panel 2019. International Human Papillomavirus (HPV) Reference Center, 2020

Fernandes A.T., Carvalho M., Avvad-Portari E., Rocha N., Russomano F., Roma E.H., Bonecini-Almeida M. A prognostic value of CD45RA+, CD45RO+, CCL20+ and CCR6+ expressing cells as 'immunoscore' to predict cervical cancer induced by HPV. *Sci. Rep.* 2021;11(1):8782. DOI 10.1038/s41598-021-88248-x

Fu D.Y., Wang Z.M., Wang B.L., Chen L., Yang W.T., Shen Z.Z., Huang W., Shao Z.M. Frequent epigenetic inactivation of the receptor tyrosine kinase *EphA5* by promoter methylation in human breast cancer. *Hum. Pathol.* 2010;41(1):48-58. DOI 10.1016/j.humpath.2009.06.007

Gao Y., Zhou J., Mao J., Jiang L., Li X.-P. Identification of the Thyrotropin-Releasing Hormone (TRH) as a novel biomarker in the prognosis for acute myeloid leukemia. *Biomolecules.* 2022;12(10):1359. DOI 10.3390/biom12101359

Gebrie A. Disease progression role as well as the diagnostic and prognostic value of microRNA-21 in patients with cervical cancer: a systematic review and meta-analysis. *PLoS One.* 2022;17(7):e0268480. DOI 10.1371/journal.pone.0268480

Guess J.C., McCance D.J. Decreased migration of Langerhans precursor-like cells in response to human keratinocytes expressing human Papillomavirus type 16 E6/E7 is related to reduced macrophage inflammatory protein-3α production. *J. Virol.* 2005;79(23):14852-14862. DOI 10.1128/JVI.79.23.14852-14862.2005

Hu Z., Zhu D., Wang W., Li W., Jia W., Zeng X., Ding W., Yu L., Wang X., Wang L., Shen H., Zhang C., Liu H., Liu X., Zhao Y., Fang X., Li S., Chen W., Tang T., Fu A., Wang Z., Chen G., Gao Q., Li S., Xi L., Wang C., Liao S., Ma X., Wu P., Li K., Wang S., Zhou J., Wang J., Xu X., Wang H., Ma D. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat. Genet.* 2015;47(2):158-163. DOI 10.1038/ng.3178

Jiang B., Xue M. Correlation of E6 and E7 levels in high-risk HPV16 type cervical lesions with CCL20 and Langerhans cells. *Genet. Mol. Res.* 2015;14(3):10473-10481. DOI 10.4238/2015.September.8.8

Karimzadeh M., Arlidge C., Rostami A., Lupien M., Bratman S., Hoffman M. Human papillomavirus integration transforms chromatin to drive oncogenesis. *Genome Biol.* 2023;24(1):142. DOI 10.1186/s13059-023-02926-9

Kober P., Bujko M., Olędzki J., Tysarowski A., Siedlecki J.A. Methyl-CpG binding column-based identification of nine genes hypermethylated in colorectal cancer. *Mol. Carcinog.* 2011;50(11):846-856. DOI 10.1002/mc.20763

Li J., Zhou C., Zhou H., Bao T., Gao T., Jiang X., Ye M. The association between methylated *CDKN2A* and cervical carcinogenesis, and its

diagnostic value in cervical cancer: a meta-analysis. *Ther. Clin. Risk Manag*. 2016;12:1249-1260. DOI 10.2147/TCRM.S108094

Li S., Zhu Y., Ma C., Qiu Z., Zhang X., Kang Z., Wu Z., Wang H., Xu X., Zhang H., Ren G., Tang J., Li X., Guan M. Downregulation of EphA5 by promoter methylation in human prostate cancer. *BMC Cancer*. 2015;15:18. DOI 10.1186/s12885-015-1025-3

Luan Y., Zhang W., Xie J., Mao J. *CDKN2A* inhibits cell proliferation and invasion in cervical cancer through LDHA-mediated AKT/mTOR pathway. *Clin. Transl. Oncol*. 2021;23(2):222-228. DOI 10.1007/s12094-020-02409-4

Martin J.A., Wang Z. Next-generation transcriptome assembly. *Nat. Rev. Genet*. 2011;12(10):671-682. DOI 10.1038/nrg3068

Mazibrada J., Rittà M., Mondini M., De Andrea M., Azzimonti B., Borgogna C., Ciotti M., Orlando A., Surico N., Chiusa L., Landolfo S., Gariglio M. Interaction between inflammation and angiogenesis during different stages of cervical carcinogenesis. *Gynecol. Oncol*. 2008;108(1):112-120. DOI 10.1016/j.ygyno.2007.08.095

Nandi B., Pai C., Huang Q., Prabhala R., Munshi N., Gold J. CCR6, the sole receptor for the chemokine CCL20, promotes spontaneous intestinal tumorigenesis. *PLoS One*. 2014;9(5):e97566. DOI 10.1371/journal.pone.0097566

Okunade K.S. Human papillomavirus and cervical cancer. *J. Obstet. Gynaecol*. 2020;40(5):602-608. DOI 10.1080/01443615.2019.1634030

Puttipanyalears C., Arayataweegool A., Chalertpet K., Rattanachayoto P., Mahattanasakul P., Tangjaturonsasme N., Kerekhanjanarong V., Mutirangura A., Kitkumthorn N. TRH site-specific methylation in oral and oropharyngeal squamous cell carcinoma. *BMC Cancer*. 2018;18(1):786. DOI 10.1186/s12885-018-4706-x

Qi D., Li H., Wang S., Wang S., Zheng R., Liu N., Han B., Liu L. Construction of ceRNA network and key gene screening in cervical squamous intraepithelial lesions. *Medicine* (*Baltimore*). 2022; 101(48):e31928. DOI 10.1097/MD.0000000000031928

Robertson K.D., Jones P.A. Tissue-specific alternative splicing in the human INK4a/ARF cell cycle regulatory locus. *Oncogene*. 1999; 18(26):3810-3820. DOI 10.1038/sj.onc.1202737

Royse K., Zhi D., Conner M., Clodfelder-Miller B., Srinivasasainagendra V., Vaughan L., Skibola C., Crossman D., Levy S., Shrestha S. Differential gene expression landscape of co-existing cervical pre-cancer lesions using RNA-seq. *Front. Oncol*. 2014;4:339. DOI 10.3389/fonc.2014.00339

Schubert M., Bauerschlag D., Muallem M., Maass N., Alkatout I. Challenges in the diagnosis and individualized treatment of cervical cancer. *Medicina* (*Kaunas*). 2023;59(5):925. DOI 10.3390/medicina59050925

Siddiqi S., Matushansky I. Piwis and piwi-interacting RNAs in the epigenetics of cancer. *J. Cell. Biochem*. 2012;113(2):373-380. DOI 10.1002/jcb.23363

Soltanian S., Dehghani H. BORIS: a key regulator of cancer stemness. *Cancer Cell Int*. 2018;18:154. DOI 10.1186/s12935-018-0650-8

Su Z.D., Huang Y., Zhang Z.Y., Zhao Y.W., Wang D., Chen W., Chou K.C., Lin H. iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics*. 2018;34(24):4196-4204. DOI 10.1093/bioinformatics/bty508

Sung H., Ferlay J., Siegel R.L., Laversanne M., Soerjomataram I., Jemal A., Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin*. 2021;71(3):209-249. DOI 10.3322/caac.21660

Suzuki R., Honda S., Kirino Y. PIWI expression and function in cancer. *Front. Gene*. 2012;3:204. DOI 10.3389/fgene.2012.00204

Walch-Ruckheim B., Mavrova R., Henning M., Vicinus B., Kim Y.J., Bohle R., Juhasz-Boss I., Solomayer E.F., Smola S. Stromal fibroblasts induce CCL20 through IL6/C/EBPβ to support the recruitment of Th17 cells during cervical cancer progression. *Cancer Res*. 2015;75(24):5248-5259. DOI 10.1158/0008-5472.CAN-15-0732

Wang X., Gao X.H., Hong Y., Li X., Chen H.D. Local hyperthermia decreases the expression of CCL-20 in condyloma acuminatum. *Virol. J*. 2010;7:301. DOI 10.1186/1743-422X-7-301

Wijetunga N.A., Belbin T., Burk R., Whitney K., Abadi M., Greally J., Einstein M., Schlecht N. Novel epigenetic changes in *CDKN2A* are associated with progression of cervical intraepithelial neoplasia. *Gynecol. Oncol*. 2016;142(3):566-573. DOI 10.1016/j.ygyno.2016.07.006

Xu C.Q., Zhu S.T., Wang M., Guo S.L., Sun X.J., Cheng R., Xing J., Wang W.H., Shao L.L., Zhang S.T. Pathway analysis of differentially expressed genes in human esophageal squamous cell carcinoma. *Eur. Rev. Med. Pharmacol. Sci*. 2015;19(9):1652-1661

Xu Y., Sun Y., Song X., Ren J. The mechanisms and diagnostic potential of lncRNAs, miRNAs, and their related signaling pathways in cervical cancer. *Front. Cell Dev. Biol*. 2023;11:1170059. DOI 10.3389/fcell.2023.1170059

Yamazaki T., Yang X., Chung Y., Fukunaga A., Nurieva R., Pappu B., Martin-Orozco N., Kang H.S., Ma L., Panopoulos A., Craig S., Watowich S., Jetten A., Tian Q., Dong C. CCR6 regulates the migration of inflammatory and regulatory T cells. *J. Immunol*. 2008;181(12): 8391-8401. DOI 10.4049/jimmunol.181.12.8391

Yan X., Hu Z., Feng Y., Hu X., Yuan J., Zhao S., Zhang Y., Yang L., Shan W., He Q., Fan L., Kandalaft L., Tanyi J., Li C., Yuan C.X., Zhang D., Yuan H., Hua K., Lu Y., Katsaros D., Huang O., Montone K., Fan Y., Coukos G., Boyd J., Sood A., Rebbeck T., Mills G., Dang C., Zhang L. Comprehensive genomic characterization of long non-coding RNAs across human cancers. *Cancer Cell*. 2015;28(4): 529-540. DOI 10.1016/j.ccell.2015.09.006

Yen M.C., Chou S.K., Kan J.Y., Kuo P.L., Hou M.F., Hsu Y.L. New insight on solute carrier family 27 member 6 (SLC27A6) in tumoral and non-tumoral breast cells. *Int. J. Med. Sci*. 2019;16(3):366-375. DOI 10.7150/ijms.29946

Yu Q., Lou X.M., He Y. Preferential recruitment of Th17 cells to cervical cancer via CCR6-CCL20 pathway. *PLoS One*. 2015;10(3): e0120855. DOI 10.1371/journal.pone.0120855

Zhao L., Zhang Z., Lou H., Liang J., Yan X., Li W., Xu Y., Ou R. Exploration of the molecular mechanisms of cervical cancer based on mRNA expression profiles and predicted microRNA interactions. *Oncol. Lett*. 2018;15(6):8965-8972. DOI 10.3892/ol.2018.8494

Allele-specific PCR with fluorescently labeled probes: criteria for selecting primers for genotyping

V.A. Devyatkin, A.A. Shklyar, A.Zh. Fursova (iD), Yu.V. Rumyantseva, O.S. Kozhevnikova (iD) (✉)

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

Abstract. Single-nucleotide polymorphisms (SNPs) can serve as reliable markers in genetic engineering, selection, screening examinations, and other fields of science, medicine, and manufacturing. Whole-genome sequencing and genotyping by sequencing can detect SNPs with high specificity and identify novel variants. Nonetheless, in situations where the interest of researchers is individual specific loci, these methods become redundant, and their cost, the proportion of false positive and false negative results, and labor costs for sample preparation and analysis do not justify their use. Accordingly, accurate and rapid methods for genotyping individual alleles are still in demand, especially for verification of candidate polymorphisms in analyses of association with a given phenotype. One of these techniques is genotyping using TaqMan allele-specific probes (TaqMan dual labeled probes). The method consists of real-time PCR with a pair of primers and two oligonucleotide probes that are complementary to a sequence near a given locus in such a way that one probe is complementary to the wild-type allele, and the other to a mutant one. Advantages of this approach are its specificity, sensitivity, low cost, and quick results. It makes it possible to distinguish alleles in a genome with high accuracy without additional manipulations with DNA samples or PCR products; hence the popularity of this method in genetic association studies in molecular genetics and medicine. Due to advancements in technologies for the synthesis of oligonucleotides and improvements in techniques for designing primers and probes, we can expect expansion of the possibilities of this approach in terms of the diagnosis of hereditary diseases. In this article, we discuss in detail basic principles of the method, the processes that influence the result of genotyping, criteria for selecting optimal primers and probes, and the use of locked nucleic acid modifications in oligonucleotides as well as provide a protocol for the selection of primers and probes and for PCR by means of rs11121704 as an example. We hope that the presented protocol will allow research groups to independently design their own effective assays for testing for polymorphisms of interest.

Key words: genotyping; single-nucleotide polymorphisms; TaqMan probes; LNA modifications; allele-specific PCR.

Аллель-специфичная ПЦР с флуоресцентно-мечеными зондами: критерии подбора праймеров для генотипирования

В.А. Девяткин, А.А. Шкляр, А.Ж. Фурсова (iD), Ю.В. Румянцева, О.С. Кожевникова (iD) (✉)

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

Аннотация. Однонуклеотидные полиморфизмы (SNP) могут служить надежными маркерами в генной инженерии, селекции, скрининговых обследованиях и других областях науки, медицины и производства. Полногеномное секвенирование и генотипирование при помощи секвенирования могут высокоспецифично детектировать SNP и выявлять новые аллели. Однако в ситуациях, когда интерес исследователей направлен на отдельные конкретные локусы, эти методы становятся избыточными, а их цена, доля ложноположительных и ложноотрицательных результатов и трудозатраты на пробоподготовку и анализ не оправдывают их применения. Поэтому точные и быстрые методы генотипирования отдельных аллелей все еще остаются востребованными, особенно при проверке кандидатных полиморфизмов в анализах ассоциации с определенным фенотипом. Один из таких методов – генотипирование с использованием аллель-специфичных зондов TaqMan (TaqMan dual labeled probes). Метод заключается в реакции ПЦР в реальном времени с использованием пары праймеров и двух олигонуклеотидных зондов, комплементарных последовательности вблизи данного локуса таким образом, что один зонд комплементарен аллелю дикого типа, а другой – мутантному аллелю. Преимуще-

ства метода заключаются в его специфичности, чувствительности, невысокой стоимости и быстроте получения результатов. Он позволяет с высокой точностью различать аллели в геноме в одностадийной ПЦР без дополнительного этапа разделения продуктов реакции, что делает его востребованным в исследованиях генетических ассоциаций в молекулярной генетике и медицине. Благодаря развитию технологий синтеза олигонуклеотидов и совершенствованию методов подбора праймеров и зондов можно ожидать расширения возможностей применения этого подхода в диагностике наследственных заболеваний. В настоящей статье мы разобрали основные принципы метода, процессы, влияющие на результат генотипирования, критерии подбора оптимальных праймеров и зондов, использование LNA-модификаций в олигонуклеотидах, а также привели протокол подбора праймеров, зондов и ПЦР на примере SNP rs11121704. Мы надеемся, что представленный протокол позволит исследовательским группам самостоятельно подбирать собственные эффективные тест-системы для проверки интересующих полиморфизмов.

Ключевые слова: генотипирование; однонуклеотидные полиморфизмы; зонды TaqMan; LNA-модификации; аллель-специфичная ПЦР.

Introduction

Single-nucleotide polymorphisms (SNPs) are actively used as reliable markers in genetic engineering, selection, screening examinations, and other fields of science, medicine, and industrial production. It is clear that whole-genome sequencing and genotyping by sequencing can detect SNPs with high specificity and identify novel variants. On the other hand, in situations where the interest of researchers is focused on individual specific loci, these methods become redundant, and their cost, the proportion of false positive and false negative results, and the labor costs for sample preparation and analysis do not justify their use. Accordingly, accurate and rapid techniques for genotyping individual alleles are still in demand, especially for verification of candidate polymorphisms in analyses of association with a given phenotype (Kalendar et al., 2022).

Currently, methods based on allele-specific PCR allow to obtain the most accurate results at a low cost and do not require highly qualified personnel or expensive laboratory equipment. In this work, we analyze principles of work with one of these approaches: genotyping by means of allele-specific probes based on the TaqMan method (TaqMan dual labeled probes). It was first described 15 years ago (Hui et al., 2008) and remains one of the most popular for the detection of SNPs. The accuracy of the method ensures genotyping error of less than one case per 2,000 (Ranade et al., 2001). The correct choice of primers and probes is possible for most genome sequences and in more than 90 % of cases allows to obtain fairly accurate genotyping results for high-quality DNA without further optimization.

The method of allele-specific PCR with TaqMan probes can separate genotypes even with small amounts of an initial sample, does not require post-PCR processing, and correlates well with other methods (Broccanello et al., 2018). Nonetheless, kits that are commercially available and developed for a specific SNP are expensive, and recommendations for creating and optimizing one's own assays are described rather superficially in most of literary sources.

In this work, we provide a detailed analysis of this technique with a description of processes that can influence genotyping results as well as recommendations for designing your own assays.

Description of the method

The method consists of real-time PCR involving a pair of primers (forward and reverse, between which a polymorphic locus of interest is located) and two oligonucleotide probes complementary to a sequence near this locus in such a way that one probe is complementary to the wild-type allele, and the other to a mutant one. Each probe has a distinct fluorescent dye at the 5′ end and a fluorescence quencher and phosphate group at the 3′ end. The phosphate group prevents the probes from acting as primers in the PCR. Due to the proximity of the quencher and dye, an intact probe does not yield a signal because of Förster resonance energy transfer (FRET) and fluorescence quenching. At the elongation stage, a Taq DNA polymerase molecule that has reached the probe bound to the fully complementary template hydrolyzes it owing to 5′–3′ exonuclease activity, thereby uncoupling the quencher and dye, and the fluorescence is detected by an instrument.

Hybridization of the probe with the template is more effective in the case of complete complementarity; moreover, in the case of an unpaired base (mismatch), when a probe corresponding to one allele binds to the template corresponding to another allele, the polymerase preferentially displaces it entirely without separating the chromophores. Therefore, signal accumulation will occur much more efficiently in the case of complete complementarity between the probe and the template. Thus, the ratio of fluorescence levels of the different dyes depends on the ratio of the alleles (corresponding to the probes labeled with these dyes) in the initial template (Hui et al., 2008).

In a situation close to ideal, an allelic discrimination plot, where X- and Y-axes correspond to the fluorescence levels of the first and second dye for each sample, looks like the one in Figure 1. Samples having the same genotype form a cloud of dots distant from other clusters. The fluorescence level of each dye is zero for homozygotes that do not have the allele labeled with this dye and is almost twice as high for homozygotes of a given allele as compared to heterozygotes.

In Figure 2, an outcome of a less specific reaction is shown, when annealing and subsequent restriction of probes additionally take place on the template corresponding to the other allele. Under this scenario, fluorescence intensity of both dyes is non-zero for all samples. On the other hand, the genotypes of the samples can still be well discriminated with high accuracy. It should be noted that during the reaction, the concentration of the specific probe decreases, the concentration of the template increases, and the probe with the mismatch remains intact, which shifts the equilibrium toward the formation of a duplex containing the mismatch. In this case, using a smaller amount of the initial template in the reaction or reading an allelogram at earlier cycles can help.
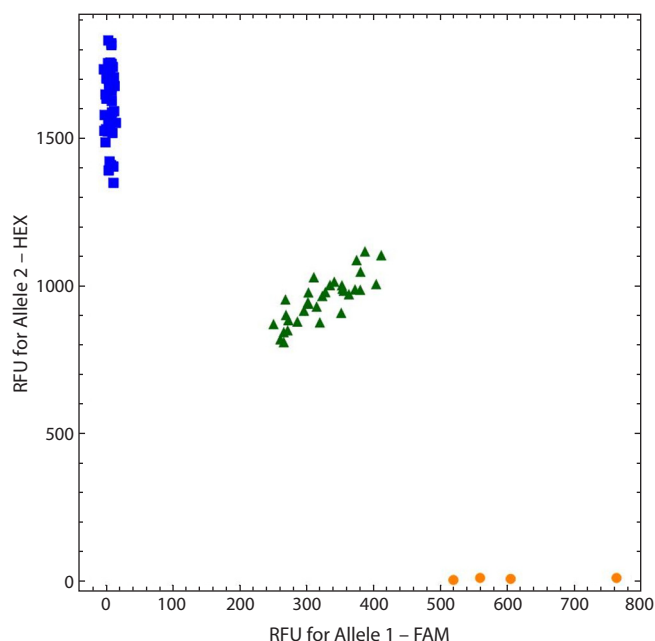
В.А. Девяткин, А.А. Шкляр, А.Ж. Фурсова
Ю.В. Румянцева, О.С. Кожевникова

Аллель-специфичная ПЦР с флуоресцентно-мечеными
зондами: критерии подбора праймеров для генотипирования

2024
28•3

Fig. 1. A high-quality allelic discrimination plot.

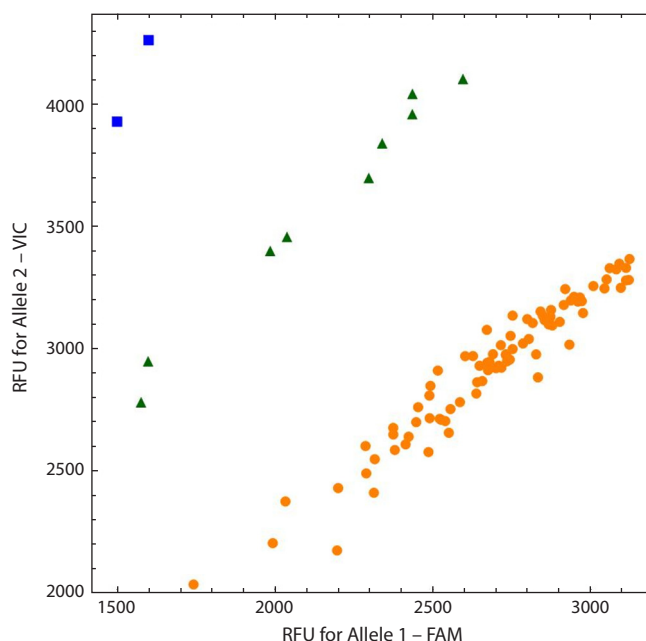Here and in Fig. 2: Created in Bio-Rad CFX Manager software. RFU: relative fluorescence units.

Fig. 2. The allelic discrimination plot with a low ratio of target fluorescence to the background signal, but reliable discrimination of the genotypes is still possible.

Based on the ratio between the dye signals in a single sample, it is impossible to directly determine which probe binds better and, accordingly, what the genotype is, because the dyes have different fluorescence intensities, probes differ in binding efficiency, and this process can be influenced by other random factors. Therefore, for the assay, it is necessary to employ several samples, among which there are different genotypes. It is recommended to analyze at least 20 samples per instrument run for reliable discrimination. By means of how much one signal increases more intensely than the other (i.e., by means of the angle between the *X*- and *Y*-axes in the plot for each sample), you can separate all samples into groups, and if three groups are detected, you can be confident in accurate determination of the genotype of each group. When assessing the results in the absence of a certain genotype among the samples, one should rely on the match between the actual frequency of genotypes and the one expected according to the literature or the Hardy–Weinberg equilibrium. To ensure correct allelic discrimination, it is advisable to verify samples of each genotype by Sanger sequencing.

Criteria for designing optimal primers and probes for genotyping

Design of primers
– GC content within 30–80 % (ideally 40–60 %).
– Stretches of a repeated nucleotide, especially four or more consecutive Gs, should be avoided.
– Melting point (T_m) in the range of 58–60 °C. The difference in T_m between the forward and reverse primers is no more than 2 °C.
– Among the five nucleotides at the 3′ end, more than two Gs and/or Cs are not recommended. T should be avoided at

the 3′ end. G or C at the last position at the 3′ end is a more suitable binding site for DNA polymerases.
– Length 18–30 nucleotides.
– The primers and probes must not overlap.
– The recommended amplicon length according to the literature is 80–120 nucleotides.

Increasing the PCR product length reduces reaction efficiency and nuclease activity of Taq polymerase (Debode et al., 2017). In some cases, depending on nucleotide composition, to meet other criteria, amplicon length can be increased to 1,000. The minimum length is determined by the total length of the primers and probes. It is best to keep the distance between a probe and the primer annealing to the same strand shorter to the extent feasible, no more than 20 nucleotides if possible. Our experience shows that amplicon length of 98 to 469 bp and a 22–348 bp distance from a probe to the primer complementary to the same strand have no visible effect on discrimination accuracy.

From the point of view of simplicity and low cost of the experiment, we recommend selecting primers by taking into account the possibility of Sanger sequencing from the same primers. In this case, it is desirable that there be at least 50 nucleotides from the sequencing primer to the polymorphic site and from the polymorphic site to the other end of the maximal read (i.e., the regions upstream and downstream of the polymorphic site).

Design of probes
– Both probes must anneal to the same strand and not be complementary to each other.
– T_m should be approximately 6–8 °C higher than that of the primers (as opposed to qPCR involving only one probe, for which the T_m should be 10 °C higher than that of the

primers). The reason is that as the reaction mixture cools, the oligonucleotide probes must anneal to the DNA template before the primers do.

– The greater the difference in T_m between the probe fully paired with the template and the probe carrying an unpaired base, the more efficiently can the alleles be separated. The minimum difference that has allowed us to separate alleles is 3 °C, but we advise designing probes to achieve a minimum of 4–5 °C whenever possible. The wider the temperature window, the easier it is to select an annealing temperature at which the annealing probability of a probe carrying a mismatch is negligible compared to that of a fully complementary probe.

– Do not place G at the 5′ end because it will quench the fluorophore attached to it after cleavage the probe. Furthermore, Taq polymerase cleaves such probes worse (Huang, Li, 2009). The cleavage of the oligonucleotide begins with the appearance of 1–2 unpaired nucleotides at the 5′ end, which are recognized by the nuclease domain. An unpaired G at the end dramatically disrupts complementarity, thereby sometimes causing complete strand separation faster than Taq polymerase can begin to cut the probe; thus, the probe is displaced entirely as if it was not fully complementary to begin with.

– Of the two strands, choose one such that the probes contain more Cs than Gs because empirical data indicate that such probes are more likely to produce a strong signal (https://www.thermofisher.com/order/catalog/product/450025).

– The polymorphic site should be located approximately in the middle third of the probe.

– GC content within 20–80 % (ideally 30–70 %).

– It is advisable to select the positions of the start and end of each probe so that T_m for both probes becomes approximately the same.

– The length of the probes is 18–30 nucleotides, and the optimal length is 20 nucleotides. These restrictions are due to the fact that a probe must bind specifically to only one region within the amplified fragment and satisfy T_m requirements. The longer the entire probe, the smaller is the contribution of the polymorphic site to the melting temperature, the smaller is the percentage difference in T_m for cases of complete complementarity and mismatch, and less effectively are the alleles discriminated. A length greater than 30 nucleotides is acceptable, but in such cases, the quencher should not be located at the 3′ end but inside the probe at a distance of approximately 18–25 nucleotides from the 5′ end. The reason is that at distances between the dye and quencher greater than 100 Å (corresponding to approximately 30 bp in B-form DNA structure), FRET is disrupted and intact probes can emit fluorescence, thus lowering the signal-to-background ratio.

– The reporter fluorophores (dyes) must have different emission spectra. Fluorophores from a list of those compatible with the instrument at hand should be chosen for different channels. Probes labeled with FAM and HEX are cleaved more efficiently than probes labeled with ROX or CY5.

– It is best to label with a brighter dye the probe having lower T_m or GC content or containing an A/T allele, which binds worse to the template (for example, FAM gives a stronger signal than HEX does).

Use of locked nucleic acid (LNA) modifications in oligonucleotides

Commercially available TaqMan probes can be conjugated to a minor groove binder (MGB) motif, e. g., dihydrocyclopyrroloindole tripeptide (DPI3), to increase the probe's binding affinity for the target sequence. This approach allows to increase melting temperature of the probe without increasing its length, thus improving discrimination between the complementary probe and noncomplementary probe.

Commercially available alternatives to this technology are locked nucleic acids (LNAs or bridged nucleic acids, BNAs), which are analogs of RNA with ribose locked in the 3′-endo conformation due to a 2′-O, 4′-C methylene bridge. The presence of these modified bases in the oligonucleotide enhances the thermal stability and specificity of hybridization (Owczarzy et al., 2011). Such nucleotides are usually marked as [+X] or +X (where X = A, T, G, or C). By replacing individual nucleotides in a probe with their LNA analogs, it is possible to make the probe itself shorter, and the contribution that the polymorphic site makes to the total T_m is greater, which will facilitate the discrimination of alleles. Typically, a modification of a single nucleotide at the SNP position is made, but for each sequence, effects of different variants may differ (You et al., 2006).

The choice of tools for calculating T_m of oligonucleotides with an LNA modification is narrower than in the case of unmodified bases; you can use OligoEvaluator services (http://www.oligoevaluator.com/LoginServlet) or the OligoAnalyzer Tool (https://www.idtdna.com/calc/analyzer).

Unfortunately, there are currently no available services for calculating the T_m difference between the fully paired duplex and the heteroduplex containing unpaired bases for oligonucleotides with an LNA modification. Previously, the OligoAnalyzer Tool has allowed for such calculations, but due to low accuracy, this option has been removed. To roughly estimate the effect of LNA on mismatches, you can use data from articles on the thermodynamics of oligonucleotides with LNA modifications. In some cases, an LNA modification even reduces the match vs. mismatch difference as compared to the unmodified oligonucleotide; therefore, these modifications require caution (You et al., 2006).

Design example

By changing the amplicon length, melting temperature, GC content, and positions and lengths of primers and probes, it is possible to obtain combinations that satisfy the above criteria and to select the best one. SNP-containing sequences being analyzed do not always permit designing primers and probes that meet all the aforementioned criteria, but this does not mean that the selected assay will not work in practice.

Let's consider the algorithm for designing primers and probes for analysis of the rs11121704 polymorphism.

1. Find the polymorphism in the dbSNP database (https://www.ncbi.nlm.nih.gov/snp/), and go to the page with a detailed description (https://www.ncbi.nlm.nih.gov/snp/rs11121704).

2. Find the substitution you need, and pay attention to the genome assembly for which the position is specified. For example, in our case, the latest one is currently GRCh38. At the next stage, we need information about the chromosome

В.А. Девяткин, А.А. Шкляр, А.Ж. Фурсова
Ю.В. Румянцева, О.С. Кожевникова

Аллель-специфичная ПЦР с флуоресцентно-мечеными
зондами: критерии подбора праймеров для генотипирования

2024
28 • 3

(NC_000001.11), the position on it (11233902), and the substitution (C>T). One rsID can correspond to several variants at one locus; all of them are listed on the page (C>A/C>T). Typically, most people have either the reference allele or the most common alternative allele. The necessary information can be found on the "Frequency" tab (Fig. 3).

In this genotyping method, we regard all polymorphisms as biallelic, and we are usually interested in the most common substitution at a given position because variants with a near-zero frequency can occur only in large study populations.

3. To select primers, we use the open online resource Primer-Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/)

In the "Enter accession" field, specify the chromosome with the polymorphism you are interested in (NC_000001.11).

In the "Range" fields, we indicate the boundaries within which the forward and reverse primers should lie near the SNP position (11233902). We set the boundaries of the primer-binding site no closer than 15 nucleotides to the SNP (because otherwise, primer landing may be impeded by the probe) and not farther than 200 (so that the amplicon is not too long and the reaction efficiency is higher): the forward primer from

11233702 (11233902-200) to 11233887 (11233902-15), and the reverse one from 11233917 (11233902+15) to 11234102 (11233902+200). "PCR product size" is set to 100–250.

In the "Database Refseq" field, choose "Refseq representative genomes" or "Genomes for selected organisms (primary reference assembly only)". Both databases contain primary assemblies of chromosomal sequences with minimal redundancy, and "representative genomes" also include alternative loci and mitochondrial genomes, if available.

The "Advanced parameters" option provides access to additional parameters, among which, we are interested in "Primer GC content (%)", for which we set the range to 40–60 %.

We leave the remaining parameters unchanged. Before clicking the "Get primers" button, select the "Show results in a new window" option so that after the results are displayed, it is easier to change individual launch parameters for a second search (Fig. 4).

4. Go to the page with the primer search results. We select primers that are suitable for the position, T_m, and specificity. You can work with nonspecific primers, but you must then check that the probe binds only to the specific amplicon and not to side products.

Fig. 3. Basic information about the rs11121704 variant: chromosome, position, genome assembly, and frequency of nucleotide substitutions.

Fig. 4. An example of settings for designing primers in the Primer-Blast tool.

We chose the sequence 5′-TTTTTCCTCATTTTGGGC
GA-3′ for the forward primer and 5′-TATCAGTTGCAG
GAAAGTGC-3′ for the reverse primer. The "results" page
shows that the selected primers give a target specific product
130 nucleotides long (Fig. 5). There is also one potential non-
specific PCR product with a length of 1,186 nucleotides, which
will not be synthesized due to incomplete complementarity of
the binding sites to the primer sequences (Fig. 5).

5. Next, in the "Tracks" option, select the "Configure
Tracks" suboption, find and check the boxes "Common va-
riations (MAF>=0.01)", "Cited Variations", and "ClinVar
variants with precise endpoints" and add them to the display
with the "Configure" button. A probe should not overlap with
polymorphisms other than the one of interest.

6. In the search results, select the sequence area around
the SNP (nucleotides –20...+20) and click "copy sequence
(selection)" (Fig. 6).

Additionally, build a complementary strand to this se-
quence. This can be done manually or use any available
service, for example, (https://www.bioinformatics.org/sms/
rev_comp.html).

Forward strand sequence:

5′-TTCTCCTTTCCAAACATCTG(C)GATGATGTGCC
TGAAGCATT-3′

Reverse strand sequence:

5′-AATGCTTCAGGCACATCATC(G)CAGATGTTTGG
AAAGGAGAA-3′

The position of the SNP in question is indicated in brackets.

В.А. Девяткин, А.А. Шкляр, А.Ж. Фурсова
Ю.В. Румянцева, О.С. Кожевникова

Аллель-специфичная ПЦР с флуоресцентно-мечеными
зондами: критерии подбора праймеров для генотипирования

2024
28 • 3

Primer pair 1

	Sequence (5'->3')	Template strand
Forward primer	TTTTTCCTCATTTTGGGCGA	Plus
Reverse primer	TATCAGTTGCAGGAAAGTGC	Minus
Product length	130	

Products on intended targets
>NC_000001.11 Homo sapiens chromosome 1, GRCh38.p14 Primary Assembly

```
product length = 130
Features associated with this product:
    serine/threonine-protein kinase mtor isoform 1

    serine/threonine-protein kinase mtor isoform x3

Forward primer  1          TTTTTCCTCATTTTGGGCGA  20
Template        11233836   ....................  11233855

Reverse primer  1          TATCAGTTGCAGGAAAGTGC  20
Template        11233965   ....................  11233946
```

Products on potentially unintended templates
>NC_000013.11 Homo sapiens chromosome 13, GRCh38.p14 Primary Assembly

```
product length = 1186
Features associated with this product:
    fibroblast growth factor 14 isoform 1a

    fibroblast growth factor 14 isoform 3

Forward primer  1           TTTTTCCTCATTTTGGGCGA  20
Template        101877194   ..A.G............T..  101877213

Reverse primer  1           TATCAGTTGCAGGAAAGTGC  20
Template        101878379   CTGA....T...........  101878360
```

Fig. 5. The selected primers in the Primer-Blast tool give a specific target PCR product 130 nucleotides long containing SNP rs11121704, and one potential PCR product 1,186 nucleotides long, which should not form under normal conditions.

7. Within the sequence near the SNP, select a fragment of suitable length and composition. Based on the GC content, it is worthwhile taking the reverse strand sequence, because in this case, there will be more Cs than Gs in the probe:

5′-CAGGCACATCATC(G)CAGATGTTT-3′

Given that we are not taking the strand in which the SNP is shown to be located, you should remember that for our sequence, the C>T substitution in the reverse strand corresponds to the G>A substitution.

We select the boundaries of the second probe so as to equalize the probes' T_m:

5′-CAGGCACATCATC(A)CAGATGTTTG-3′

8. It is recommended to check T_m by means of several services and to average it (see the Table).

For example, we used Oligo Calc (http://biotools.nubic. northwestern.edu/OligoCalc.html), OligoEvaluator (http://www.oligoevaluator.com/LoginServlet), and OligoAnalyzer Tool (https://www.idtdna.com/calc/analyzer).

9. To compare the melting temperatures between the fully complementarily bound probe and the probe forming an unpaired base, select the "Tm mismatch" option in the OligoAnalyzer Tool.

For the first probe CAGGCACATCATC(G)CAGATGTTT, the mismatch is the nucleotide complementary to the second probe (CAGGCACATCATC(A)CAGATGTTTG), i. e., select the letter "T", and click "Use Exact Complement Tm" and "Calculate". The greater the difference in the T_m between the
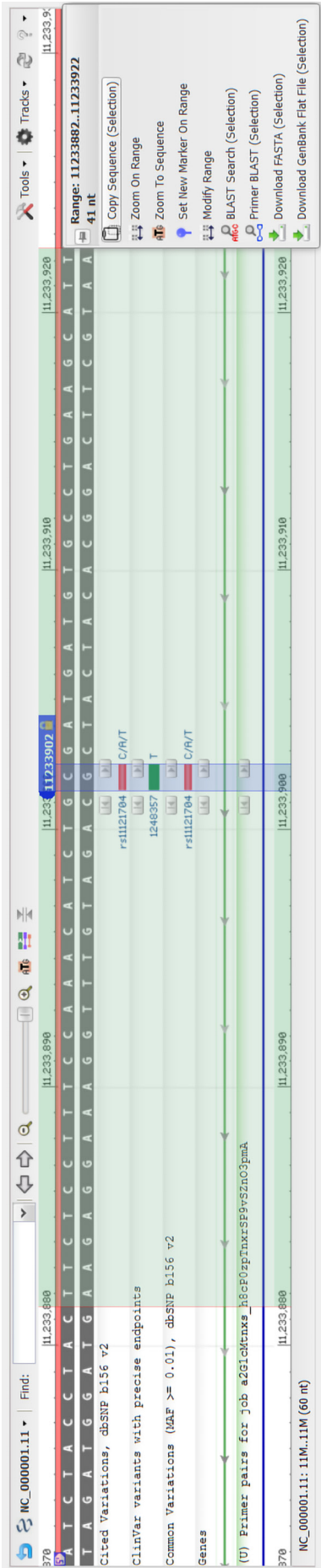

Fig. 6. A nucleotide sequence that flanks the SNP and does not overlap with other known polymorphic sites.

Main characteristics of primers and probes for rs11121704

Oligonucleotide	T_m OligoEvaluator	T_m OligoCalc	T_m Oligo-Analyzer Tool	T_m average	T_m mismatch	delta T_m
5'-TTTTTCCTCATTTTGGGCGA-3'	66.3	54.3	63.0	61.2		
5'-TATCAGTTGCAGGAAAGTGC-3'	60.5	56.4	62.4	59.8		
5'-FAM-CAGGCACATCATCGCAGATGTTT-BHQ1-3'	70.2	62.9	66.4	66.5	62.4	4.0
5'-VIC-CAGGCACATCATCACAGATGTTTG-BHQ1-3'	68.6	63.6	64.9	65.7	60.3	4.6

Note. Melting temperatures (T_m) predicted by several services and their averages are listed. The T_m mismatch is the melting temperature of the probe in the noncomplementary duplex with the template. "delta T_m" is the difference between "T_m (Oligo-Analyzer Tool)" and "T_m mismatch".

fully complementary oligonucleotide and the probe having the noncomplementary base ("deltaTm") (and accordingly, the lower the proportion of the bound mismatched probe compared to the fully complementary probe at the probe annealing stage), the more accurate the allele discrimination will be.

10. To check the specificity of the newly designed probes, we use the Blast service (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome).

In the "Enter accession number(s), gi(s), or FASTA sequence(s)" field, insert the probe sequence; "Database" should be "Refseq representative genomes", "Organism" should be "human (taxid:9606)". Select option "Show results in a new window," and press the "Blast" button. It is important for us that the probe does not bind to nonspecific PCR products (if any exist) and binds to the single region of the target amplicon.

11. It is worthwhile to check primers and probes for complementarity to each other and for self-complementarity and hairpin formation in the OligoAnalyzer Tool software according to their recommendations (https://www.idtdna.com/pages/education/decoded/article/designing-pcr-primers-and-probes). In this software, we check a parameter called ΔG (change in Gibbs free energy) of secondary-structure formation. At ΔG values more positive than –9 kcal/mol, secondary structures do not have a significant effect on PCR, and values greater than zero indicate that under these conditions, secondary structures do not form (https://www.gene-quantification.de/oligo_architect_glossary.pdf). Therefore, when checking the primers and probes, we select those with ΔG ≥ –9 kcal/mol for potential secondary structures.

PCR execution and choosing PCR conditions

Fluorophore-labeled probes should be stored in the dark to avoid photobleaching (https://assets.thermofisher.com/TFS-Assets/LSG/Application-Notes/cms_043004.pdf).

Prepare the PCR mixture on ice; for one reaction you need:
– 10 µl of a buffer (we have used BioMaster HS-qPCR (2x), (Biolabmix, Russia), but it can be replaced with any available analog),
– 3.5 pmol of forward primer,
– 3.5 pmol of reverse primer,
– 1.5 pmol of FAM-labeled probe,
– 1.5 pmol of VIC-labeled probe,
– 10 ng of DNA,
– double-distilled H_2O up to 20 µl.

Mix all the listed components, except for the DNA sample, in a microtube while taking into consideration the number of the samples (with a 10 % excess of the mix). Place the DNA samples directly into wells of the PCR plate, then add 18 µl of the mix into each well, vortex, and centrifuge down.

Check the performance of the new primers and probes by means of several DNA samples and select optimal annealing temperatures first. Optimal concentrations of primers and probes may also differ from those given above, but the final concentrations of probes are usually at least 2 times lower than those of primers (https://www.bioline.com/mwdownloads/download/link/id/3301//p/i/pi-50201_sensifast_probe_hi-rox_one-step_kit_v11.pdf).

PCR program:
1. Initial denaturation, 95 °C for 3 min,
2. Amplification and detection (40 cycles):
   denaturation, 95 °C for 10 s,
   primer annealing and elongation with signal detection, 60 °C for 30 s.

The outcome of PCR with the chosen primers and probes for SNP rs11121704 is presented in Figure 7.

Because T_m of an oligonucleotide is the temperature at which half the population of oligo molecules is molten and half is double-stranded, the recommended annealing temperature should be approximately 5 °C lower than the lower T_m between the two primers, because under such conditions, both primers will bind almost completely to the complementary strands. In practice, due to possible inaccuracy of T_m calculation or discrepancies between the reaction conditions and the conditions for which the calculation was performed, the optimal temperature is selected empirically. We recommend checking the interval [Tm_{av} – 5 °C…Tm_{av} + 5 °C], where Tm_{av} is the average T_m value of the two primers. We also recommend choosing conditions for two PCR mix versions: with probes or with an intercalating dye, for example, SYBR Green. The melting curve plot will identify possible nonspecific PCR products.

Elongation typically takes ~1 min per 1,000 bp. Often, if the annealing temperature is greater than 60 °C, this step is combined with the previous one, and elongation occurs at the annealing temperature. Although the temperature optimum for most Taq polymerases is approximately 75–80 °C, elongation cannot occur at temperatures higher than the melting point of the probes.
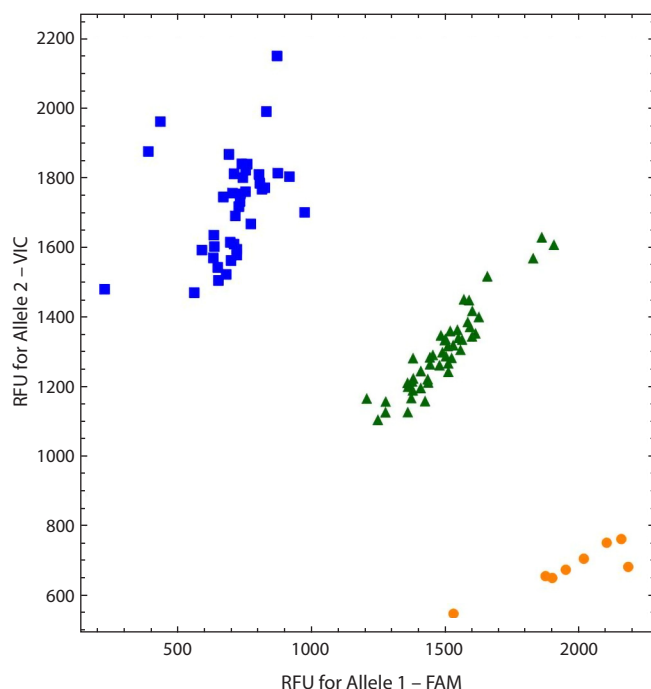
В.А. Девяткин, А.А. Шкляр, А.Ж. Фурсова
Ю.В. Румянцева, О.С. Кожевникова

Аллель-специфичная ПЦР с флуоресцентно-мечеными
зондами: критерии подбора праймеров для генотипирования

2024
28 • 3

Fig. 7. The result of allelic discrimination using probes for SNP rs11121704.

Orange dots represent homozygotes of the reference allele (C/C), green triangles represent heterozygotes (C/T), and blue squares represent homozygotes of the alternative allele (T/T).

Conclusion

Genotyping by allele-specific PCR is an effective and accurate way to detect genetic variants. Advantages of this method are its specificity, sensitivity, low cost, and quick results. It makes it possible to distinguish different alleles in the genome by one-step PCR without additional product separation steps; accordingly, it is particularly useful for genetic association studies in molecular genetics and medicine.

Thanks to developments in technologies for the synthesis of oligonucleotides and improvements in methods for designing primers and probes, we can expect expansion of the possibilities offered by this approach in the diagnosis of hereditary diseases. In this article, we discussed in detail the criteria and conditions for optimizing successful design primers and oligonucleotide probes for allele-specific PCR. We hope that the presented protocol will enable research groups to independently design their own effective assays for testing for polymorphisms of interest.

References

Broccanello C., Chiodi C., Funk A., McGrath J.M., Panella L., Stevanato P. Comparison of three PCR-based assays for SNP genotyping in plants. *Plant Methods*. 2018;14:28. DOI 10.1186/s13007-018-0295-6

Debode F., Marien A., Janssen E., Bragard C., Berben G. The influence of amplicon length on real-time PCR results. *Biotechnol. Agron. Soc. Environ*. 2017;27(1):3-11. DOI 10.25518/1780-4507.13461

Huang Q., Li Q. Characterization of the 5′ to 3′ nuclease activity of *Thermus aquaticus* DNA polymerase on fluorogenic double-stranded probes. *Mol. Cell. Probes*. 2009;23(3-4):188-194. DOI 10.1016/j.mcp.2009.04.002

Hui L., DelMonte T., Ranade K. Genotyping using the TaqMan assay. *Curr. Protoc. Hum. Genet*. 2008;56(2):2.10.1-2.10.8. DOI 10.1002/0471142905.hg0210s56

Kalendar R., Shustov A.V., Akhmetollayev I., Kairov U. Designing allele-specific competitive-extension PCR-based assays for high-throughput genotyping and gene characterization. *Front. Mol. Biosci*. 2022;9:773956. DOI 10.3389/fmolb.2022.773956

Owczarzy R., You Y., Groth C.L., Tataurov A.V. Stability and mismatch discrimination of locked nucleic acid-DNA duplexes. *Biochemistry*. 2011;50(43):9352-9367. DOI 10.1021/bi200904e

Ranade K., Chang M.S., Ting C.T., Pei D., Hsiao C.F., Olivier M., Pesich R., Hebert J., Chen Y.D., Dzau V.J., Curb D., Olshen R., Risch N., Cox D.R., Botstein D. High-throughput genotyping with single nucleotide polymorphisms. *Genome Res*. 2001;11(7):1262-1268. DOI 10.1101/gr.157801

You Y., Moreira B.G., Behlke M.A., Owczarzy R. Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res*. 2006;34(8):e60. DOI 10.1093/nar/gkl175