2020 • 24 • 8

Print ISSN 2500-0462
Online ISSN 2500-3259

Научный рецензируемый журнал

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.
Периодичность 8 выпусков в год
DOI 10.18699/VJ20.677

Учредители

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

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

Сибирское отделение Российской академии наук

2020 • 24 • 8

Print ISSN 2500-0462
Online ISSN 2500-3259

Scientific Peer Reviewed Journal

VAVILOV JOURNAL OF GENETICS AND BREEDING

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

Founded in 1997
Published 8 times annually
DOI 10.18699/VJ20.677

Founders
Federal State Budget Scientific Institution "The Federal Research Center Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences"
The Vavilov Society of Geneticists and Breeders
Siberian Branch of the Russian Academy of Sciences

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

СОДЕРЖАНИЕ • 2020 • 24 • 8

Генетика и селекция растений

Генетика и селекция животных

Генетика человека

Актуальные технологии

VAVILOV JOURNAL OF GENETICS AND BREEDING

CONTENTS • 2020 • 24 • 8

Plant genetics and breeding

Animal genetics and breeding

Human genetics

Mainstream technologies

Вавиловский журнал генетики и селекции. 2020;24(8):813-820
DOI 10.18699/VJ20.678

Physiological responses to water deficiency in bread wheat (*Triticum aestivum* L.) lines with genetically different leaf pubescence

S.V. Osipova[1, 4] ✉, A.V. Rudikovskii[1], A.V. Permyakov[1], E.G. Rudikovskaya[1], M.D. Permyakova[1], V.V. Verkhoturov[3], T.A. Pshenichnikova[2]

[1] Siberian Institute of Plant Physiology and Biochemistry of Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia
[2] Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] National Research Irkutsk State Technical University, Irkutsk, Russia
[4] Irkutsk State University, Irkutsk, Russia
✉ e-mail: svetlanaosipova2@mail.ru

Abstract. Studying the relationship between leaf pubescence and drought resistance is important for assessing *Triticum aestivum* L. genetic resources. The aim of the work was to assess resistance of common wheat genotypes with different composition and allelic state of genes that determine the leaf pubescence phenotype. We compared the drought resistance wheat variety Saratovskaya 29 (S29) with densely pubescent leaves, carrying the dominant alleles of the *Hl1* and *Hl3* genes, and two near isogenic lines, i: S29 *hl1*, *hl3* and i: S29 *Hl2*^{aesp}, with the introgression of the additional pubescence gene from diploid species *Aegilops speltoides*. Under controlled conditions of the climatic chamber, the photosynthetic pigments content, the activity of ascorbate-glutathione cycle enzymes and also the parameters of chlorophyll fluorescence used to assess the physiological state of the plants photosynthetic apparatus were studied in the leaves of S29 and the lines. Tolerance was evaluated using the comprehensive index D, calculated on the basis of the studied physiological characteristics. The recessive state of pubescence genes, as well as the introduction of the additional *Hl2*^{aesp} gene, led to a 6-fold decrease in D. Under the water deficit influence, the fluorescence parameters profile changed in the lines, and the viability index decreased compared with S29. Under drought, the activity of ascorbate peroxidase, glutathione reductase and dehydroascorbate reductase in the line i: S29 *hl1*, *hl3* decreased 1.9, 3.3 and 2.3 times, in the line i: S29 *Hl2*^{aesp} it decreased 1.8, 3.6 and 1.8 times respectively, compared with S29. In a hydroponic greenhouse, line productivity was studied. Compared with S29, the thousand grains mass in the line i: S29 *hl1*, *hl3* under water deficit was reduced. The productivity of the line i: S29 *Hl2*^{aesp} was significantly reduced regardless of water supply conditions in comparison with S29. Presumably, the revealed effects are associated with violations of cross-regulatory interactions between the proteins of the trichome formation network and transcription factors that regulate plant growth and stress response.

Key words: drought tolerance; leaf pubescent genes; isogenic lines; *Triticum aestivum* L.; chlorophyll fluorescence; ascorbate-glutathione cycle enzymes; productivity.

For citation: Osipova S.V., Rudikovskii A.V., Permyakov A.V., Rudikovskaya E.G., Permyakova M.D., Verkhoturov V.V., Pshenichnikova T.A. Physiological responses to water deficiency in bread wheat (*Triticum aestivum* L.) lines with genetically different leaf pubescence. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):813-820. DOI 10.18699/VJ20.678

Физиологические реакции линий пшеницы (*Triticum aestivum* L.) с генетически различным опушением листа на водный дефицит

С.В. Осипова[1, 4] ✉, А.В. Рудиковский[1], А.В. Пермяков[1], Е.Г. Рудиковская[1], М.Д. Пермякова[1], В.В. Верхотуров[3], Т.А. Пшеничникова[2]

[1] Сибирский институт физиологии и биохимии растений Сибирского отделения Российской академии наук, Иркутск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Иркутский национальный исследовательский технический университет, Иркутск, Россия
[4] Иркутский государственный университет, Иркутск, Россия
✉ e-mail: svetlanaosipova2@mail.ru

Аннотация. Изучение взаимосвязи опушения листьев и засухоустойчивости важно для оценки генетических ресурсов *Triticum aestivum* L. Целью работы был анализ устойчивости к дефициту воды генотипов мягкой пшеницы с различными составом и аллельным состоянием генов, определяющих фенотип опушения листьев. Мы сравнили засухоустойчивый сорт пшеницы Саратовская 29 (С29) с густо опушенными листьями, несущий доминантные аллели генов *Hl1* и *Hl3*, и две почти изогенные линии – i: С29 *hl1*, *hl3* и i: С29 *Hl2*^{aesp} с интрогрессией дополнительного гена опушения от диплоидного злака *Aegilops speltoides*. В контролируемых условиях климатической камеры изучены параметры флуоресценции хлорофилла, используемые для оценки физиологическо-

го состояния фотосинтетического аппарата растений, содержание фотосинтетических пигментов и активность ферментов аскорбат-глутатионового цикла в листьях С29 и линий. Устойчивость определяли с помощью комплексного индекса D, рассчитанного на основе изученных физиологических признаков. Рецессивное состояние генов опушения, как и введение дополнительного гена *Hl2*^{aesp}, привели к 6-кратному снижению значений D. Под воздействием водного дефицита у линий менялся профиль параметров флуоресценции и снижался индекс жизнеспособности по сравнению с С29. Активность аскорбатпероксидазы, глутатионредуктазы и дегидроаскорбатредуктазы в листьях линии i: С29 *hl1*, *hl3* уменьшалась в 1.9, 3.3 и 2.3 раза, в листьях линии i: С29 *Hl2*^{aesp} – в 1.8, 3.6 и 1.8 раза соответственно по сравнению с С29. В условиях гидропонной теплицы изучена продуктивность линий. По сравнению с С29 у линии i: С29 *hl1*, *hl3* при водном дефиците была снижена масса тысячи зерен. Продуктивность линии i: С29 *Hl2*^{aesp} была значительно ниже независимо от условий водоснабжения в сравнении с С29. Предполагается, что выявленные эффекты связаны с нарушениями перекрестных регуляторных взаимодействий белков сети формирования трихом и факторов транскрипции, которые контролируют рост растений и реакцию на стресс.
Ключевые слова: засухоустойчивость; гены опушения листа; изогенные линии; *Triticum aestivum* L.; флуоресценция хлорофилла; ферменты аскорбат-глутатионового цикла; продуктивность.

Introduction

The spring bread wheat (*Triticum aestivum* L.) variety Saratovskaya 29 (S29) is one of the most famous varieties created in Russia, as it has high drought tolerance and outstanding grain quality (Ilyina, 1989). These properties characterize S29 as a valuable genetic resource, used for obtaining not less than 155 other varieties. One of the characteristic features of the variety is the dense pubescence of the leaf blade. Among 47 genotypes of bread wheat and relative species studied for the diversity of this trait, the leaf pubescence in S29 was distinguished by its high density and trichomes length (Pshenichnikova et al., 2017). Obviously, such morphological adaptations make a significant contribution to the drought tolerance of this variety.

The trichomes are best known as excess sunlight reflectors (Ehleringer et al., 1976). The recent studies showed that trichomes can play a significant role in the water balance of leaves, affecting their wettability, droplet retention, and water absorption (Bickford, 2016). The dense trichomes layer can increase water use efficiency indirectly, promoting dew formation and reducing the difference in water potential inside the leaves and in the air. This allows stomata to be kept open longer, allowing for an influx of carbon dioxide without excessive water loss (Konrad et al., 2015).

Among the cultivated plant species, the physiological role of leaf pubescence is poorly studied. In *Oryza sativa* L. introgression of a chromosome segment from the wild species *Oryza nivara* increased leaf pubescence, reduced transpiration rate and increased water use efficiency due to increased stability of the boundary air layer (Hamaoka et al., 2017). The only experiment in *T. aestivum* L. showed that the stomatal conductivity and the photosynthetic rate in substituted and near-isogenic lines with genetically different leaf pubescence were inversely proportional to the density and trichomes length (Pshenichnikova et al., 2019).

In bread wheat, several genes are known today that determine a different phenotype of leaf pubescence. The *Hl1* and *Hl2* genes were localized and mapped on chromosomes 4B and 7B, respectively (Maystrenko, 1976; Taketa et al., 2002; Dobrovolskaya et al., 2007). The *Hl3* gene not yet assigned to a specific chromosome was genetically detected in the spring cultivar S29 (Doroshkov et al., 2011). In addition to them, the gene *Hl2*^{aesp} allelic to the gene *Hl2* was identified, introgressed into bread wheat from the species *Ae. speltoides* (Pshenichnikova et al., 2007). *Hl1* and *Hl3* affect to a greater extent on trichomes initiation and growth, while *Hl2* regulates the length of trichomes (Doroshkov et al., 2016). Knowledge of the relationship of these genes with the physiological characteristics of drought tolerance and grain productivity is necessary for their including in the breeding process.

Two near-isogenic lines with a different composition and allelic state of *Hl* genes were developed on the genetic base of the drought-tolerant wheat cultivar S29. The line i: S29 *hl1, hl3* carry the recessive alleles of *Hl1* and *Hl3* genes which are dominant in the recipient. The line i: S29 *Hl2*^{aesp} carries the gene for a long pubescence in addition to the two own dominant genes of the recipient. Previously, photosynthetic indicators were studied in these two lines under natural light and contrasting water supply. The lines were found to be contrast in terms of gas exchange (Pshenichnikova et al., 2019). However, no clear answer was obtained in respect of the pubescence influence on the parameters of chlorophyll fluorescence, which describe the physiological state of the plants photosynthetic apparatus (Goltsev et al., 2016).

The aim of this work was to assess the drought resistance of wheat by a wide range of physiological characteristics and productivity, depending on the presence of dominant or recessive alleles of the genes or the additional *Hl2*^{aesp} gene, which determine the phenotype of leaf pubescence. Among the physiological traits were chlorophyll fluorescence indicators, including the OJIP-test parameters, the content of photosynthetic pigments and the effectiveness of the ascorbate-glutathione cycle, which, as know, is a powerful defense of cellular structures from oxidative damage (Foyer, Shigeoka, 2011). The resistance to drought was assessed with using the comprehensive score of drought D (Cao et al., 2015), calculated on the basis of the tolerance indexes of physiological traits.

Materials and methods

Plant material. The object of the research was the drought-tolerant wheat spring cultivar S29 carrying two genes (*Hl1* and *Hl3*) for leaf pubescence and two near-isogenic lines with contrasting leaf pubescence. Line i: S29 *hl1, hl3* was obtained by crossing the S29 cultivar with the non-pubescent Rodina cultivar carrying the recessive alleles of these genes. In the process of 8-fold backcrossing on the recipient cultivar, non-pubescent plants were selected. Line i: S29 *Hl2*^{aesp} was

obtained by crossing the S29 cultivar with the introgressed line 102/00^{i}, which carries the *Hl2^{aesp}* gene from the *Ae. speltoides*. Then, an 8-fold backcrossing was carried out on the recipient cultivar with the selection of plants bearing introgressed pubescence. The line i: S29 *hl1, hl3* has a poor pubescence, while the leaves of the second line were densely pubescent. The origin, genetic characteristics and the quantitative characteristics of the pubescence in leaves of the near-isogenic lines have been described in detail earlier (Doroshkov et al., 2016; Pshenichnikova et al., 2019).

Experimental conditions. Physiological parameters were studied under controlled conditions of the climatic chamber CLF PlantMaster (CLF Plant Climatics GmbH, Wertingen, Germany), mounted in the phytotron of Siberian Institute of Plant Physiology and Biochemistry of Siberian Branch of the Russian Academy of Sciences (Irkutsk, Russia). The mixture consisted of humus, sand and peat (1: 1: 1) was used as soil for plant growing. A 16-hour photoperiod was maintained with a light intensity of 300 μmol (photon)/$m^{-2} \cdot s^{-1}$, a day/night temperature 23/16 °C and a relative humidity of 60 %. Each pot (19 cm diameter, 0.24 cm high, containing 4 kg soil) was planted with ten grains. For each line, one pot was maintained in a state of optimal water supply, which was 60 % of the total soil moisture capacity (control), while in the second pot, starting from the stage of three leaves, watering was limited until the water content in the soil decreased to 30 % from the full moisture capacity of the soil (water shortage or drought). This model of drought corresponds to the climatic conditions of Western and Eastern Siberia in the spring.

Yield components of the lines was studied in a hydroponic greenhouse in the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia) during two seasons. Plants were grown in the bathtubs (size: 4 × 1 × 0.35 m) filled with artificial soil "ceramzit" (expanded clay), Knop's solution was used for plant nutrition. The near-isogenic lines and S29 were grown in rows in two independent replicates consisted of seven plants. From seedlings to tillering stages, all plants in the bathtubs were watered twice a day. After the beginning of tillering, two water supply regimes were created in the bathtubs. At a control regime, plants were continued to water twice a day until the end of a season. In the second regime, water supply was stopped. Moisture level was measured once a week on the depth 6 cm using a moisture meter MG-44 ("AKVASENSOR", Kharkov, Ukraine). The moisture value in the control variant was 28–30 % in average during the season. In the second variant, the moisture level gradually decreased from the control level and after a month of drought reached the constant value of 10–12 % in average. The following yield components were measured: number of tillers, stem and spike length, the number and weight of grains in main and secondary spikes. Thousand grain weight was a calculated value.

Determination of Chl fluorescence parameters. The measurements of the Chl fluorescence of leaves were carried out using a portable impulse fluorometer PAM-2500 (Walz, Effelrich, Germany). A total of 33 Chl fluorescence parameters were measured and calculated. 13 of them were most sensitive to water scarcity, and are shown in Fig. 1. In order to register the minimal fluorescence yield of the dark-adapted state (F_0), we darkened the leaves for 30 min and then illuminated them with modulated measuring light of low frequency (5 Hz) and low intensity (630 nm). The intensity of the Chl fluorescence under conditions of closed reactive centers (F_m) was measured after the exposure of a light impulse of high intensity (25,000 μmol (photon)/$m^{-2} \cdot s^{-1}$, 630 nm). In addition, we calculated the rate of electron transport (ETR), the real quantum yield of PSII (Y(II)), quantum yield of unregulated fluorescence quenching (Y(NO)), coefficient of non-photochemical fluorescence quenching (qN), coefficient of photochemical fluorescence quenching (qP). Parameters lk and ETR_{max} were calculated from the Chl fluorescence light curve (PAR range from 0 to 2,000 μmol (photon)/$m^{-2} \cdot s^{-1}$).

The quantitative analysis of the characteristics of photosynthesis primary processes based on parameters of fluorescence kinetic curve was conducted using the OJIP-test, based on the theory of energy pathways (Strasser et al., 2004). The following parameters were calculated:

- $V_I = (F_{30ms} - F_0) / F_v$ – relative variable fluorescence at 30 ms;
- $PI^{abs} = (RC / ABS) \times [\varphi Po / (1 - \varphi Po)] \times [\Psi 0 / (1 - \Psi 0)]$ – performance index, an indicator of the functional activity of PSII;
- $Mo = 4 \times (F_{0.3ms} - F_0) / (F_m - F_0)$ – the parameter reflects the closing speed of the reaction centers of PSII;
- $Rfd = (F_m - F_t) / F_t$ – viability index (Lichtenthaler et al., 2005).

Determination of photosynthetic pigments content and enzymes activity in leaves. After determining the photosynthetic parameters, the leaf pieces were frozen with liquid nitrogen and stored at the temperature of –80 °C. The content of pigments per gram of leaves dry mass and activities of superoxide dismutase (SOD), glutathione reductase (GR), dehydroascorbate reductase (DHAR) and ascorbate peroxidase (APX) were determined and calculated as it was previously described (Osipova et al., 2016).

Statistical analysis. Chl fluorescence was measured on the flag leaves of four plants per line. The content of pigments and the enzymes activity were determined in three biological and three analytical replicates. One plant of each line was taken for the biological replicate. Yield components were studied in each season, in two replicates; in all, the measurements were made for twenty-four plants of each line under drought and in control conditions. All the comparisons were made with S29. Microsoft Excel 2010 (Microsoft Corp., Redmond, WA, USA) was used for data processing and histogram plotting. The statistical significance of the differences between the recipient variety and the wheat lines from the measured parameters was compared with the Student's test. Means were considered to be significantly different when $p < 0.05$. The statistics package PAST (Hammer et al., 2001) was used for principal component analysis (PCA). The drought tolerance index (IT, %) for each parameter was calculated as shown in the following formula:

$$IT\ (\%) = \frac{\text{the value of the drought}}{\text{the value of the control}} \times 100\ \%.$$

The data from PCA were used in further calculations of comprehensive drought tolerance values D (Cao et al., 2015).

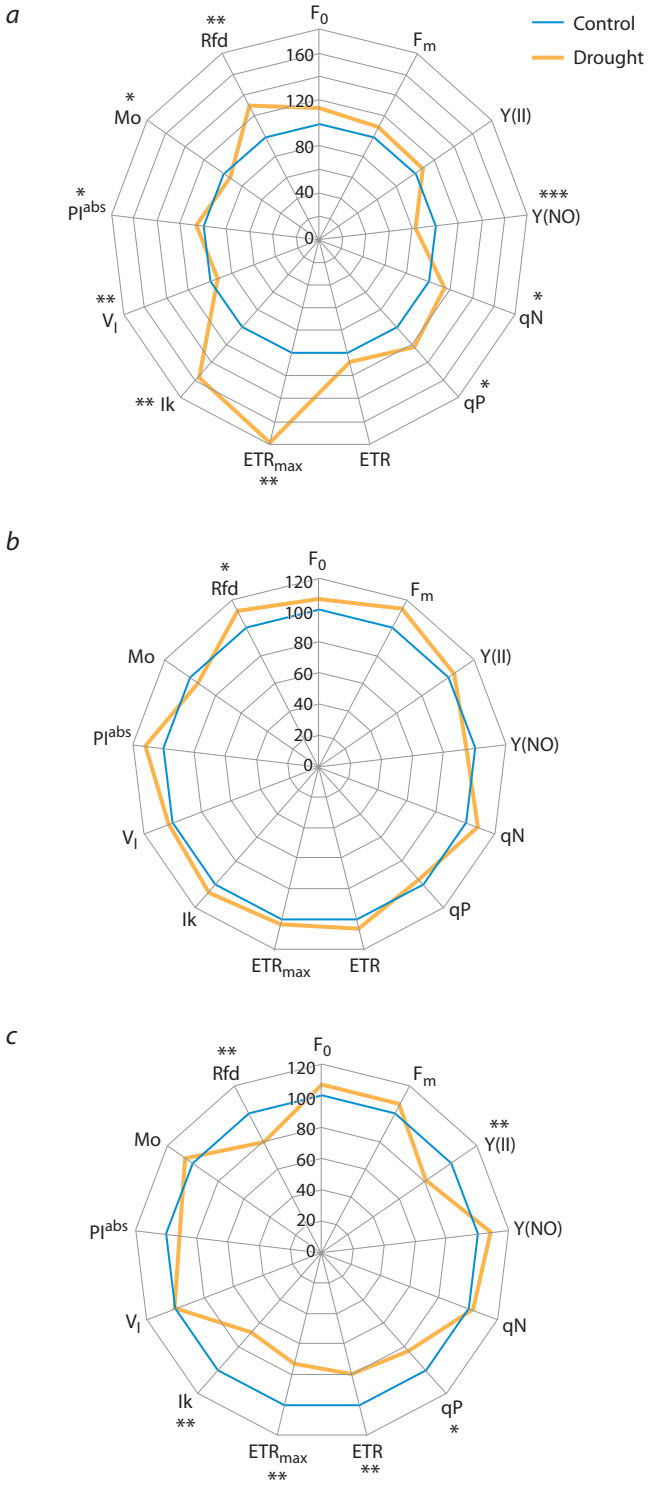


Fig. 1. Relative deviation of chlorophyll fluorescence parameters (in %) under drought compared to control (watering 100%) in S29 (*a*) and lines i: S29 *Hl2^{aesp}* (*b*) and i: S29 *hl1, hl3* (*c*).

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

F_0 – minimal fluorescence yield of the dark-adapted state; F_m – maximal fluorescence yield of the dark-adapted state; Y(II) – real quantum yield of PSII; Y(NO) – quantum yield of unregulated fluorescence quenching; qN – coefficient of non-photochemical fluorescence quenching; qP –coefficient of photochemical fluorescence quenching; ETR – rate of electron transport provided by PSII; ETR_{max} – maximum electron transport rate; lk – intensity of illumination, expressing the beginning of PAR saturation; V_I – relative variable fluorescence at 30 ms; Mo relative the closing rate of the reaction centers of PSII; PI^{abs} – PSII performance index; Rfd – PSII vitality index.

Results

Effect of water deficit on fluorescence parameters of S29 and lines i: S29 *Hl2^{aesp}* and i: S29 *hl1, hl3*. The reaction of the photosynthetic apparatus to water deficiency was significantly different in the studied wheat genotypes. The thirteen parameters most sensitive to water deficit are shown in Fig. 1.

In S29, the most noticeable changes were an increase of ETR_{max} and lk, statistically significant increase in qP, qN, PI^{abs} and vitality index Rfd. Parameters Y(NO), Mo, and V_I in variety S29 decreased under conditions of water deficiency (see Fig. 1, *a*). In the i: S29 *Hl2^{aesp}* line the chlorophyll fluorescence parameters remained unchanged under water deficiency, with the exception of an increase in the viability index Rfd (see Fig. 1, *b*). In line i: S29 *hl1, hl3*, under water deficiency, the parameters Y(II), qP, ETR, ETR_{max}, lk, and Rfd decreased statistically significantly compared to the control (see Fig. 1, *c*).

Effect of water deficit on the antioxidant enzymes activities in leaves of S29 and lines i: S29 *Hl2^{aesp}* and i: S29 *hl1, hl3*. Under adaptation to water deficiency of cv. S29, the activity of APX, GR, and DHAR in leaves were higher than under optimal conditions (Fig. 2, *b–d*).

In the leaves of line i: S29 *Hl2^{aesp}* the activities of this enzymes were significantly lower compared to S29 under drought conditions; moreover, GR and DHAR activities in this line were lower during drought compared to the control. The activities of APX, GR and DHAR in the line i: S29 *hl1, hl3* were significantly reduced compared to S29 regardless of water supply conditions. SOD activity was also reduced compared to S29 under drought conditions in lines, most significantly in the line i: S29 *hl1, hl3* (see Fig. 2, *a*).

Effects of water deficit stress on the photosynthetic pigments content in leaves of S29 and lines i: S29 *Hl2^{aesp}* and i: S29 *hl1, hl3*. The content of chlorophylls and carotenoids in the leaves of S29 did not change depending on the water supply conditions (Suppl. Material 1)[1]. Under optimal irrigation conditions, the lines significantly exceeded the initial variety in the content of photosynthetic pigments. Under conditions of water deficiency, the content of chlorophylls and carotenoids in the lines decreased. In the line i: S29 *Hl2^{aesp}*, the decrease in the content of chlorophyll b and carotenoids was significantly lower than in S29. Regardless of the conditions, the ratio chlorophyll a+b/carotenoids was higher in the line i: S29 *hl1, hl3* compared to the original variety. This is due to the higher content of chlorophylls in the leaves of this line. The tolerance index of photosynthetic pigment content in both lines was reduced compared to S29 (Suppl. Material 2).

Principal component analysis and calculation of the comprehensive evaluation value. The drought tolerance coefficients for 14 physiological traits were involved into PCA (see Suppl. 2). The cumulative contribution rates of PC_1 and PC_2 accounted for 100 % of the total variation (Suppl. Material 3). PC_1 accounted for 84.9 % of total variation and was constituted mainly by ITs of GR and DHAR activities and ETR_{max}. PC_2 explained 15.1 % of the total variation with SOD and DHAR activities, ETR_{max} and lk being the largest

[1] Supplementary Materials 1–3 are available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/SupplOsipova_engl.pdf

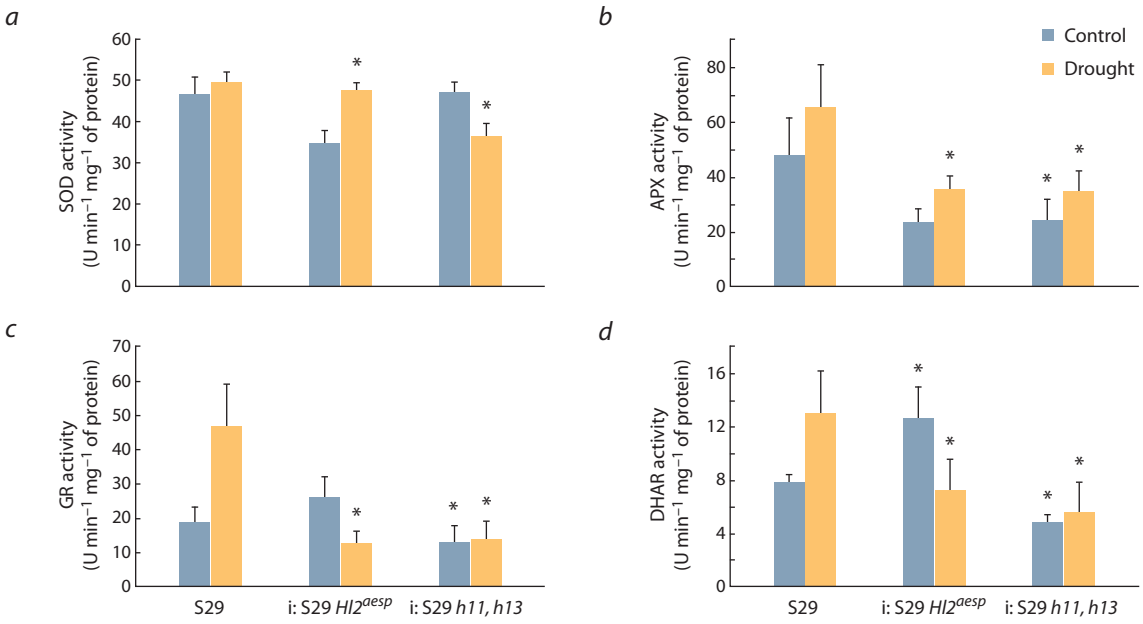


Fig. 2. The average activity of superoxide dismutase (SOD), ascorbate peroxidase (APX), glutathione reductase (GR) and dehydroascorbate reductase (DHAR) in the leaves of S29 and lines i: S29 *Hl2*^{aesp} and i: S29 *hl1, hl3* under normal irrigation and drought.

* Significant differences with S29, $p < 0.05$.

contributors. Using the formulas of X. Cao et al. (2015), the comprehensive evaluation value, D, was calculated. The D value indicated the relative level of drought tolerance in the different wheat genotypes subjected to drought stress. Based on this criterion, S29 with a D value of 0.948 had the highest drought tolerance. The lines i: S29 *Hl2*^{aesp} and i: S29 *hl1, hl3* had D value of 0.150 and 0.160, respectively. Thus, a comprehensive value D, which takes into account 14 physiological indicators, showed that both near-isogenic lines had a significantly reduced level of resistance compared to the original cultivar. The greatest contribution to these differences was made by such indicators as the activity of GR and DHAR, as well as the parameters of the light curve ETR_{max} and lk.

The productivity evaluation of S29 and lines i: S29 *Hl2*^{aesp} and i: S29 *hl1, hl3* in different conditions of water supply. The recipient variety significantly exceeded the line with the additional gene *Hl2*^{aesp} for leaf pubescence. Most of the yield components of the line i: S29 *Hl2*^{aesp} were significantly reduced, regardless of the water supply conditions (Table).

Inhibition of plant development was observed starting from tillering; a reduced yield was formed both on the primary and secondary spikes. This line was also significantly inferior in productivity to the line with recessive genes *Hl1* and *Hl3* for leaf pubescence. Cv. S29 reduced the productivity of secondary tillers under drought, but thousand grain weight decreased slightly. The line i: S29 *hl1, hl3* differed from S29 in reduced productivity of the secondary tillers under irrigation conditions (see Table). Under drought, it exceeded the recipient in length of the stem and the main spike, the number of grains in the secondary spikes and the total number of grains. However, thousand grain weight was lower compared to the original cultivar. That is, the line i: S29 *hl1, hl3* formed smaller grains under water deficiency.

Discussion

Dense pubescence of the leaf blade is a morphological component of adaptation of cv. S29 to drought conditions (Ilyina, 1989; Pshenichnikova et al., 2017, 2019). On its basis, the two near-isogenic lines with genetically modified morphology of leaf pubescence were obtained. The line i: S29 *hl1, hl3* with recessive genes for this trait is characterized by a significant decrease in the density (6.6–14 times) and length (2.5–4.7 times) of trichomes on different sides of the leaf and under different conditions compared with S29. In line i: S29 *Hl2*^{aesp} with gene introgression from *Ae. speltoides*, the density of trichomes on different sides of the leaf and under different conditions increased 1.08–1.17 times, and the length of trichomes increased 1.6 times compared to the recipient (Doroshkov et al., 2016; Pshenichnikova et al., 2019). The lines are a convenient model for studying the role that pubescence genes play in wheat stress tolerance. Previously, we used the lines to assess the relationship between the density and trichomes length and gas exchange parameters (Pshenichnikova et al., 2019). In a greenhouse with natural light, the transpiration rate, stomatal conductivity, and the rate of photosynthesis of S29 and lines were inversely proportional to the density and length of trichomes, which is consistent with the data of N. Hamaoka et al. (2017) for rice. The highest water use efficiency, calculated as the relation photosynthesis rate / transpiration rate was in S29. The water use efficiency at line i: S29 *hl1, hl3* was 1.9 times lower under optimal conditions and 1.5 times lower under drought compared to S29, since increased transpiration led to water loss (Pshenichnikova et al., 2019).

An analysis of the chlorophyll fluorescence parameters in this experiment showed that, when adapting to drought, S29 was characterized by a significant increase in the light curve parameters ETR_{max} and lk, an increase in the qP parameter, and a significant (30%) increase in the PSII viability coef-

Average values of yield components in cv. Saratovskaya 29 and the lines i: S29 *Hl2^{aesp}* and i: S29 *hl1*, *hl3* under normal watering and drought grown under hydroponic green-house conditions on artificial soil

Yield component	Saratovskaya 29			i: S29 *Hl2^{aesp}*			i: S29 *hl1*, *hl3*		
	Normal	Drought	IT, %	Normal	Drought	IT, %	Normal	Drought	IT, %
Number of tillers	4.8 ± 1.0	2.3 ± 0.6	47.9	3.8 ± 0.8***	2.4 ± 0,7	63.2	4.3 ± 1.0	2.6± 0.6*	60.5
Stem length, cm	95.8 ± 6.8	79.9 ± 10.4	83.4	89.6 ± 9.1*	82.2 ± 11.0	91.7	97.9 ± 9.5	89.4 ± 7.2*	91.3
The main spike									
Length, cm	7.4 ± 0.4	6.7 ± 0.6	90.5	7.2 ± 0.6	6.7 ± 0.7	93.1	7.6 ± 0.5	7.2 ± 0.4***	94.7
Grain number	26.5 ± 3.1	25.8 ± 3.7	97.4	22.5 ±3.9***	21.2 ± 6.5**	94.2	25.3 ±4.6	25.9 ±3.5	102.0
Grain weight, g	1.12 ± 0.2	0.88 ± 0.2	78.6	0.84 ± 0.2***	0.65 ± 0.2***	77.4	1.09 ± 0.2	0.88 ± 0.2	80.7
Secondary tillers									
Grain number	77.3 ± 18.6	26.4 ± 7.4	34.2	49.6 ± 16.2***	20.8 ± 8.7*	41.9	65.8 ± 21.4	34.0 ± 10.5**	51.7
Grain weight, g	2.9 ± 0.9	0.83 ± 0.3	28.6	1.6 ± 0.4***	0.67 ± 0.2*	41.9	2.5 ± 0.8*	1.0 ± 0.4	40.0
Total productivity of plant									
Grain number	101.3 ± 24.7	50.1 ± 8.5	49.5	61.8 ± 26.7***	39.2 ± 15.9**	63.4	93.6 ± 26.5	57.6 ± 12.3*	61.5
Grain weight, g	4.0 ± 1.0	1.7 ± 0.4	42.5	2.5 ± 0.5***	1.3 ± 0.5**	52.0	3.6 ± 1.0	1.9 ± 0.6	52.8
Thousand grain weight, g	39.1 ± 5.6	33.8 ± 4.8	86.4	34.8 ± 3.9**	29.8 ± 5.9*	85.6	39.7 ± 3.6	30.4 ± 4.2*	76.6

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ in comparison with S29 on corresponding watering regime.

ficient Rfd. These data indicate that S29 can stably support PSII functions, increasing the fraction of light energy used for photochemical reactions and the rate of assimilation of photosynthetic CO_2 under drought conditions (Lichtenthaler et al., 2005). The content of photosynthetic pigments was also stable. A significant increase in APX, DHAR, and GR activity was observed in S29 leaves under drought which contributed to the maintenance of structural and functional the integrity of the photosynthetic apparatus and the maintenance of the ascorbic acid (Asc) pool (Foyer, Shigeoka, 2011). Under conditions of water deficiency Asc, in addition to the antioxidant role, can be the donor of electrons in the photosynthetic electron transport chain (Tóth et al., 2013). Thus, at the cellular level, the high drought tolerance of S29 was associated with a high antioxidant ability and preservation of the functions of the photosynthetic apparatus (PhA).

Introduction of the additional pubescence gene *Hl2^{aesp}* into the genotype of S29 led to an increase in the length of trichomes (Pshenichnikova et al., 2019) and significant changes in the physiological responses to water deficiency. Unlike S29, the chlorophyll fluorescence parameters of the line i: S29 *Hl2^{aesp}* did not change under drought compared to optimal conditions except for a slight increase in the Rfd index. At the same time, APX, GR, and DHAR activities in the line were reduced 1.8, 3.6, and 1.8 times, respectively, compared with the recipient. Since maintaining the redox state of Asc through recycling is critical under stressful conditions (Gallie, 2013), a significant decrease in DHAR activity in the line i: S29 *Hl2^{aesp}* could lead to a decrease in the Asc content. At low concentrations of Asc, the activity of APX into chloroplasts is rapidly lost in the presence of H_2O_2. These in turn limits the effectiveness of photosynthesis under stressful conditions (Ishikawa, Shigeoka, 2008). The high content of chlorophylls and carotenoids comparable to the recipient S29 under optimal conditions did not retained under drought. It is likely that a significant decrease in the line productivity, both under favorable conditions and under drought, is associated with the observed inhibition of physiological processes.

The recessive state of the *Hl1* and *Hl3* genes in the line i: S29 *hl1, hl3* also led to a significant weakening of the antioxidant potential. As in the previous line, SOD, APX, GR, and DHAR activities under drought were reduced at the same manner: 1.4, 1.9, 3.3, and 2.3 times, respectively, compared with the recipient S29. Chlorophyll fluorescence parameters indicated disturbances in the functioning of PhA under stress, since the real efficacy of PSII, ETR and ETR_{max}, PSII viability coefficient (Rfd), and photosynthetic fluorescence quenching (qP) significantly decreased. The content of leaf pigments also decreased. Trichomes formation and accumulation of phenolic compounds are interconnected at the molecular level (Pattanaik et al., 2014; Zhang, Schrader, 2017). Due to the diffuse deposition of phenolic compounds in the cell walls, trichomes provide a protection against ultraviolet radiation by acting as optical filters, shielding wavelengths that can damage sensitive tissues (Karabourniotis et al., 2020). Therefore a further increase of a light load may lead to even more dramatic changes in the operation of the photosynthetic apparatus of the i: S29 *hl1, hl3* line. Changes in yield components of the line i: S29 *hl1, hl3* were less pronounced compared to the line i: S29 *Hl2^{aesp}*. Under drought, it was even more productive than the recipient cultivar. The increase in productivity was due to the number of grains of the secondary spikes. However, the line gave smaller grains which negatively affects the output of flour. It can be assumed that the decrease in the

stability of the photosynthetic apparatus found in the line led to a disruption in the synthesis of simple carbohydrates in line i: S29 *hl1, hl3*. This, in turn, reduced the level of starch synthesis associated with 1,000 grain mass and productivity (Wang et al., 2019). The values of the comprehensive drought tolerance index D, calculated on the basis of physiological parameters, in the lines were 6 times lower compared to S29.

The genetic regulation of trichomes formation in wheat has not been studied enough to unambiguously explain the reasons for the negative impact of manipulations with the *Hl1*, *Hl3*, and *Hl2^{aesp}* genes on tolerance to water deficiency. A well-studied genetic network for the development and differentiation of Arabidopsis trichomes may be a model in this regard. Dozens of genes are involved in this network. The vast majority of the products of these genes are transcription factors. They are components of the regulatory network of trichomes initiation, root hairs formation, and flavonoid biosynthesis involved in a large number of cross-regulatory protein-protein interactions (Pesch, Hülskamp, 2004, 2009; Pattanaik et al., 2014; Zhang, Schrader, 2017). For example, P. Achard et al. (2008) showed that transcription levels of Cu/Zn superoxide dismutase are positively modulated by proteins of the DELLA regulatory protein family. However, DELLA proteins interact with the WD-repeat/bHLH/MYB complex, which is involved in the regulation of development of trichomes (Qi et al., 2014). The transcription factors GIS and GIS2 play an important role in the integration of cytokinin and gibberellin signaling and have regulatory interactions with the proteins of the trichomes initiation network GL1, SRY and GL3, thereby affecting the functioning of the initiating complex of trichomes formation (Gan et al., 2007). These and other examples available in the literature indicate that genes that control the development of trichomes are linked by cross-regulatory interactions with transcription factors that regulate hormonal signaling, stress responses, including antioxidant response and developmental programs. Based on knowledge of the regulation of trichomes formation in Arabidopsis, we assume that the effects identified in our work, namely, the negative impact on physiological stability and yield of wheat of the recessive state of the *Hl1* and *Hl3* genes or the introgression of the *Hl2^{aesp}* gene, are probably associated with violations of cross-regulatory protein-protein interactions prevailing in the genotype of the recipient cultivar S29.

Conclusion

Changes in the composition and allelic state of *Hl* genes influenced not only the quantitative characteristics of leaf pubescence, but also stability of photosynthetic pigments content, chlorophyll fluorescence indexes, activity of ascorbate-glutathione cycle enzymes, and productivity of near-isogenic lines of bread wheat. The comprehensive drought tolerance index D, calculated on the basis of physiological indicators, was 6 times lower in the lines compared to S29. Regardless of the water supply conditions, all yield components significantly decreased in the line i: S29 *Hl2^{aesp}*, and in the line i: S29 *hl1, hl3* the weight of 1,000 grains decreased as compared to S29. It is assumed that these effects are associated with changes in the cross regulatory interactions of proteins involved in the formation of trichomes, and transcription factors that regulate growth, development, and reactions to stress factors.

References

Achard P., Renou J-P., Berthomé R., Harberd N.P., Genschik P. Plant DELLAs restrain growth and promote survival of adversity by reducing the levels of reactive oxygen species. *Curr. Biol.* 2008;18:656-660. DOI 10.1016/j.cub.2008.04.034.

Bickford C.B. Ecophysiology of leaf trichomes. *Funct. Plant Biol.* 2016;43(9):807-814. Available at: https://digital.kenyon.edu/biology_publications/85.

Cao X., Jiang F., Wang X., Zang Y., Wu Z. Comprehensive evaluation and screening for chilling-tolerance in tomato lines at the seedling stage. *Euphytica.* 2015;205:569-584. DOI 10.1007/s10681-015-1433-0.

Dobrovolskaya O.B., Pshenichnikova T.A., Arbuzova V.S., Lohwasser U., Röder M.S., Börner A. Molecular mapping of genes determining hairy leaf character in common wheat with respect to other species of the Triticeae. *Euphytica.* 2007;155(3):285-293. DOI 10.1007/s10681-006-9329-7.

Doroshkov A.V., Afonnikov D.A., Dobrovolskaya O.B., Pshenichnikova T.A. Interactions between leaf pubescence genes in bread wheat as assessed by high throughput phenotyping. *Euphytica.* 2016;207:491-500. DOI 10.1007/s10681-015-1520-2.

Doroshkov A.V., Pshenichnikova T.A., Afonnikov D.A. Morphological and genetic characteristics of leaf hairiness in wheat (*Triticum aestivum* L.) as analyzed by computer aided phenotyping. *Russian J. Genet*. 2011;47:739-743. DOI 10.1134/S1022795411060093.

Ehleringer J., Björkman O., Mooney H. Leaf pubescence: effects on absorptance and photosynthesis in a desert shrub. *Science.* 1976;192(4237):376-377. Available at: https://doi.org/10.1126/science.192.4237.376.

Foyer C.H., Shigeoka S. Understanding oxidative stress and antioxidant functions to enhance photosynthesis. *Plant Physiol.* 2011;155:93-100. DOI 10.1104/pp.110.166181.

Gallie D.R. The role of L-ascorbic acid recycling in responding to environmental stress and in promoting plant growth. *J. Exp. Bot.* 2013;64(2):433-43. DOI 10.1093/jxb/ers330.

Gan Y., Liu C., Yu H., Broun P. Integration of cytokinin and gibberellin signalling by Arabidopsis transcription factors GIS, ZFP8 and GIS2 in the regulation of epidermal cell fate. *Development.* 2007;134(11):2073-2081. DOI 10.1242/dev.005017.

Goltsev V.N., Kalaji H.M., Paunov M., Baba W., Horaczek T., Mojski J., Kociel H., Allakhverdiev S.I. Variable chlorophyll fluorescence and its use for assessing physiological condition of plant photosynthetic apparatus. *Russ. J. Plant Physiol.* 2016;63:869-893. DOI 10.1134/S1021443716050058.

Hamaoka N., Yasui H., Yamagata Y., Inoue Y., Furuya N., Araki T., Ueno O., Yoshimura A. A hairy-leaf gene, BLANKET LEAF, of wild *Oryza nivara* increases photosynthetic water use efficiency in rice. *Rice.* 2017;10(1):20. DOI 10.1186/s12284-017-0158-1.

Hammer O., Harper D.A.T., Ryan P.D. PAST: paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 2001. http://palaeo-electronica.org/2001_1/past/issue1_01.htm.

Ilyina L.G. Breeding of spring bread wheat in southeastern regions. Saratov: Saratov University Publ., 1989. (in Russian)

Ishikawa T., Shigeoka S. Recent advances in ascorbate biosynthesis and the physiological significance of ascorbate peroxidase in photosynthesizing organisms. *Biosci. Biotechnol. Biochem.* 2008;72:1143-1154. DOI 10.1271/bbb.80062.

Karabourniotis G., Liakopoulos G., Nikolopoulos D., Bresta P. Protective and defensive roles of non-glandular trichomes against multiple stresses: structure–function coordination. *J. For. Res.* 2020;31:1-12. DOI 10.1007/s11676-019-01034-4.

Konrad W., Burkhardt J., Ebner M., Roth-Nebelsick A. Leaf pubescence as a possibility to increase water use efficiency by promoting condensation. *Ecohydrology.* 2015;8:480-492. DOI 10.1002/eco.1518.

Lichtenthaler H.K., Buschmann C., Knapp M. How to correctly determine the different chlorophyll fluorescence parameters and

the chlorophyll fluorescence decrease ratio Rfd of leaves with the PAM fluorometer. *Photosynthetica.* 2005;43:379-393. DOI 10.1007/s11099-005-0062-6.

Maistrenko O.I. Identification and localization of genes controlling leaf pubescence of young common wheat plants. *Russian J. Genet.* 1976; 12(1):5-15.

Osipova S., Permyakov A., Permyakova M., Pshenichnikova T., Verkhoturov V., Rudikovsky A., Rudikovskaya E., Shishparenok A., Doroshkov A., Börner A. Regions of the bread wheat D genome associated with variation in key photosynthesis traits and shoot biomass under both well watered and water deficient conditions. *J. Appl. Genet.* 2016;6:553-559. DOI 10.1007/s13353-015-0315-4.

Pattanaik S., Patra B., Singh S.K., Yuan L. An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. *Front. Plant Sci.* 2014;5:259. DOI 10.3389/fpls.2014.00259.

Pesch M., Hülskamp M. Creating a two-dimensional pattern de novo during Arabidopsis trichome and root hair initiation. *Curr. Opin. Genet. Dev.* 2004;14:422-427. DOI 10.1016/j.gde.2004.06.007.

Pesch M., Hülskamp M. One, two, three... models for trichome patterning in Arabidopsis? *Curr. Opin. Plant Biol.* 2009;12(5):587-592. DOI 10.1016/j.pbi.2009.07.015.

Pshenichnikova T.A., Doroshkov A.V., Osipova S.V., Permyakov A.V., Permyakova M.D., Efimov V.M., Afonnikov D.A. Quantitative characteristics of pubescence in wheat (*Triticum aestivum* L.) are associated with the parameters of gas exchange and chlorophyll fluorescence under conditions of normal and limited water supply. *Planta.* 2019;249(3):839-847. DOI 10.1007/s00425-018-3049-9.

Pshenichnikova T.A., Doroshkov A.V., Simonov A.V., Afonnikov D.A., Börner A. Diversity of leaf pubescence in bread wheat and relative species. *Genet. Resour. Crop Evol.* 2017;64:1761-1773. DOI 10.1007/s10722-016-0471-3.

Pshenichnikova T.A., Lapochkina I.F., Shchukina L.V. The inheritance of morphological and biochemical traits introgressed into common wheat (*Triticum aestivum* L.) from *Aegilops speltoides* Tausch. *Genet. Resour. Crop. Evol.* 2007;54:287. DOI 10.1007/s10722-005-4499-z.

Qi T., Huang H., Wu D., Yan, Qi Y., Song S, Xie D. Arabidopsis DELLA and JAZ proteins bind the WD-Repeat/bHLH/MYB complex to modulate gibberellin and jasmonate signaling synergy. *Plant Cell.* 2014;26(3):1118-1133. DOI 10.1105/tpc.113.121731.

Strasser R.J., Tsimilli-Michael M., Sriyastaya A. Analysis of the chlorophyll a fluorescence transient. In: Papageorgiou G.C., Govindjee (Eds.). Chlorophyll a fluorescence: a signature of photosynthesis, advances in photosynthesis and respiration. Springer, Dordrecht, 2004;19:321-362.

Taketa S., Chang C.L., Ishii M., Takeda K. Chromosome arm location of the gene controlling leaf pubescence of a Chinese local wheat cultivar 'Hong-mang-mai'. *Euphytica.* 2002;125:141-147. DOI 10.1023/A:1015812907111.

Tóth S.Z., Schansker G., Garab G. The physiological roles and metabolism of ascorbate in chloroplasts. *Physiol. Plant.* 2013;148(2): 161-175. DOI 10.1111/ppl.12006.

Wang Y., Hou J., Liu H., Li T., Wang K., Hao Ch., Liu H., Zhang X. *TaBT1*, affecting starch synthesis and thousand kernel weight, underwent strong selection during wheat improvement. *J. Exp. Bot.* 2019; 70(5):1497-1511. DOI 10.1093/jxb/erz032.

Zhang B., Schrader A. TRANSPARENT TESTA GLABRA 1-dependent regulation of flavonoid biosynthesis. *Plants.* 2017;6(4):65. DOI 10.3390/plants6040065.

ORCID ID

S.V. Osipova orcid.org/0000-0001-7025-8590
A.V. Rudikovskii orcid.org/0000-0001-8567-0318
A.V. Permyakov orcid.org/0000-0003-0002-5470
E.G. Rudikovskaya orcid.org/0000-0001-5709-2370
M.D. Permyakova orcid.org/0000-0002-0259-8531
V.V. Verkhoturov orcid.org/0000-0003-2979-9867
T.A. Pshenichnikova orcid.org/0000-0001-5639-916X

Acknowledgements. This work was supported by the Russian Foundation for Basic Research (grant No. 18-04-00481).

The research was carried out using the equipment of The Core Facilities Center "Bioanalitika" of Siberian Institute of Plant Physiology and Biochemistry of Siberian Branch of the Russian Academy of Sciences (Irkutsk, Russia). The used bread wheat lines were grown in the Common Use Center "Laboratory of artificial plant growing" of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences with the support of state budget project No. 0324-2019-0039-C-01 and are maintained in Institute's of Cytology and Genetics collection GenAgro (Novosibirsk, Russia). We are grateful to Nina Kovaleva and Angela Chistyakova for excellent technical assistance.

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

Received June 23, 2020. Revised August 28, 2020. Accepted September 6, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):821-828
DOI 10.18699/VJ20.679

Возможности и перспективы формирования генетической защиты мягкой пшеницы от стеблевой ржавчины в Западной Сибири

В.Н. Кельбин ✉, Е.С. Сколотнева, Е.А. Салина

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ e-mail: kelbin@bionet.nsc.ru

Аннотация. Современные исследования проблемы устойчивости мягкой пшеницы к стеблевой ржавчине включают два основных направления: оценку устойчивости коллекций мягкой пшеницы к заболеванию с помощью молекулярных маркеров к известным генам устойчивости в дополнение к полевому скринингу материала и лабораторным тестам к образцам различных популяций гриба; поиск источников и доноров новых генов и генных локусов, в том числе среди культурных и дикорастущих родичей пшеницы. Для достижения адекватного генетического контроля заболевания важен интегральный подход, включающий как данные об источниках устойчивости, так и актуальные сведения о действующих в регионе патогенных популяциях, их расовом составе и динамике генов вирулентности. Результаты анализа экспериментальных данных полевого скрининга устойчивости к стеблевой ржавчине сортов мягкой пшеницы из коллекции питомников CIMMYT в условиях Омской и Новосибирской областей, а также лабораторного тестирования образцов инфекции на международном наборе пшеничных линий-дифференциаторов позволяют предполагать, что на территории Западной Сибири и Алтайского края существует обособленная, «азиатская», популяция *Puccinia graminis* f. sp. *tritici*. При этом практический интерес для современных программ опережающей селекции пшеницы на иммунитет к стеблевой ржавчине в условиях Западной Сибири представляют гены устойчивости *Sr2*, *Sr6Ai#2*, *Sr24*, *Sr25*, *Sr26*, *Sr31*, *Sr39*, *Sr40*, *Sr44* и *Sr57*. В настоящем обзоре проанализированы источники генов, сохраняющих эффективность к западносибирской популяции *P. graminis*, с целью упрощения первичного этапа отбора селекционного материала для создания устойчивого генотипа путем пирамидирования генов. Описаны основные требования, предъявляемые к фитопатологическому тестированию селекционного материала. Составлен список молекулярных маркеров к указанным генам устойчивости – как широко применяющихся в маркер-ориентированной селекции, так и требующих верификации.
Ключевые слова: мягкая пшеница; стеблевая ржавчина; гены устойчивости; маркер-ориентированная селекция; фитопатологическое тестирование.

Для цитирования: Кельбин В.Н., Сколотнева Е.С., Салина Е.А. Возможности и перспективы формирования генетической защиты мягкой пшеницы от стеблевой ржавчины в Западной Сибири. *Вавиловский журнал генетики и селекции*. 2020;24(8):821-828. DOI 10.18699/VJ20.679

Challenges and prospects for developing genetic resistance in common wheat against stem rust in Western Siberia

V.N. Kelbin ✉, E.S. Skolotneva, E.A. Salina

Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ e-mail: kelbin@bionet.nsc.ru

Abstract. Current studies on bread wheat resistance to stem rust have two main subjects: complex analysis for resistance of bread wheat germplasm using molecular markers, field screening and laboratory tests against samples of different fungal populations, and searching for sources and donors of new genes and gene loci, including cultivated and wild relatives of wheat. To achieve adequate genetic control of the disease, an integral approach is important, incorporating both data on sources of resistance and relevant information on pathogenic populations existing in the region, their race composition and dynamics of virulence genes. The analysis of experimental data on field screening of bread wheat varieties from the CIMMYT nursery germplasm for stem rust resistance in the Omsk and Novosibirsk regions, together with laboratory testing of infection samples on the international set of wheat differential lines, suggests that a separate "Asian" population of *Puccinia graminis* f. sp. *tritici* exists in Western Siberia and the Altai Territory. Wheat resistance genes *Sr2*, *Sr6Ai#2*, *Sr24*, *Sr25*, *Sr26*, *Sr31*, *Sr39*, *Sr40*, *Sr44*, and *Sr57* are of practical interest for advanced wheat breeding programs for stem rust immunity in Western Siberia. This review provides an analysis of the gene sources that remain effective against the West Siberian population of *P. graminis*, in order to facilitate the initial stage of selection of breeding material to develop a stable genotype by gene pyramiding. The basic requirements for conducting a phytopathological test of breeding material are presented. A list of

molecular markers for the mentioned resistance genes, both widely used in marker-assisted selection and requiring verification, has been compiled.
Key words: bread wheat; stem rust; resistance genes; marker-assisted selection; phytopathological test.

For citation: Kelbin V.N., Skolotneva E.S., Salina E.A. Challenges and prospects for developing genetic resistance in common wheat against stem rust in Western Siberia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):821-828. DOI 10.18699/VJ20.679 (in Russian)

Введение

До конца прошлого века значимость болезни, вызванной биотрофным грибом *P. graminis*, была снижена повсеместно успешными программами селекции на иммунитет. Однако в последнее время для регионов возделывания мягкой пшеницы характерно ухудшение фитопатологической обстановки, связанной со стеблевой ржавчиной: Северной и Южной Америки (Singh R.P. et al., 2016), Восточной Африки (Patpour et al., 2016), Австралии (Addai et al., 2018), Западной Европы (Lewis et al., 2018) и Казахстана (Рсалиев А.С., Рсалиев Ш.С., 2018). Развитию патогена главным образом способствуют благоприятные климатические условия (Shamanin et al., 2013; Morgounov et al., 2014). Причиной массовых эпифитотий пшеницы в Уганде (1998–1999), Кении и Эфиопии (2005–2006), Йемене (2006), Иране (2007) и Пакистане (2009) стало появление и быстрое распространение новой агрессивной расы Ug99 (Prasad et al., 2016). Существует реальная угроза поражения расой и ее модификациями, так называемым семейством рас Ug99 (TTKSK, TTKSF, TTKST, TTTSK, TTKSP, PTKSK, PTKST, TTKSF+, TTKTT, TTKTK и TTHSK), производственных посевов пшеницы в Казахстане и Западной Сибири, на Урале и в других регионах Российской Федерации (Шаманин и др., 2015). В этой связи проводят оценку устойчивости существующих коллекций мягкой яровой пшеницы как к местным популяциям патогенов, так и популяциям патогенов, распространенных в соседних регионах.

Выявление источников устойчивости является повсеместной задачей. В результате скрининга коллекций мягкой яровой пшеницы в Индии (Sharma et al., 2015) и Эфиопии (Soresa, 2018) получены сходные данные: доля устойчивых генотипов к местному возбудителю стеблевой ржавчины, а также Ug99 в коллекциях оказалась минимальной. В России над созданием исходного материала и сортов с устойчивостью к стеблевой ржавчине успешно работают научный коллектив Федерального исследовательского центра «Немчиновка» (Московская область) под руководством д-ра биол. наук И.Ф. Лапочкиной и специалисты Всероссийского научно-исследовательского института защиты растений (Санкт-Петербург, Пушкин). На базе исходного материала, выделенного из коллекции генетических ресурсов растений ВИР и коллекции «Арсенал», созданы линии озимой пшеницы, устойчивые к стеблевой ржавчине в условиях Нечерноземной зоны России. Значительные результаты по генотипированию сортов яровой мягкой пшеницы, а также интрогрессивных линий с генетическим материалом от чужеродных видов (*Aegilops speltoides*, *Agropyron elongatum*, *Aegilops triuncialis*, *Secale cereale*) селекции Федерального аграрного научного центра Юго-Востока получены канд. биол. наук О.А. Барановой. У исследуемых интрогрессивных линий постулированы гены *Sr31*/*Lr26*, *Sr25*/*Lr19*, *Sr28*, *Sr57*/*Lr34* и *Sr38*/*Lr37*. Сочетание генов *Sr31*/*Lr26* и *Sr25*/*Lr19* идентифицировано у 26.3 % линий и сортов, возделываемых на территории Поволжья (Baranova et al., 2019).

Для адекватного генетического контроля заболеваний важен интегральный подход, включающий как данные об источниках устойчивости, так и актуальные сведения о действующих в регионе патогенных популяциях, их расовом составе и динамике генов вирулентности. Целью данного обзора был интегральный анализ фитосанитарной ситуации по стеблевой ржавчине в Западной Сибири в отношении перспективы формирования генетической защиты мягкой пшеницы.

Эффективность известных генов *Sr* в условиях Западно-Сибирского региона

С 2007–2009 гг. значение стеблевой ржавчины в фитопатогенном комплексе пшеницы Западной Сибири возросло (Сочалова, Лихенко, 2015). Путем сравнения расового состава образцов инфекции из Омска и Новосибирска выяснено, что первичной зоной формирования инокулюма *P. graminis* является Омская область (Сколотнева и др., 2020). Комплексное исследование культивируемых на территории Западной Сибири сортов мягкой пшеницы показало, что большинство из них восприимчивы к заболеванию, а остальные защищены небольшим количеством генов устойчивости: *Sr25*, *Sr31*, *Sr36*, *Sr6Ai*, *Sr6Ai#2* (Shamanin et al., 2016; Leonova et al., 2020). При этом оценка коллекции мягкой яровой пшеницы Омского государственного аграрного университета на естественном инфекционном фоне лесостепи Западной Сибири продемонстрировала, что только 10 % сортов коллекции устойчивы к местному патогену. По данным скрининга коллекции к Ug99 в Кении, доля устойчивых к агрессивной расе сортов также не превышает 10 % (Шаманин и др., 2015).

Изучение образцов стеблевой ржавчины Западной Сибири в последние несколько десятилетий выявило изменчивость вирулентности к генам устойчивости пшеницы *Sr6*, *Sr7b*, *Sr8a*, *Sr9e*, *Sr11*, *Sr21*, *Sr30* и *Sr36* (Сколотнева и др., 2020). Результаты полевого скрининга набора сортов с генами *Sr* из коллекции питомников CIMMYT в условиях Омской и Новосибирской областей, а также лабораторного анализа образцов инфекции на международном наборе пшеничных линий-дифференциаторов позволяют предположить, что на территории Западной Сибири и Алтайского края представлена обособленная (так называемая азиатская) популяция *P. graminis* (Shamanin et al., 2020). Она отличается высокой вирулентностью к генам *Sr5*, *Sr9a*, *Sr9b*, *Sr9d*, *Sr9g*, *Sr10*, *Sr17*, *Sr38* и *SrMcN*. Интерес для современных программ опережающей селекции на иммунитет к стеблевой ржавчине в условиях Западной

Сибири представляют гены устойчивости *Sr2*, *Sr6Ai#2*, *Sr24*, *Sr25*, *Sr26*, *Sr31*, *Sr39*, *Sr40*, *Sr44*, *Sr57*, для которых показано преобладание авирулентных клонов в местных субпопуляциях гриба (Shamanin et al., 2016; Skolotneva et al., 2018; Сколотнева и др., 2020).

Присутствие генов *Sr* в мировом селекционном материале

Sr2 является одним из наиболее важных генов в современной селекции на иммунитет к стеблевой ржавчине, так как обеспечивает длительную устойчивость взрослых растений (McIntosh, 1988; Roelfs, 1988; Simmonds, Rajaram, 1988). В 1979 г. описано замедленное течение патогенеза *P. graminis* на растениях, несущих ген *Sr2*, что позволило отнести обеспечиваемую им устойчивость к неспецифическому типу (Hare, McIntosh, 1979). Физиологическим маркером гена *Sr2* является характерное почернение чешуй колоса (pseudo black chaff). Кроме того, ассоциированным признаком, проявляющимся при тепличной температуре выше 22 °С, является хлороз листьев на стадии проростков (Brown, 1993). За исключением Канады ген *Sr2* обеспечивает эффективную устойчивость повсеместно с момента его введения в гексаплоидную пшеницу в 1920-х гг. (McFadden, 1930). Во время эпифитотий в Северной Америке в 1950-х гг. сорта пшеницы Regent, Renown и Redman с геном *Sr2* показали умеренную восприимчивость. Сорта Pavon 76 и Buck Buck с комбинацией генов *Sr2* и *Sr23*, испытанные в условиях Западной Сибири, продемонстрировали устойчивость к локальной популяции стеблевой ржавчины (Шаманин и др., 2015). Тестирование на естественном инфекционном фоне Нечерноземной зоны России новых линий озимой пшеницы, созданных на основе коллекций ВИР и «Арсенал», позволило выявить эффективное сочетание двух и более генов ювенильной устойчивости (*Sr22*, *Sr32*, *Sr39* и *Sr40)* с геном *Sr2* (Лапочкина и др., 2018). В настоящий момент база данных GRIS (http://wheatpedigree.net/) содержит 1762 наименования сортов и линий пшеницы, которые несут ген устойчивости *Sr2.*

Sr6Ai#2 находится в составе группы генов *Lr6Ai#2*/*Sr6Ai#2*/*Pm6Ai#2*, обеспечивающих устойчивость к комплексу листостебельных заболеваний пшеницы, и расположен в хромосоме 6Ai#2, которая интрогрессирована в мягкую пшеницу от *Thinopyrum intermedium*. Цитогенетическое исследование сортов Тулайковская-5, Тулайковская-10 и Тулайковская-100, имеющих многолетнюю историю культивирования в различных регионах России, показало, что хромосома 6Ai#2 сохранила свою целостность в данных сортах (Salina et al., 2015). Среди отечественного материала присутствие гена *Sr6Ai#2* продемонстрировано для линий и сортов саратовской и самарской селекции в сочетании с генами *Sr31* и *Sr25* (Shamanin et al., 2016).

Sr24, ген устойчивости к стеблевой ржавчине, вместе с *Lr24*, геном устойчивости к бурой ржавчине, перенесен в пшеницу от *Ag. elongatum*. Известна спонтанная транслокация (3Ag) в хромосоме 3DL, описанная в сорте Agent (Smith et al., 1968). Получены рекомбинантные линии, в которых удалось нарушить сцепление генов устойчивости и признака красной пигментации зерна (Sears, 1973), что позволило интрогрессировать ген *Sr24* в белозерную пшеницу.

Комплекс генов *Lr24*/*Sr24* обеспечивает эффективную защиту от основных ржавчинных заболеваний пшеницы по всему миру за исключением Южной Африки (Roux, 1985; Pretorius et al., 2010), Индии (Bhardwaj et al., 2010; Manjunatha et al., 2015) и Кении (Jin et al., 2008), где выявлены новые вирулентные расы стеблевой ржавчины. Вирулентные к гену *Lr24* клоны возбудителя бурой ржавчины зарегистрированы в Австралии, Чехии, Иране и США (Park et al., 2002; Kolmer, 2019; Hanzalova et al., 2020; Nemati et al., 2020). В России, в Омской области, недавно выявлены расы *P. graminis*, вирулентные к гену *Sr24* (Shamanin et al., 2020; Skolotneva et al., 2020). В базу данных GRIS загружена информация о 903 сортах и линиях пшеницы, которые несут ген устойчивости *Sr24.*

Sr25, ген расоспецифической устойчивости, перенесен в длинные плечи хромосом 7D и 7A от *Thinopyrum ponticum* с комплексом генов резистентности к бурой ржавчине *Lr19* и геном, контролирующим желтую окраску муки (Friebe et al., 1996; Zhang et al., 2005).

Ген *Sr25* введен в австралийские сорта пшеницы и использован в программе селекции пшеницы CIMMYT, где одним из его основных источников является сорт Wheatear (Bariana et al., 2007). Показана повсеместная эффективность гена *Sr25* по отношению к расам семейства Ug99 (Singh R.P. et al., 2011). Среди отечественного материала присутствие гена *Sr25* отмечено для линий и сортов саратовской, самарской и омской селекции (Shamanin et al., 2016). База данных GRIS в настоящий момент содержит 92 наименования сортов и линий пшеницы, которые несут этот ген.

Sr26, эффективный в отношении семейства рас Ug99, перенесен от *Ag. elongatum* в дистальную область длинного плеча хромосомы 6A (Knott, 1961). Транслокация 6AS.6AL-6Ae#1L заметно влияла на урожайность созданных линий и сортов, потери составляли 9 %. Ген *Sr26* использован в качестве источника устойчивости к стеблевой ржавчине, в основном в Австралии, где был создан сорт Eagle. В настоящее время созданы новые линии с укороченными фрагментами транслокаций, показавшие высокие показатели качества и урожайности (Dundas et al., 2007). Таким образом, исторически сложившаяся низкая частота *Sr26* среди современных сортов и создание донорских линий с короткими чужеродными сегментами делает ген *Sr26* наиболее подходящим для использования в селекционных программах. В базу данных GRIS загружена информация о 61 сорте и линии пшеницы, которые несут ген *Sr26.*

Sr31 унаследован от ржи сорта Petkus в составе транслокации 1BL.1RS вместе с генами, контролирующими устойчивость растений к другим грибным патогенам: бурой ржавчине (*Lr26*), желтой ржавчине (*Yr9*) и мучнистой росе (*Pm8*) (Singh N.K. et al., 1990; McIntosh et al., 1993). Пшенично-ржаная транслокация нашла интенсивное применение в селекционных программах различных регионов мира, в которых родительскими формами были отечественные сорта пшеницы Kavkaz и Aurora, носители этой транслокации (Rabinovich, 1998; Zhou et al., 2003; Schlegel, 2010). За последние сорок лет селекционеры

широко использовали транслокацию 1BL.1RS для улучшения агрономических характеристик мягкой пшеницы, особенно урожайности зерна. Ген *Sr31* присутствует во многих сортах, районируемых в России, Европе, Китае и США, а также в селекционном материале, распространяемом программой CIMMYT, например в сортах Bobwhite и Veery (Carver, Rayburn, 1994; Lelley et al., 2004; Shamanin et al., 2016). В базе данных GRIS содержится 1119 наименований сортов и линий пшеницы, которые несут пшенично-ржаную транслокацию 1BL.1RS.

В большинстве регионов низких широт ген *Sr31* утратил актуальность в связи с распространением из стран Северо-Восточной Африки рас семейства Ug99, вирулентных к этому гену (Singh R.P. et al., 2006), однако *Sr31* остается эффективным на территории России, в том числе в Западно-Сибирском регионе (Волкова и др., 2014; Сколотнева и др., 2020).

Sr39 обеспечивает устойчивость ко всем известным в настоящее время патотипам *P. graminis,* в том числе к семейству рас Ug99 (Mago et al., 2009). Ген перенесен в хромосому 2B сорта Marquis из генома *Ae. speltoides* в составе большой транслокации вместе с геном устойчивости к бурой ржавчине *Lr35* (Kerber, Dyck, 1990). Разные авторы сообщают как об отрицательном, так и о положительном влиянии на хозяйственно важные характеристики транслокационных линий. Например, показано увеличение гигроскопичности муки у южноафриканской линии пшеницы Karee*6/RL6082 с геном *Sr39* (Labuschagne et al., 2002). Разработаны транслокационные линии с уменьшенными чужеродными сегментами, которые обеспечивают групповую устойчивость к ржавчинным болезням за счет генов *Sr39* и *Lr35* (рекомбинант #247) (Mago et al., 2009). Всего шесть линий с геном *Sr39* представлено в базе данных GRIS, из которых четыре имеют канадское происхождение, а линия Line-292 является результатом отечественной селекции.

Sr40 (*SrA*), обеспечивающий высокий уровень ювенильной и возрастной устойчивости к семейству рас Ug99, интрогрессирован в пшеницу от *Triticum timopheevii* ssp. *armeniacum* в составе транслокации T2BL/2G#2S (Friebe et al., 1996; Wu et al., 2009). После скрининга селекционного материала на устойчивость к расам из семейства Ug99 в условиях сильной инфекционной нагрузки кенийских полевых питомников ген *Sr40* рекомендован для использования в коммерческих сортах пшеницы (Jin et al., 2007). В настоящий момент база данных GRIS содержит 9 наименований сортов и линий пшеницы, которые несут ген *Sr40*.

Sr44 (*SrAgi*) перенесен в геном мягкой пшеницы от *Thinopyrum intermedium* в составе транслокации 7Ai#1S (Cauderon et al., 1973; Friebe et al., 1996). Как и гены *Sr25*, *Sr26*, *Sr39* и *Sr40*, ген *Sr44* способен эффективно защищать растения от поражения расами семейства Ug99 (Liu W. et al., 2013). В базу данных GRIS внесена информация о четырех сортах мягкой пшеницы, которые несут ген *Sr44.* В их числе сорт ростовской селекции Донская полукарликовая.

Sr57 (*Lr34*/*Yr18*/*Pm38*/*Bdv1*) – плейотропный ген, обеспечивающий неспецифическую устойчивость к биотрофным патогенам, в том числе стеблевой ржавчине, локализован в хромосоме 7DS (Krattinger et al., 2009; Lagudah et al., 2009; Dakouri et al., 2010). Эффект защитной реакции генотипов с *Sr57* на различных инфекционных фонах описан как возрастная устойчивость (adult plant resistance) (McIntosh et al., 2010). Источником гена *Sr57* являются стародавние итальянские сорта мягкой пшеницы Ardito и Mentana, созданные в 1900-х гг., при этом он сохраняет эффективность в течение столетия (Kolmer et al., 2008). Среди репрезентативной коллекции западноевропейских сортов ген обнаружен только у сорта Кавказ, однако *Sr57* широко распространен среди американских, канадских и австралийских сортов (Kolmer et al., 2008), а также среди украинских сортов озимой мягкой пшеницы (Karelov et al., 2011). Большинство сортов омской и казахской селекции с геном *Sr57* принадлежат к группам Ekada и Fiton соответственно (Shamanin et al., 2016). В последнее время неспецифический ген *Sr57* с успехом применяют для создания генотипов с длительной устойчивостью методом пирамидирования генов. База данных GRIS содержит 2171 наименование сортов и линий пшеницы, которые несут ген *Sr57*.

Методы постулирования генов *Sr*

До разработки первых молекулярных маркеров присутствие в селекционном материале генов устойчивости определяли эмпирически с помощью фитопатологического постулирования в соответствии с законом «ген на ген» – взаимодействия хозяина и патогена (Flor, 1947). Суть постулата заключается в том, что каждому гену устойчивости или восприимчивости растения-хозяина соответствует определенный комплементарный ген вирулентности или авирулентности паразита. К использованию фитопатологического постулирования (фитопатологического тестирования) прибегают до сих пор как к альтернативному подходу, позволяющему верифицировать молекулярные маркеры. Кроме того, данный метод остается единственно возможным в случае идентификации генов, для которых ДНК-маркеры не разработаны. Обязательным условием фитопатологического тестирования селекционного материала является поддержание в лаборатории рабочей коллекции чистых линий гриба, обладающих противоположными аллелями генов *Avr*, в данном случае изолятов *P. graminis*, вирулентных и авирулентных к искомому гену устойчивости *Sr*. При этом исследование проводят на ювенильной стадии растений, оценивая и сравнивая реакции (инфекционные типы) на заражение системами изолятов гриба (McVey, Roelfs, 1975). Проявление высокой восприимчивости (инфекционные типы 3 и 4 по балльной шкале, разработанной E.C. Stakman и коллегами (1962)) у тестируемой линии свидетельствует об отсутствии в генотипе генов устойчивости, к которым изолят *P. graminis* авирулентен. Так, восприимчивый тип реакции на заражение изолятом, авирулентным к *Sr5*, сообщает о том, что тестируемая линия не несет ген *Sr5*. Присутствие гена было бы сопряжено с устойчивостью (инфекционные типы 0, 1 и 2).

Необходимо учитывать подготовительный этап подбора контрольных изолятов или патотипов *P. graminis* с определенной вирулентностью. Их используют, чтобы исключить присутствие гена устойчивости у исследуемого

DNA markers for stem rust resistance genes verified on the worldwide wheat germplasm pool

Gene	DNA marker	Marker type	References
Sr2	csSr2	CAPS	Mago et al., 2011
	wMAS000005	KASP	
	Xgwm533	SSR	Spielmeyer et al., 2003
Sr6Ai#2	TNAC1752	PLUG	Salina et al., 2015
	Xicg6A#2	STS	Salina et al., 2016 (RF patent 2598275)
Sr24	Sr24#12	AFLP	Mago et al., 2005
	Sr24#50		
	Xbarc71	SSR	
Sr25	BF145935	STS	Liu S. et al., 2010
	Gb		Prins et al., 2001
Sr26	Sr26#43	STS	Mago et al., 2005
	BE518379		Liu S. et al., 2010
Sr31	$SCSS30.2_{576}$	SCAR	Das et al., 2006
	$SCSS26.1_{1100}$		
	SCM9	SSR	Weng et al., 2007
Sr39	Sr39#22r	STS	Mago et al., 2009
	BE500705		
	Sr39#50s		
Sr44	Xbe404728	CAPS	Liu W. et al., 2013
	Xbe473884		
Sr57	csLV34	STS	Lagudah et al., 2006
	cssfr1	SSR	Lagudah et al., 2009
	wMAS000003	KASP	

Note. AFLP – amplified fragment length polymorphism; CAPS – cleaved amplified polymorphic sequences; KASP – kompetitive allele specific PCR; PLUG – PCR-based landmark unique gene; SCAR – sequence characterized amplified region; SSR – simple sequence repeats; STS – sequence tagged site.

сорта, если сорт восприимчив хотя бы к одному патотипу, авирулентному к гену; предположить наличие гена на основании совпадения реакции совместимости патотипов с изучаемым сортом и линией, имеющей ген устойчивости. Методы закладки опытов и заражения проростков пшеницы стеблевой ржавчиной, в том числе уход за опытными растениями и оптимальный температурно-временной режим в период экспериментов, подробно описаны в разных публикациях (Jin et al., 2007; Li et al., 2016; Рсалиев А.С., Рсалиев Ш.С., 2018; Flath et al., 2018).

Выявленный эмпирическим путем ген должен быть обязательно подтвержден дополнительными исследованиями, такими как генетический и/или цитогенетический анализы. В настоящее время использование молекулярных маркеров является достойной альтернативой, позволяющей сократить время анализа с помощью оптимизированных протоколов. Список маркеров к генам *Sr2*, *Sr6Ai#2*, *Sr24*, *Sr25*, *Sr26*, *Sr30*, *Sr31*, *Sr39*, *Sr40*, *Sr44*, *Sr57*, сохраняющих эффективность к западносибирской популяции *P. graminis* и верифицированных в отечественных лабораториях, приведен в таблице. Для гена *Sr40* в литературе предложены ДНК-маркеры различного типа, чаще всего микросателлитные (SSR) (Bernardo et al., 2013). Однако SSR-маркеры требуют оценки степени достоверности выявления гена на широком генетическом материале.

Заключение

В обзоре сопоставлены результаты исследований популяции возбудителя стеблевой ржавчины с актуальными данными генов устойчивости мягкой пшеницы, эффективными в условиях Западной Сибири. Интерес для опережающей селекции на иммунитет представляют гены устойчивости *Sr2*, *Sr6Ai#2*, *Sr24*, *Sr25*, *Sr26*, *Sr31*, *Sr39*, *Sr40*, *Sr44*, *Sr57*, для которых показано преобладание авирулентных клонов в местных субпопуляциях гриба. Наиболее популярными при создании пирамидированного генотипа мягкой пшеницы являются гены *Sr2*, *Sr24*, *Sr31*, *Sr57*, о чем свидетельствует высокий удельный вес сортов и селекционных линий, представленных в международных базах данных, таких как GRIS. Весьма перспективным для селекции на иммунитет является интродуцированный ген *Sr6Ai#2*, широко представленный в современном отечественном материале. Однако важно подчеркнуть, что после районирования и интенсивного внедрения в производство сорта быстро теряют устойчивость из-за появления новых вирулентных рас патогена. В большинстве случаев широко распространенные коммерческие сорта мягкой пшеницы оказываются восприимчивыми к ржавчине через 7–10 лет (Коваль и др., 2010). Интегральный подход к отбору селекционного материала, включающий данные о генотипах растения-хозяина и патогена, повышает гарантии длительного иммунитета у нового сорта или селекционной линии.

Список литературы / References

Волкова Г.В., Шумилов Ю.В., Синяк Е.В., Ваганова О.Ф., Данилова А.В. Эффективные гены устойчивости пшеницы и ячменя к возбудителям ржавчины и их идентификация в перспективных сортообразцах. Труды 8-й Международной конференции «Биологическая защита растений – основа стабилизации агроэкосистем». Краснодар, 16–18 сентября 2014. Краснодар. 2014; 346-348.

[Volkova G.V., Shumilov Yu.V., Sinyak Ye.V., Vaganova O.F., Danilova A.V. Effective genes for resistance of wheat and barley to rust pathogens and their identification in promising varieties. Proceedings of the 8th International Conference “Biological plant protection as the basis of agroecosystem stabilization.” Krasnodar, 16-18 September 2014. Krasnodar. 2014;346-348. (in Russian)]

Коваль С.Ф., Шаманин В.П., Коваль А.С. Стратегия и тактика отбора в селекции растений. Омск: Изд-во ФГБОУ ВПО ОмГАУ, 2010.

[Koval S.F., Shamanin V.P., Koval A.S. Selection strategy and tactics in plant breeding. Omsk: Omsk State Agrarian University Publ., 2010. (in Russian)]

Лапочкина И.Ф., Баранова О.А., Гайнуллин Н.Р., Волкова Г.В., Гладкова Е.В., Ковалева Е.О., Осипова А.В. Создание линий озимой пшеницы с несколькими генами устойчивости к *Puccinia graminis* Pers. f. sp. *tritici* для использования в селекционных программах России. *Вавиловский журнал генетики и селекции.* 2018;22(6):676-684. DOI 10.18699/VJ18.410.

[Lapochkina I.F., Baranova O.A., Gainullin N.R., Volkova G.V., Gladkova E.V., Kovaleva E.O., Osipova A.V. The development of winter wheat lines with several genes for resistance to *Puccinia graminis* Pers. f. sp. *tritici* for use in breeding programs in Russia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2018;22(6):676-684. DOI 10.18699/VJ18.410. (in Russian)]

Рсалиев А.С., Рсалиев Ш.С. Основные подходы и достижения в изучении расового состава стеблевой ржавчины пшеницы.

Вавиловский журнал генетики и селекции. 2018;22(8):967-977. DOI 10.18699/VJ18.439.
[Rsaliyev A.S., Rsaliyev Sh.S. Principal approaches and achievements in studying race composition of wheat stem rust. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2018;22(8):967-977. DOI 10.18699/VJ18.439. (in Russian)]

Сколотнева Е.С., Кельбин В.Н., Моргунов А.И., Бойко Н.И., Шаманин В.П., Салина Е.А. Расовый состав новосибирской популяции *Puccinia graminis* f. sp. *tritici*. *Микология и фитопатология.* 2020;54(1):49-58. DOI 10.31857/S0026364820010092.
[Skolotneva E.S., Kelbin V.N., Morgunov A.I., Boyko N.I., Shamanin V.P., Salina E.A. Races composition of the Novosibirsk population of *Puccinia graminis* f. sp. *tritici*. *Mikologiya i Fitopatologiya = Mycology and Phytopathology.* 2020;54(1):49-58. DOI 10.31857/S0026364820010092. (in Russian)]

Салина Е.А., Леонова И.Н., Щербань А.Б., Стасюк А.И. Способ создания линий озимой мягкой пшеницы с комплексной устойчивостью к бурой и стеблевой ржавчине и мучнистой росе. Патент РФ № 2598275. 2016.
[Salina E.A., Leonova I.N., Shcherban A.B., Stasyuk A.I. A method of the development of common wheat winter lines with complex resistance to leaf and stem rusts and to powdery mildew. Russian Federation patent No. 2598275. 2016. (in Russian)]

Сочалова Л.П., Лихенко И.Е. Генетическое разнообразие яровой пшеницы по устойчивости к мигрирующим заболеваниям. Новосибирск: ООО Междуречье, 2015.
[Sochalova L.P., Lichenko I.E. The genetic diversity of spring wheat in resistance to migratory diseases. Novosibirsk: Mezhdurechye Publ., 2015. (in Russian)]

Шаманин В.П., Потоцкая И.В., Кузьмина С.П., Трущенко А.Ю., Чурсин А.С. Селекция яровой мягкой пшеницы на устойчивость к стеблевой ржавчине в Западной Сибири. Омск: Изд-во ФГБОУ ВПО ОмГАУ, 2015.
[Shamanin V.P., Pototskaya I.V., Kuzmina S.P., Trushchenko A.Yu., Chursin A.S. Spring soft wheat breeding for stem rust resistance in Western Siberia. Omsk: Omsk State Agrarian University Publ., 2015. (in Russian)]

Addai D., Hafi A., Randall L., Tennant P., Arthur T., Gomboso J. Potential economic impacts of the wheat stem rust strain Ug99 in Australia. ABARES research report, prepared for the Plant Biosecurity Branch, Department of Agriculture and Water Resources. 2018. DOI 10.13140/RG.2.2.17341.72164.

Baranova O.A., Sibikeev S.N., Druzhin A.E. Molecular identification of the stem rust resistance genes in the introgression lines of spring bread wheat. *Vavilov J. Genet. Breed*. 2019;23(3):296-303. DOI 10.18699/VJ19.494.

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

Bernardo A.N., Bowden R.L., Rouse M.N., Newcomb M.S., Marshall D.S., Bai, G. Validation of molecular markers for new stem rust resistance genes in US hard winter wheat. *Crop Sci*. 2013;53(3): 755-764. DOI 10.2135/cropsci2012.07.0446.

Bhardwaj S.C., Prashar M., Jain S.K., Kumar S., Sharma Y.P. Physiologic specialization of *Puccinia triticina* on wheat (*Triticum species*) in India. *Indian J. Agric. Sci.* 2010;80(9):805.

Brown G.N. A seedling marker for gene *Sr2* in wheat. Proceedings of the 10th Australian plant breeding conference. 1993;2:139-140. DOI 10.1007/s00122-010-1482-7.

Carver B.F., Rayburn A.L. Comparison of related wheat stocks possessing 1B or 1RS. 1BL chromosomes: agronomic performance. *Crop Sci.* 1994;34(6):1505-1510. DOI 10.2135/cropsci1994.0011183X003400060017x.

Cauderon Y., Saigne B., Dauge M. The resistance to wheat rusts of *Agropyron intermedium* and its use in wheat improvement. Proceedings of 4th Int. Wheat Genet. Symp. Columbia, Missouri, USA. 1973;401-407.

Dakouri A., McCallum B.D., Walichnowski A.Z., Cloutier S. Fine-mapping of the leaf rust *Lr34* locus in *Triticum aestivum* (L.) and characterization of large germplasm collections support the ABC transporter as essential for gene function. *Theor. Appl. Genet.* 2010; 121(2):373-384. DOI 10.1007/s00122-010-1316-7.

Das B.K., Saini A., Bhagwat S.G., Jawali N. Development of SCAR markers for identification of stem rust resistance gene *Sr31* in the homozygous or heterozygous condition in bread wheat. *Plant Breed.* 2006;125(6):544-549. DOI 10.1111/j.1439-0523.2006.01282.x.

Dundas I.S., Anugrahwati D.R., Verlin D.C., Park R.F., Bariana H.S., Mago R., Islam A.K.M.R. New sources of rust resistance from alien species: meliorating linked defects and discovery. *Aust. J. Agric. Resour. Econ*. 2007;58(6):545-549. DOI 10.1071/AR07056.

Flath K., Miedaner T., Olivera P.D., Rouse M.N., Yue J. Genes for wheat stem rust resistance postulated in German cultivars and their efficacy in seedling and adult-plant field tests. *Plant Breed.* 2018; 137(3):301-312. DOI 10.1111/pbr.12591.

Flor H.H. Inheritance of reaction to rust in flax. *J. Agric. Res.* 1947; 74(9):41.

Friebe B., Jiang J., Raupp W.J., McIntosh R.A., Gill B.S. Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. *Euphytica.* 1996;91(1):59-87. DOI 10.1007/BF00035277.

Hanzalova A., Dumalasova V., Zelba O. Wheat leaf rust (Puccinia triticina Eriks.) virulence frequency and detection of resistance genes in wheat cultivars registered in the Czech Republic in 2016–2018. *Czech J. Genet. Plant Breed.* 2020;56:87-92. DOI 0.17221/86/2019-CJGPB.

Hare R.A., McIntosh R.A. Genetic and cytogenetic studies of the durable adult plant resistance in Hope and related cultivars to wheat rusts. *Z. Pflanzenzuchtg.* 1979;83:350-67.

Jin Y., Singh R.P., Ward R.W., Wanyera R., Kinyua M., Njau P., Fetch T., Pretorius Z.A., Yahyaoui A. Characterization of seedling infection types and adult plant infection responses of monogenic *Sr* gene lines to race TTKS of *Puccinia graminis* f. sp. *tritici*. *Plant Dis.* 2007;91(9):1096-1099. DOI 10.1094/PDIS-91-9-1096.

Jin Y., Szabo L.J., Pretorius Z.A., Singh R.P., Ward R., Fetch T., Jr. Detection of virulence to resistance gene *Sr24* with in race TTKS of *Puccinia graminis* f. sp. *tritici*. *Plant Dis.* 2008;92:923-926. DOI 10.1094/PDIS-92-6-0923.

Karelov A.V., Pirko Y.V., Kozub N.A., Sozinov I.A., Pirko N.N., Litvinenko N.A., Lyfenko S.F., Koliuchii V.T., Blume Ya.B., Sozinov A.A. Identification of the allelic state of the *Lr34* leaf rust resistance gene in soft winter wheat cultivars developed in Ukraine. *Cytol. Genet.* 2011;45(5):271. DOI 10.3103/S0095452711050069.

Kerber E.R., Dyck P.L. Transfer to hexaploid wheat of linked genes for adult-plant leaf rust and seedling stem rust resistance from an amphiploid of *Aegilops speltoides* × *Triticum monococcum*. *Genome.* 1990;33(4):530-537. DOI 10.1139/g90-079.

Knott D.R. The inheritance of rust resistance. VI. The transfer of stem rust resistance from *Agropyron elongatum* to common wheat. *Can. J. Plant Sci.* 1961;41(1):109-123. DOI 10.4141/cjps61-014.

Kolmer J.A. Virulence of Puccinia triticina, the wheat leaf rust fungus, in the United States in 2017. *Plant Dis.* 2019;103(8):2113-2120. DOI 10.1094/PDIS-09-18-1638-SR.

Kolmer J.A., Singh R.P., Garvin D.F., Viccars L., William H.M., Huerta-Espino J., Ogbonnaya F.C., Raman H., Orford S., Bariana H.S., Lagudah E.S. Analysis of the *Lr34*/*Yr18* rust resistance region in wheat germplasm. *Crop Sci.* 2008;48(5):1841-1852. DOI 10.2135/cropsci2007.08.0474.

Krattinger S.G., Lagudah E.S., Spielmeyer W., Singh R.P., Huerta-Espino J., McFadden H., Bossolini E., Selter L.L., Keller B. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science.* 2009;323(5919):1360-1363. DOI 10.1126/science.1166453.

Labuschagne M.T., Pretorius Z.A., Grobbelaar B. The influence of leaf rust resistance genes *Lr29*, *Lr34*, *Lr35* and *Lr37* on breadmaking quality in wheat. *Euphytica.* 2002;124(1):65-70. DOI 10.1023/A:1015683216948.

Lagudah E.S., Krattinger S.G., Herrera-Foessel S.A., Singh R.P., Huerta-Espino J., Spielmeyer W., Brown-Guedira G., Selter L.L., Keller B. Gene-specific markers for the wheat gene *Lr34*/*Yr18*/*Pm38* which confers resistance to multiple fungal pathogens. *Theor. Appl. Genet.* 2009;119(5):889-898. DOI 10.1007/s00122-009-1097-z.

Lagudah E.S., McFadden H., Singh R.P., Huerta-Espino J., Bariana H.S., Spielmeyer W. Molecular genetic characterization of the *Lr34*/*Yr18* slow rusting resistance gene region in wheat. *Theor. Appl. Genet.* 2006;114(1):21-30. DOI 10.1007/s00122-006-0406-z.

Lelley T., Eder C., Grausgruber H. Influence of 1BL. 1RS wheat-rye chromosome translocation on genotype by environment interaction. *J. Cereal Sci.* 2004;39(3):313-320. DOI 10.1016/j.jcs.2003.11.003.

Leonova I.N., Skolotneva E.S., Orlova E.A., Orlovskaya O.A., Salina E.A. Detection of genomic regions associated with resistance to stem rust in Russian spring wheat varieties and breeding germplasm. *Int. J. Mol. Sci.* 2020;21(13):4706. DOI 10.3390/ijms21134706.

Lewis C.M., Persoons A., Bebber D.P., Kigathi R.N., Maintz J., Findlay K., Bueno-Sancho V., Corredor-Moreno P., Harrington S.A., Kangara N., Berlin A., García R., Germán S.E., Hanzalová A., Hodson D.P., Hovmøller M.S., Huerta-Espino J., Imtiaz M., Mirza J.I., Justesen A.F., Niks R.E., Omrani A., Patpour M., Pretorius Z.A., Roohparvar R., Sela H., Singh R.P., Steffenson B., Visser B., Fenwick P.M., Thomas J., Wulff B.B.H., Saunders D.G.O. Potential for re-emergence of wheat stem rust in the United Kingdom. *Commun. Biol.* 2018;1(1):1-9. DOI 10.1038/s42003-018-0013-y.

Li T.Y., Wu X.X., Xu X.F., Wang W.L., Cao Y.Y. Postulation of seedling stem rust resistance genes of Yunnan wheat cultivars in China. *Plant Prot. Sci.* 2016;52:242-249. DOI 10.17221/137/2015-PPS.

Liu S., Yu L.X., Singh R.P., Jin Y., Sorrells M.E., Anderson J.A. Diagnostic and co-dominant PCR markers for wheat stem rust resistance genes *Sr25* and *Sr26*. *Theor. Appl. Genet.* 2010;120(4):691-697. DOI 10.1007/s00122-009-1186-z.

Liu W., Danilova T.V., Rouse M.N., Bowden R.L., Friebe B., Gill B.S., Pumphrey M.O. Development and characterization of a compensating wheat-*Thinopyrum intermedium* Robertsonian translocation with *Sr44* resistance to stem rust (Ug99). *Theor. Appl. Genet.* 2013;126(5):1167-1177. DOI 10.1007/s00122-013-2044-6.

Mago R., Bariana H.S., Dundas I.S., Spielmeyer W., Lawrence G.J., Pryor A.J., Ellis J.G. Development of PCR markers for the selection of wheat stem rust resistance genes *Sr24* and *Sr26* in diverse wheat germplasm. *Theor. Appl. Genet.* 2005;111(3):496-504. DOI 10.1007/s00122-005-2039-z.

Mago R., Simkova H., Brown-Guedira G., Dreisigacker S., Breen J., Jin Y., Singh R., Appels R., Lagudah E.S., Ellis J., Dolezel J., Spielmeyer W. An accurate DNA marker assay for stem rust resistance gene *Sr2* in wheat. *Theor. Appl. Genet.* 2011;122(4):735-744. DOI 10.1007/s00122-010-1482-7.

Mago R., Zhang P., Bariana H.S., Verlin D.C., Bansal U.K., Ellis J.G., Dundas I.S. Development of wheat lines carrying stem rust resistance gene *Sr39* with reduced *Aegilops speltoides* chromatin and simple PCR markers for marker-assisted selection. *Theor. Appl. Genet.* 2009;119(8):1441-1450. DOI 10.1007/s00122-009-1146-7.

McFadden E.S. A Successful Transfer of Emmer Characters to Vulgare Wheat 1. *Agron. J.* 1930;22(12):1020-1034. DOI 10.2134/agronj1930.00021962002200120005x.

McIntosh R.A. The role of specific genes in breeding for durable stem rust resistance in wheat and triticale. Breeding strategies for resistance to the rusts of wheat. *CIMMYT.* Mexico. 1988;1-9.

McIntosh R.A., Dubcovsky J., Rogers J.W., Morris C.F., Appels R., Xia X.C. Catalogue of gene symbols for wheat: 2011 supplement. *Annual Wheat Newsletter*. 2010;57.

McIntosh R.A., Hart G., Gale M. Catalogue of gene symbols for wheat. Proc. of the 8th Intern. Wheat Genet. Symp. China. 1993;1333-1500.

McVey D.V., Roelfs A.P. Postulation of genes for stem rust resistance in the entries of the fourth international winter wheat performance nursery. *Crop Sci.* 1975;15(3):335-337. DOI 10.2135/cropsci1975.0011183X001500030016x.

Manjunatha C., Aggarwal R., Bhardwaj S.C., Sharma S. Virulence analysis and molecular characterization of *Puccinia triticina* pathotypes causing wheat leaf rust in India. *J. Biotech. Res.* 2015;10:98-107.

Morgounov A., Abugalieva A., Martynov S. Effect of climate change and variety on long-term variation of grain yield and quality in winter wheat in Kazakhstan. *Cereal Res. Commun.* 2014;42(1):163-172.

Nemati Z., Mostowfizadeh-Ghalamfarsa R., Dadkhodaie A., Mehrabi R., Steffenson B.J. Virulence of Leaf Rust Physiological Races in Iran from 2010 to 2017. *Plant Dis.* 2020;104(2):363-372. DOI 10.1094/PDIS-06-19-1340-RE.

Park R.F., Bariana H.S., Wellings C.R., Wallwork H. Detection and occurrence of a new pathotype of *Puccinia triticina* with virulence for *Lr24* in Australia. *Crop Pasture Sci.* 2002;53(9):1069-1076. DOI 10.1071/AR02018.

Patpour M., Hovmøller M.S., Justesen A.F., Newcomb M., Olivera P.D., Jin Y., Szabo L.J., Hodson D., Shahin A.A., Wanyera R., Haburema I., Wobibi S. Emergence of virulence to *SrTmp* in the Ug99 race group of wheat stem rust, *Puccinia graminis* f. sp. *tritici*, in Africa. *Plant Dis.* 2016;100(2):522-522. DOI 10.1094/PDIS-06-15-0668-PDN.

Prasad P., Bhardwaj S.C., Khan H., Gangwar O.P., Kumar S., Singh S.B. Ug99: saga, reality and status. *Curr. Sci.* 2016;110(9):1614-1616.

Pretorius Z.A., Bender C.M., Visser B., Terefe T. First report of a *Puccinia graminis* f. sp. *tritici* race virulent to the *Sr24* and *Sr31* wheat stem rust resistance genes in South Africa. *Plant Dis.* 2010;94:784. DOI 10.1094/PDIS-94-6-0784C.

Prins R., Groenewald J.Z., Marais G.F., Snape J.W., Koebner R.M.D. AFLP and STS tagging of *Lr19*, a gene conferring resistance to leaf rust in wheat. *Theor. Appl. Genet.* 2001;103(4):618-624. DOI 10.1007/PL00002918.

Rabinovich S.V. Importance of wheat-rye translocations for breeding modern cultivar of *Triticum aestivum* L. *Euphytica.* 1998;100(1-3):323-340. DOI 10.1023/A:1018361819215.

Roelfs A.P. Resistance to leaf and stem rusts in wheat. Breeding strategies for resistance to the rusts of wheat. (CIMMYT) 29 Jun – 1 Jul 1987. El Batan, Mexico, 1988.

Roux J.L. First report of a *Puccinia graminis* f. sp. *tritici* race with virulence for *Sr24* in South Africa. *Plant Dis.* 1985;69(11).

Salina E.A., Adonina I.G., Badaeva E.D., Kroupin P.Y., Stasyuk A.I., Leonova I.N., Shishkina A.A., Divashuk M.G., Starikova E.V., Khuat T.M., Syukov V.V., Karlov G.I. A *Thinopyrum intermedium* chromosome in bread wheat cultivars as a source of genes conferring resistance to fungal diseases. *Euphytica.* 2015;204(1):91-101. DOI 10.1007/s10681-014-1344-5.

Schlegel R. Current list of wheats with rye and alien introgression. V. 05. 2010;8:1-14.

Sears E.R. Agropyron-wheat transfers induced by homoeologous pairing. Proceedings. Fourth International Wheat Genetics Symposium. Columbia. MO. Agriculture Experiment Station. College of Agriculture. University of Missouri. Columbia. 1973;191-199.

Shamanin V., Morgounov A., Petukhovskiy S., Likhenko I. The problem of climate warming and the objectives of spring soft wheat breeding in Western Siberia. Intern. Plant Breeding Cong.: Abstract book. 10-14 November 2013. Antalya, Turkey. 2013;217.

Shamanin V., Pototskaya I., Shepelev S., Pozherukova V., Salina E., Skolotneva E., Hodson D., Hovmøller M., Patpour M., Morgounov A. Stem rust in Western Siberia – race composition and effective resistance genes. *Vavilov J. Genet. Breed.* 2020;24(2):131-138 DOI 10.18699/VJ20.608.

Shamanin V., Salina E., Wanyera R., Zelenskiy Y., Olivera P., Morgounov A. Genetic diversity of spring wheat from Kazakhstan and Russia for resistance to stem rust Ug99. *Euphytica.* 2016;212(2):287-296. DOI 10.1007/s10681-016-1769-0.

Sharma A.K., Saharan M.S., Bhardwaj S.C., Prashar M., Chatrath R., Tiwari V., Singh M., Sharma I.N.D.U. Evaluation of wheat (*Triticum aestivum*) germplasm and varieties against stem rust (*Puccinia graminis* f. sp. *tritici*) pathotype Ug99 and its variants. *Indian Phytopathology.* 2015;68(2):134-138.

Simmonds N.W., Rajaram S. (Ed.). Breeding strategies for resistance to the rusts of wheat. CIMMYT, 1988.

Singh N.K., Shepherd K.W., McIntosh R.A. Linkage mapping of genes for resistance to leaf, stem and stripe rusts and ω-secalins on the short arm of rye chromosome 1R. *Theor. Appl. Genet.* 1990;80(5): 609-616. DOI 10.1007/BF00224219.

Singh R.P., Hodson D.P., Jin Y., Huerta-Espino J., Kinyua M., Wanyera R., Njau P., Ward R. Current status, likely migration and strategies to mitigate the threat to wheat production from race Ug99 (TTKS) of stem rust pathogen. CAB reviews: perspectives in agriculture, veterinary science, nutrition and natural resources. 2006;1(54):1-13. DOI 10.1079/PAVSNNR20061054.

Singh R.P., Hodson D.P., Huerta-Espino J., Jin Y., Bhavani S., Njau P., Herrera-Foessel S., Singh P.K., Singh S., Govindan V. The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annu. Rev. Phytopathol*. 2011;49:465-481. DOI 10.1146/annurev-phyto-072910-095423.

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

Skolotneva E.S., Kosman E., Patpour M., Kelbin V.N., Morgounov A., Shamanin V.P., Salina E.A. Virulence Phenotypes of Siberian Wheat Stem Rust Population in 2017-2018. *Front. Agron.* 2020;2:6. DOI 10.3389/fagro.2020.00006.

Skolotneva E.S., Leonova I.N., Bukatich E.Y., Boiko N.I., Piskarev V.V., Salina E.A. Effectiveness of leaf rust resistance genes against *Puccinia triticina* populations in Western Siberia during 2008–2017. *J. Plant Dis. Prot.* 2018;125(6):549-555. DOI 10.1007/s41348-018-0191-3.

Smith E.L., Schlehuber A.M., Young Jr H.C., Edwards L.H. Registration of Agent Wheat (Reg. No. 471). *Crop Sci.* 1968;8(4):511-512. DOI 10.2135/cropsci1968.0011183X000800040039x.

Soresa D.N. Evaluation of bread wheat (*Triticum aestivum* L.) genotypes for resistance against stem rust (*Puccinia graminis* f. sp. *tritici*) diseases at seedling and adult stages. *Afr. J. Agric. Res.* 2018;13(52): 2904-2910. DOI 10.5897/AJAR2018.13244.

Spielmeyer W., Sharp P.J., Lagudah E.S. Identification and validation of markers linked to broad-spectrum stem rust resistance gene *Sr2* in wheat (*Triticum aestivum* L.). *Crop Sci.* 2003;43(1):333-336. DOI 10.2135/cropsci2003.0333.

Stakman E.C., Stewart D.M., Loegering W.Q. Identification of physiologic races of *Puccinia graminis* var. *tritici*. Washington: USDA, 1962.

Weng Y., Azhaguvel P., Devkota R.N., Rudd J.C. PCR-based markers for detection of different sources of 1AL.1RS and 1BL.1RS wheat-rye translocations in wheat background. *Plant Breed*. 2007;126(5): 482-486. DOI 10.1111/j.1439-0523.2007.01331.x.

Wu S., Pumphrey M., Bai G. Molecular mapping of stem-rust-resistance gene *Sr40* in wheat. *Crop Sci.* 2009;49(5):1681-1686. DOI 10.2135/cropsci2008.11.0666.

Zhang W., Lukaszewski A.J., Kolmer J., Soria M.A., Goyal S., Dubcovsky J. Molecular characterization of durum and common wheat recombinant lines carrying leaf rust resistance (*Lr19*) and yellow pigment (*Y*) genes from *Lophopyrum ponticum*. *Theor. Appl. Genet.* 2005;111:573-582. DOI 10.1007/s00122-005-2048-y.

Zhou Y., He Z.H., Liu J.J., Liu L. Distribution of 1BL/1RS translocation in Chinese winter wheat and its effect on noodle quality. Proc. of 10th Intern. Wheat Genet. Symp. Paestum. 2003;3:1419-1421. DOI 10.1016/j.fcr.2011.11.008.

ORCID ID

V.N. Kelbin orcid.org/0000-0002-3455-5704
E.S. Skolotneva orcid.org/0000-0001-8047-5695
E.A. Salina orcid.org/0000-0001-8590-847X

Acknowledgements. This study was supported by the Russian Foundation for Basic Research, project 19-316-90051, and State Budgeted Project 0259-2019-0001-C-01.

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

Received June 23, 2020. Revised October 15, 2020. Accepted October 15, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):829-835
DOI 10.18699/VJ20.680

Selection of an optimal method for screening the collection of narrow-leaved lupine held by the Vavilov Institute for the qualitative and quantitative composition of seed alkaloids

A.V. Kushnareva[1], T.V. Shelenga[1], I.N. Perchuk[1], G.P. Egorova[1], L.L. Malyshev[1], Yu.A. Kerv[1], A.L. Shavarda[1, 2, 3], M.A. Vishnyakova[1] ✉

[1] Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia
[2] V.L. Komarov Botanical Institute of the Russian Academy of Sciences, St. Petersburg, Russia
[3] Saint-Petersburg State University, St. Petersburg, Russia
✉ e-mail: m.vishnyakova.vir@gmail.com

Abstract. Narrow-leaved lupine (*Lupinus angustifolius* L.) is a widely cultivated leguminous forage and green manure crop with a potential for human nutrition. However, the presence of secondary metabolites – alkaloids – in lupine seeds considerably affects the quality of raw produce, reducing its nutritive value; in addition, high concentrations of alkaloids are toxic to humans and animals. Therefore, plant breeders working with lupine need to gain knowledge about the variability of alkaloid content in seeds of different genotypes and search for the sources of their low concentrations in the crop's gene pool. The collection of narrow-leaved lupine genetic resources held by the N.I. Vavilov Institute of Plant Genetic Resources (VIR) offers wide opportunities for such search by means of mass screening. For its part, largescale gene pool screening requires the selection of an optimal technique to measure alkaloid content in seeds, so that it would be easily reproducible and as little labor-, time- and fund-consuming as possible. The results of the search for such method are presented. Qualitative and quantitative indices were compared when target compounds had been extracted with multicomponent mixtures and individual reagents (chloroform, methanol, etc.) and the extracts analyzed using gas chromatography-mass spectrometry. High-performance liquid chromatography combined with mass spectrometry was also employed. Five major alkaloids were found to be present in all types of extracts: lupanine, 13-hydroxylupanine (dominant ones), angustifoline, sparteine, and isolupanine. The fullest extraction of alkaloids was observed when the extractant with an added alkaline agent was used (425 mg/100 g). The lowest level of extraction was registered with chloroform (216 mg/100 g). The significance of the differences was confirmed statistically.
Key words: *Lupinus angustifolius* L.; alkaloids; extraction techniques; lupanine; 13-hydroxylupanine; angustifoline; sparteine; isolupanine.

For citation: Kushnareva A.V., Shelenga T.V., Perchuk I.N., Egorova G.P., Malyshev L.L., Kerv Yu.A., Shavarda A.L., Vishnyakova M.A. Selection of an optimal method for screening the collection of narrow-leaved lupine held by the Vavilov Institute for the qualitative and quantitative composition of seed alkaloids. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):829-835. DOI 10.18699/VJ20.680

Выбор оптимального метода скрининга генофонда люпина узколистного из коллекции ВИР по качественному и количественному составам алкалоидов семян

А.В. Кушнарева[1], Т.В. Шеленга[1], И.Н. Перчук[1], Г.П. Егорова[1], Л.Л. Малышев[1], Ю.А. Керв[1], А.Л. Шавarда[1, 2, 3], М.А. Вишнякова[1] ✉

[1] Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР), Санкт-Петербург, Россия
[2] Ботанический институт им. В.Л. Комарова Российской академии наук, Санкт-Петербург, Россия
[3] Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
✉ e-mail: m.vishnyakova.vir@gmail.com

Аннотация. Люпин узколистный (*Lupinus angustifolius* L.) – широко возделываемая кормовая и сидеральная зернобобовая культура с продовольственным потенциалом. Однако наличие в семенах люпина вторичных метаболитов – алкалоидов – значительно влияет на качество получаемого сырья, снижая его питательную ценность. Кроме того, высокие концентрации алкалоидов токсичны для человека и животных. Поэтому для селекции люпина актуальны сведения об изменчивости содержания алкалоидов в семенах у разных генотипов и поиск в генофонде источников их низкой концентрации. Коллекция генетических ресурсов люпина узколистного Всероссийского института генетических ресурсов растений им. Н.И. Вавилова предоставляет широкие возможности для такого поиска на основе массового скрининга. В свою очередь масштабный скрининг генофонда предполагает выбор оптимального метода оценки содержания алкалоидов в семенах: легко воспроизводимого, наиболее доступного с финансовой точки зрения и наименее трудо- и времязатратного. Представлены результаты поиска

такого метода. Сравнивали качественные и количественные показатели при извлечении целевых веществ многокомпонентными смесями и отдельными реагентами (хлороформом, метанолом) и анализе экстрактов посредством газовой хроматографии, сопряженной с масс-спектрометрией. Также использовали высокоэффективную жидкостную хроматографию с масс-спектрометрией. Обнаружены пять алкалоидов, присутствовавших в экстрактах всех типов: люпанин и 13-гидроксилюпанин (доминирующие), ангустифолин, спартеин и изолюпанин. Наиболее полное извлечение алкалоидов отмечено при использовании экстрагента с добавлением щелочного агента – 425 мг/100 г. Минимальная экстракция зарегистрирована при извлечении хлороформом – 216 мг/100 г. Достоверность отличий подтверждена статистически.
Ключевые слова: *Lupinus angustifolius* L.; алкалоиды; экстракция; люпанин; 13-гидроксилюпанин; ангустифолин; спартеин; изолюпанин.

Introduction

Narrow-leaved lupine (*Lupinus angustifolius* L.) is a high-protein pulse crop, well adapted to comparatively low temperatures, acidic and meager soils. Its gene pool contains plenty early-ripening forms that reach maturity under a sum of active temperatures of 1700 °C, so its effective cultivation may be expanded practically to all regions of the Russian Federation (Artyukhov, 2015; Ageeva et al., 2018). It is chiefly used as a fodder and green manure crop, but there are prospects of its utilization for food production (Krasilnikov, Pankina, 2006; Islam et al., 2011). The cost price of lupine grain production is twice lower than that of soybean (Korol', Lahmotkina, 2018). However, the production of feed and food from most of the existing genetic resources of *Lupinus angustifolius* L. is restricted by the presence of secondary metabolites – alkaloids – in their seed and biomass. The collection of narrow-leaved lupine maintained at N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR) was found to harbor a considerable variability of alkaloid content levels in its seeds (Kurlovich et al., 1995). For many years, since the 1960s, such assessment had been performed on an overwhelming majority of accessions employing the rapid method of in-field differentiation between high-alkaloid and low-alkaloid varieties with Dragendorff's reagent (Ermakova et al., 1987). For part of the accessions, the data were retrieved from published sources and descriptions of varieties submitted by plant breeders. It means that exact quantitative characterization of alkaloid content is absent for most of the accessions preserved in the collection.

Alkaloids, being biologically active compounds, have a negative effect on human and livestock organisms, worsen organoleptic properties of lupine-based products, and reduce the value of its seeds as raw materials for food and feed (Cheeke, Kelly, 1989; Resta et al., 2008). According to the production standards accepted in Russia, the content of alkaloids in lupine seeds earmarked for food and feed purposes should not exceed 0.04 % (40 mg/100 g) of the seed weight (Kuptsov, Takunov, 2006); in some European countries and Australia, no more than 0.02 % (20 mg/100 g) (Frick et al., 2017). This gave a stimulus to one of the prioritized trends in lupine breeding, aimed at the development of low-alkaloid cultivars.

Scientific plant breeding to improve this species, per se, started in the first third of the 20th century, after the development of the first low-alkaloid lupine cultivar Stamm 411 by the German researcher R. Sengbusch in 1928 (Sengbusch, 1931). By now, a substantial number of narrow-leaved lupine cultivars have been released, with the alkaloid content in their seeds not exceeding permissible levels (Kuptsov, Takunov, 2006).

It should be mentioned that lupine alkaloids are widely used in medicine and pharmacology as ganglionic blockers, antiarrhythmic agents, etc. (Hatzold et al., 1983). At present, the antimicrobial effect of lupine alkaloids is actively researched (Erdemoglu et al., 2007), so there is also a need to search for accessions with increased concentrations of total alkaloids or with higher contents of individual compounds within this group (Williams et al., 1984).

Up to 120 alkaloids have been identified in the composition of plant tissues in *Lupinus angustifolius*. Among them, lupanine is the dominating one (50–70 % of the total alkaloids); the shares of 13-hydroxylupanine and angustifoline are ca. 12–30 and 10 %, respectively (Frick et al., 2017). Minor lupine alkaloids, such as pachycarpine, lupinin and matrine, are chemical modifications of the above mentioned alkaloids (Wink, 1987). Qualitative composition of alkaloids does not differ across various forms and varieties of narrow-leaved lupine: all accessions contain the same alkaloids in different proportions (Krasilnikov, Pankina, 2006).

Selection of a method for extraction of these compounds from raw plant material is a paramount stage in the biochemical analysis of alkaloids, determining the accuracy of results during quantitative and qualitative evaluation. The ways to extract alkaloids rapidly from small amounts of material are quite numerous (Adejoke et al., 2019). However, each of them has its individual drawbacks: low performance, very labor-consuming operations, a need for large amounts of toxic and expensive solvents, etc. (Zharylgasina et al., 2014). Besides, most of these techniques would require the use of alcohols, ionic liquids or other solvents non-selective in their hydrophilic/hydrophobic pattern, which complicates purification of alkaloids by removing admixed low-molecular-weight metabolites (mono- and oligosaccharides, alcohols, free amino acids, organic acids, etc.).

Presently, the efficiency and intensity of extraction processes are increased using alternative, resource-saving technologies, specifically microwave/ultrasound treatment (Popova, Potoroko, 2018).

Basic methods to evaluate qualitative and quantitative alkaloid composition are high-performance liquid chromatography (HPLC) and gas chromatography coupled to mass spectrometry (GC-MS). Most researchers favor HPLC-MS. Gas chromatography, however, makes it possible to achieve the same resolution and precision with less time and labor expenditures. Liquid chromatography prefers extracts containing alkaloids in the form of salts, while gas chromatography in the form of bases (Markova et al., 2003).

This paper describes and characterized the alkaloid extraction techniques tested by us on narrow-leaved lupine seeds

to select an optimal one for identification of their qualitative and quantitative composition by means of GC-MS, suitable for mass screening of accessions from the lupine collection held by VIR.

Materials and methods

The research material were seeds of cultivar "Oligarkh", a high-alkaloid green-manure variety of narrow-leaved lupine (k-3814) from the collection of the N.I. Vavilov Institute of Plant Genetic Resources (VIR), grown according to the guidelines developed by VIR (Vishnyakova et al., 2018), on the experimental fields in the town of Pushkin (St. Petersburg) in 2016, and harvested in the phase of full ripeness.

Four ways to obtain alkaloid-containing extracts were tested in the process of research (hereinafter: A, B, C and D procedures). Before extraction, lupine seeds were crushed with a hammermill into thinly dispersed flour. Then alkaloids were extracted from the flour by the following methods.

Extraction procedure A makes it possible to produce alkaloids in the form of bases (Mironenko, 1966). It was divided into two versions, A_1 and A_2:

A_1 – 2 g of lupine seed flour was mixed with diethyl ether, supplemented with chloroform and a 5 % water solution of NaOH in ratio 10:5:1;

A_2 – 2 g of flour was mixed with ethyl acetate, supplemented with a concentrated ammonia solution in ratio 8:1.

Both solutions, A_1 and A_2, were treated according to the same pattern: they were left for 16–18 hours at 4 °C; after that, they were filtered through Whatman ash-free filter paper (0.45 μm, Merck, Germany) and Millipore polytetrafluoroethylene syringe filters (diameter 25 mm, pore size 22 μm, Ireland) to remove solid residues of plant material. As a result, samples A_1 and A_2 were produced.

Procedure B was employed to obtain acid salts of alkaloids: 10 mL of sample A_1 was mixed with a 1 % water solution of hydrochloric acid in ratio 1:1. After that, the water layer containing alkaloid salts was isolated (sample B) (Mironenko, 1966).

The C and D techniques make it possible to extract native forms of alkaloids as salts of organic acids using certain extractants (methanol or chloroform).

Procedure C is an intensified technique to obtain native forms of alkaloids, based on ultrasound application. Lupine seed flour (0.3 g) was mixed with methanol (1 mL). The resulting mixture was treated for 30 min with ultrasound in an Elmasonic S30H bath (Germany), ultrasonic wave length 220 nm, and later infused for 8 hours at +4–6 °C (sample C).

Procedure D: lupine seed flour (250 mg) was mixed with chloroform (1 mL). The mixture was infused for 16–18 hours at +4–6 °C (sample D) (Zharylgasina et al., 2014).

After infusion, samples C and D were centrifuged for 15 min on an Eppendorf 5415C Centrifuge (Germany) at 8000 rpm. The supernatant was collected for further analysis. Then, 100 μl portions of samples A_1, B, C and D were dried according to the same pattern using the vacuum concentrator Savant™ SpeedVac™ (USA).

The resulting solid residues of samples A_1, B, C and D were silylated by adding 20 μl of N,O-Bis(trimethylsilyl)trifluoroacetamide. The mixture of trimethylsilyl ethers was separated on an Agilent HP-5MS capillary column (5 % phenyl, 95 % methylpolysiloxane; 30.0 m, 250.00 μm, 0.25 μm), at the inert gas speed of 1.5 mL/min, employing a gas chromatograph (Agilent 6850 Network GC System) with a quadrupole mass-selective detector (Agilent 5975B VL MSD), produced by Agilent Technologies, Inc. (USA). Heating program: from +170 to +320 °C, heating rate: 4 °C/min. Mass spectrometer detector temperature: +250 °C, injector temperature: +300 °C, sample size: 1.2 μl. Sample A_2 was analyzed with GC-MS without additional conversion (without silylation).

Acid salts of alkaloids (sample B) were separated using a liquid chromatograph (Agilent Technologies Series 1200, USA) on an Agilent Zorbax SB-C18 column (150 mm; 3 mm; 1.8 μm) at a gradient elution mode from 1.000 to 0.425 deionized water/acetonitrile. Elution speed: 50.00 μl/min. Sample size: 0.5 μl.

The following commercial standards were used to identify alkaloids: 900263 for sparteine (Sigma-Aldrich, USA); ALB-RS-1465 for lupanine (ALB Technology Limited, USA); sc-481026 for angustifoline, and sc-490845 for 13-hydroxylupanine (Santa Cruz Biotechnology, USA). As an internal standard for quantitative calculation of alkaloid content, the commercial standard for caffeine, 142833 (PanReac AppliChem, ITW, USA) was used, in the 1 μg/μl concentration.

Statistical analysis. The results obtained were processed using the AMDIS and UniChrom software. Statistical data processing was made using the Statistica 7.0 software package; it included ANOVA and factor analysis of correlation matrix.

Results

Five alkaloids typical for lupine seeds were identified in the extracts produced by all tested extraction procedures (A_1, A_2, B, C and D) where GC-MS was used: lupanine, 13-hydroxylupanine, sparteine, angustifoline and isolupanine (Fig. 1, *a–e*). To control the fullness of alkaloid extraction from the plant material under different techniques of quantitative sample preparation with GC-MS, acid salts of alkaloids (sample B) were analyzed by means of HPLC, because this type of chromatography is most frequently used while studying plant alkaloids (see Fig. 1, *f*). In this case, the same set of alkaloids was identified as with GC-MS. The analysis of acid salts (sample B) with HPLC showed that the amounts of lupanine, 13-hydroxylupanine, angustifoline, sparteine and isolupanine were 259.63; 46.51; 56.00; 20.87 and 2.31 mg/100 g (67.38; 12.07; 14.53; 5.42 and 0.6 % of the total alkaloids), respectively. Using GC-MS to analyze sample B demonstrated practically the same results (257.93; 46.23; 54.92; 20.12 and 2.07 mg/100 g, or 67.65; 12.13; 14.40; 5.28 and 0.54 %, respectively).

Lupanine was the dominating alkaloid in all samples analyzed with GC-MS: its content varied from 317.86 (sample A_1) to 196.43 mg/100 g (sample D), or from 77.60 to 90.63 % of the sum of all alkaloids identified in the respective samples. The next in amount in the alkaloid composition of lupine seeds were 13-hydroxylupanine and angustifoline. The content of 13-hydroxylupanine varied from 9.67 (sample D) to 58.42 mg/100 g (sample A_2), which corresponded to 4.46 and 13.72 %. The content of angustifoline ranged from 54.92 (sample B) to 9.33 mg/100 g (sample D), and from 14.40 to 4.30 % (sample D). The levels of sparteine were significantly lower: from 20.12 (sample B) to 1/10 mg/100 g (sample D),

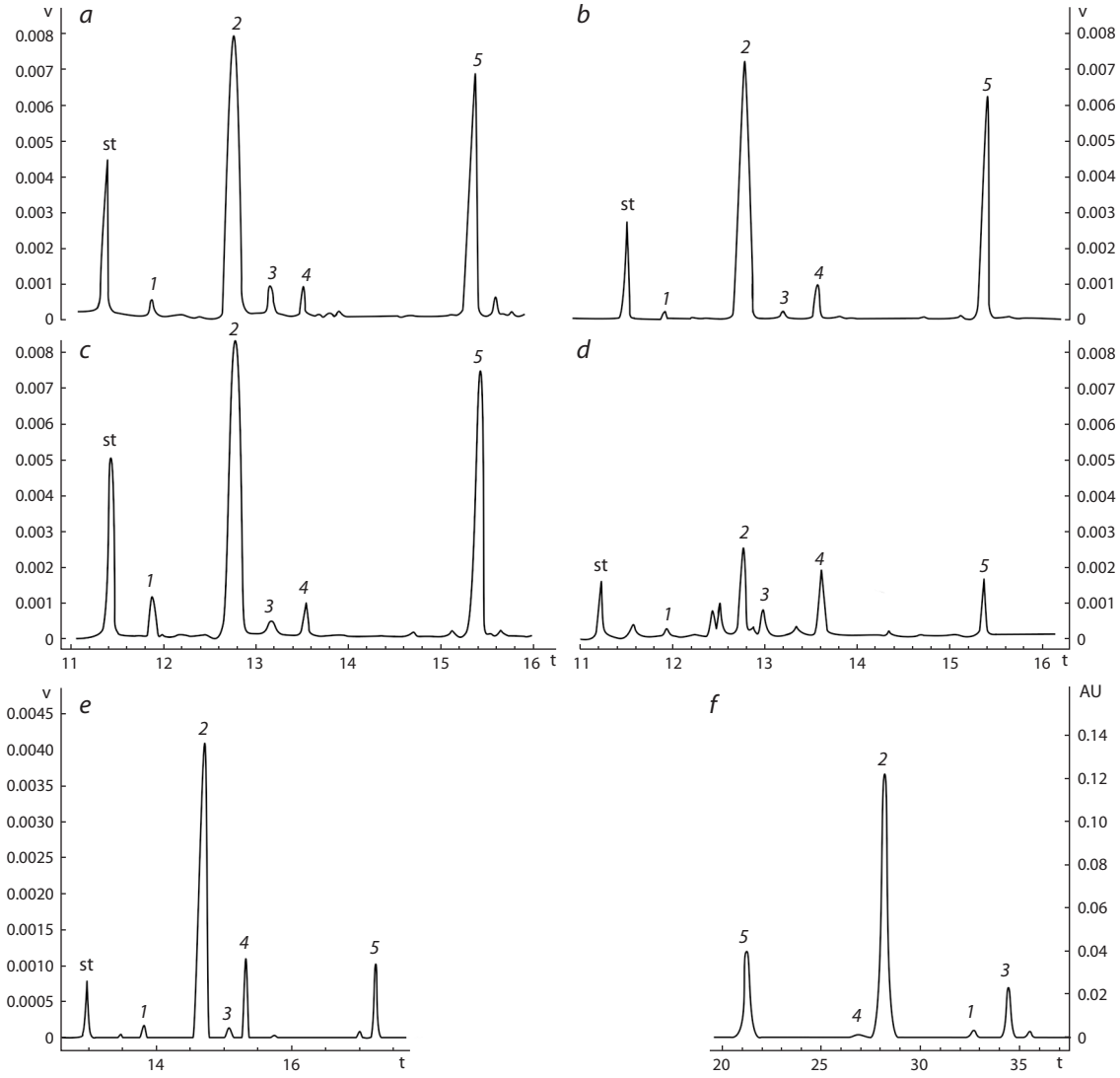


Fig. 1. Chromatograms of the samples: *a* – A_1; *b* – A_2; *c* – C; *d* – D; *e* – B, obtained by GC-MS; *f* – chromatogram of acid salts of the alkaloids obtained by HPLC-MS. Alkaloids: st – standard; *1* – sparteine; *2* – lupanine; *3* – angustifoline; *4* – isolupanine; *5* – 13-hydroxylupanine.

corresponding to 5.28 and 0.51 %, respectively. The minimum content was recorded for isolupanine: from 2.07 (sample B) to 0.12 mg/100 g (sample C), or 0.54 and 0.05 % (Table).

The results confirmed that the qualitative and quantitative composition as well as the ratios of alkaloids in the samples practically coincided in both, GC-MS and HPLC, versions (see Fig. 1, *c*, *f*). Thus, alkaloid extraction from lupine seeds had the best outcome with the A_2 and B techniques, while the C and D extraction procedures mostly isolated the dominant alkaloid, lupanine.

The least labor- and time-consuming alkaloid extraction procedure for lupine seeds was the C technique (only nine hours, requiring direct involvement of a researcher). Time expenditures for the A_1, A_2, B and D procedures were almost the same (from 19 up to 22 hours).

The most laborious way of sample preparation (the largest number of manipulations) was producing acid salts of alkaloids (the B technique).

The analysis of variance showed that the alkaloid extraction techniques tested by us on narrow-leaved lupine seeds had significant differences among them, both in the total alkaloid content and in concentrations of individual compounds (Fig. 2). For example, the A procedure in both versions (A_1 and A_2) proved the most efficient for extracting dominating alkaloids, i. e., lupanine and 13-hydroxylupanine, and the sum of alkaloids (see Fig. 2, *a*). To isolate sparteine and angustifoline, the B technique was the best (see Fig. 2, *c, b*). The amount of isolupanine extracted with all sample preparation techniques did not exceed 2.2 mg/100 g and was the lowest. In the case of isolupanine content, no significant differences were observed among the tested techniques (see Fig. 2, *c*). The use of the C and D procedures demonstrated the lowest values of both the sum of extracted alkaloids and their individual fractions, so they proved to be the least effective (Fig. 2, *a–c*).

Analyzing the system of correlations between the content and the percentage of the identified alkaloids under different extraction techniques helped to identify two factors, embracing 95.9 % of the variation in the set of the data obtained.

Factor 1 (58.5 % of variability) was associated with the variations in the content and percentage of sparteine, angus-

The content of main alkaloids in the seeds of the lupine cultivar "Oligarkh" under various extraction options, mg/100 g, and % of the total amount of identified alkaloids

Sample	Lupanine		13-hydroxylupanine		Sparteine		Angustifoline		Isolupanine		Total alkaloids
	mg/100 g	%	mg/100 g	%	mg/100 g	%	mg/100 g	%	mg/100 g	%	mg/100 g
A_1	317.9±6.1	77.6	48.3±6.2	11.8	11.8±6.9	2.8	31.0±3.1	7.6	0.62±0.1	0.2	409.6±12.3
A_2	306.7±3.3	72.0	58.4±5.4	13.7	18.6±5.1	4.3	40.6±3.9	9.5	1.59±0.1	0.4	425.9±11.1
B	257.9±4.8	67.7	46.2±7.1	12.1	20.1±5.7	5.2	54.9±6.2	14.4	2.07±0.1	0.5	381.3±13.2
HPLC	259.6±2.2	67.3	46.5±3.2	12.1	20.9±5.3	5.4	56.0±4.3	14.5	2.31±0.1	0.6	385.4±10.3
C	219.2±4.1	87.0	16.2±6.3	6.4	2.2±1.1	0.9	14.2±6.0	5.6	0.12±0.1	0.1	251.9±12.4
D	196.4±6.1	90.6	9.7±4.8	4.5	1.1±1.0	0.5	9.3±9.2	4.3	0.21±0.2	0.1	216.7±14.8

Note. HPLC – high-performance liquid chromatography.

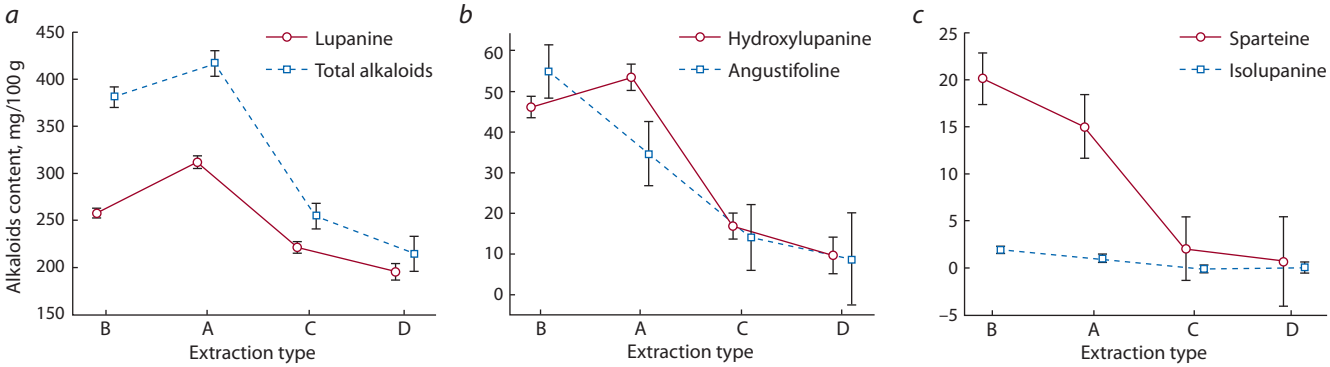


Fig. 2. ANOVA of the content (mg/100 g) of the total alkaloids and lupanine (*a*), 13-hydroxylupanine and angustifoline (*b*), sparteine and isolupanine (*c*) in the seeds of the lupine cultivar "Oligarkh" measured under various extraction techniques (A, B, C and D).

tifoline and isolupanine, while factor 2 (37.4 %) with that of lupanine and 13-hydroxylupanine. The A and B extraction techniques were found to differ considerably in the factor structure of the variables from the C and D ones. The C and D procedures formed a separate group, because the results obtained with them had no statistically significant differences between them (Fig. 3).

Discussion

Lupine alkaloids are attributed to the quinolizidinic group and contain one (lupanine and 13-hydroxylupanine) or two (sparteine) condensed quinolizidine nuclei. For all lupine alkaloids, their molar mass does not exceed 300 g/mol. A peculiar structural feature of an alkaloid molecule is the presence of an undivided pair of electrons in a nitrogen atom (Orekhov, 1955; Roberts, Wink, 2013), which explains their properties that determine the specific nature of the techniques of their extraction from plant tissues. Alkaloids are present in plants mostly as salts, because they interact with organic acids contained in plant cells, which should be taken into account while selecting an extraction method (Mironenko, 1966). In the form of bases, alkaloids are readily soluble in chloroform, ether or ethyl acetate, but practically insoluble in water; on the contrary, in the form of salts, they are water-soluble, but insoluble in organic solvents. A water solution of NaOH or, less frequently, ammonia is used as an alkaline agent for extraction of alkaloids as bases. Ultrasound is used to improve the extraction kinetics and increase the outcome of the target product (Vilkhu et al., 2008; Popova, Potoroko, 2018).

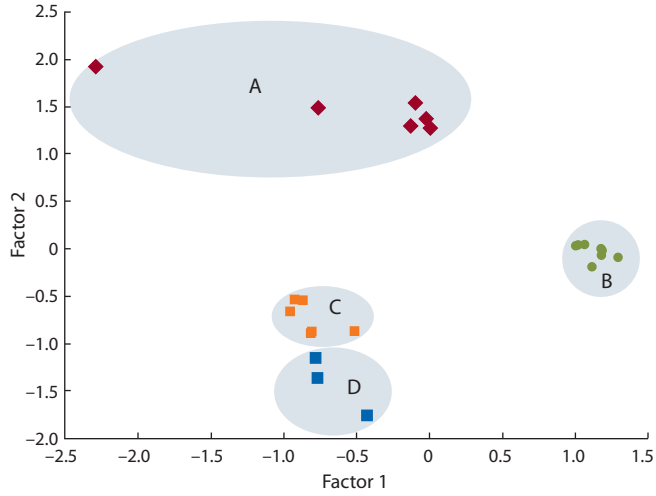


Fig. 3. Results of a factor analysis of alkaloid composition and content in the seeds of the lupine cultivar "Oligarkh" measured under various extraction techniques (A, B, C and D).

Considering these specific features, we applied different extractants to retrieve alkaloids from narrow-leaved lupine seeds: both hydrophobic (chloroform, ethyl acetate, and diethyl ether) and hydrophilic ones (methanol, and water solution of hydrochloric acid). All alkaloids typical for the studied species were identified in all extracts produced under all versions of sample preparation (Krasilnikova, Pankina, 2006; Erdemoglu et al., 2007). With all applied extraction techniques, lupanine

was the dominant alkaloid, followed in descending order by 13-hydroxylupanine, sparteine, angustifoline and isolupanine, which is also in line with the published data (Smirnova, 1938; Pankina, Borisova, 2015). However, the rates of alkaloid extraction under different sample preparation procedures were different. The maximum amounts of alkaloids were registered for the extracts obtained with the A technique (versions A_1 and A_2). Another advantage of this technique was the least amounts of extracted accompanying compounds, compared with other procedures (Zharylgasina et al., 2014). Extraction with chloroform or methanol (C and D techniques), enabling a researcher to isolate alkaloids in the form of free bases, was the cheapest and the least labor-consuming in our research; its application led mostly to the extraction of lupanine. Since long ago, lupanine and sparteine have been numbered among the most toxic alkaloids (Couch, 1926).

In our research, under all sample preparation techniques, lupanine was present in the extracts in maximum amounts – from 67.4 to 96.0 %. At the same time, the amount of sparteine was many times (from 20 to 200, or more) lower, depending on the way of extraction. Therefore, we assumed that the least expensive sample preparation procedures, when mostly lupanine was extracted, might be used for a screening assessment of "alkaloidization" in large numbers of accessions, i. e., a collection of narrow-leaved lupine genetic resources. Concentrations of lupanine in seeds, measured by such techniques, would help to understand whether such alkaloid content should be deemed fit to regard the lupine variety in question suitable for food or feed purposes. It was observed in our research that simplifying the composition of extractants to a single component, comparted with the A procedure where they were multicomponent, led to an almost twofold reduction in the extraction of the total alkaloids (see Fig. 2, *a*). The most laborious and lengthy technique, when alkaloids were extracted in the form of salts, was the B procedure. When applied, this technique led to lesser extraction of the dominating alkaloids in narrow-leaved lupine (lupanine and 13-hydroxylupanine) than with the A technique, but sparteine and angustifoline were isolated to a greater extent than with the other methods. The total sum of alkaloids was lower than with the extraction by the A technique, but higher than with the C and D procedures, which proved less labor-consuming and the most cost-effective as far as financial aspects are concerned.

Conclusion

A comparison among all tested techniques of alkaloid extraction from narrow-leaved lupine seeds has shown that the A procedure (versions A_1 and A_2) seems the most effective for quantitative assessment. This technique, involving multicomponent extractants containing an alkaline agent, has sufficient capacity and good reproducibility, which gives enough reason to regard it as reliable for evaluation of germplasm collection holdings and for most precise measurement of total alkaloid concentrations and amounts of individual alkaloids. However, as an alternative way to perform mass screening of large numbers of accessions, it is possible to employ the sample preparation procedure where only one solvent is used (methanol or chloroform). It will enable a researcher to measure the amount of lupanine, the dominant alkaloid in narrow-leaved lupine seeds, identify a permissible alkaloid content determining food or feed purposes of an accession, and categorize accessions from the collection according to their alkaloid content (high-, medium- and low-alkaloid). Despite the fact that most researchers use HPLC-MS to identify alkaloids, our study has shown that gas chromatography may be used with the same resolution and accuracy, but requires less time and labor inputs.

References

Adejoke H.T., Louis H., Amusan O.O., Apebende G. A Review on Classes, Extraction, Purification and Pharmaceutical Importance of Plants Alkaloids. *J. Med. Chem. Sci.* 2019;2(4):130-139. DOI 0.26655/JMCHEMSCI.2019.8.2.

Ageeva P.A., Pochutina N.A., Pigareva S.A. Comparative characteristics of grain and green mass quality of fodder narrow-leafed lupin varieties. *Adaptivnoe Kormoproizvodstvo = Adaptive Fodder Production*. 2018;1:42-48. (in Russian)

Artyukhov A.I. Lupin species adaptation to agrolandscape of Russia. *Zernobobovye i Krupyanye Kultury = Legumes and Groat crops*. 2015;1(13):60-67. (in Russian)

Cheeke P.R., Kelly J.D. Metabolism, toxicity and nutritional implications of quinolizidine (lupin) alkaloids. In: Recent advances of research in antinutritional factors in legume seeds. Material of 1st International Workshop on Antinutritional Factors (ANF) in Legume Seeds. Netherlands, 1989;189-201.

Couch J. Relative Toxicity of the Lupine Alkaloids. *J. Agric. Res*. 1926; XXXII:51-67.

Erdemoglu N., Ozkan S., Tosun F. Alkaloid profile and antimicrobial activity of *Lupinus angustifolius* L. alkaloid extract. *Phytochem Rev.* 2007;6(1):197-201. DOI 10.1007/s11101-006-9055-8.

Ermakov A.I., Arasimovich V.V., Yarosh N.P., Peruanski Y.V., Lukovnikova G.A., Ikonnikova M.I. Methods of biochemical study of plants. Leningrad, 1987. (in Russian)

Frick K.M., Kamphuis L.G., Siddique K.H., Singh K.B., Foley R.C. Quinolizidine alkaloid biosynthesis in lupins and prospects for grain quality improvement. *Front. Plant Sci.* 2017;8(87):1-12. DOI 10.3389/fpls.2017.00087.

Hatzold T., Elmadfa I., Gross R., Wink M., Hartmann T., Witte L. Quinolizidine alkaloids in seeds of *Lupinus mutabilis*. *J. Agric. Food Chemistry.* 1983;31(5):934-938. DOI 10.1021/jf00119a003.

Islam S., Ma W., Ma J., Buirchell B.J., Appels R., Yan G. Diversity of seed protein among the Australian narrow-leafed lupin (*Lupinus angustifolius* L.) cultivars. *Crop Past. Sci.* 2011;62(9):765-775. DOI 10.1071/CP11046.

Korol' V.F., Lahmotkina G.N. The use of ultrasound in the isolation of antialimentary substances from lupine grain. *Yuzhno-Sibirskiy Nauchnyy Vestnik = South Siberian Scientific Bulletin.* 2018;1(21): 27-34. (in Russian)

Krasilnikov V.N., Pankina I.A. Study of the chemical composition and technological properties of narrow-leafed lupine seeds with the aim of creating combined food products. Proceedings of the international conference "The problem of vegetable protein deficiency and ways to overcome it." Minsk, 2006;119-122. (in Russian)

Kuptsov N.S., Takunov I.P. Lupin (Genetic, selection, heterogeneous sowing). Bryansk, 2006. (in Russian)

Kurlovich B.S., Repiev S.I., Shchelko L.G., Budanova V.I., Petrova M.V., Buravtseva T.V., Stankevich A.K., Leokene L.V., Benken I.I., Rybnikova V.A., Kartuzova L.T., Zolotov S.V., Alexandrova T.G., Debely G.A., Taranuho G.I., Teplyakova T.E., Malysh L.K. The gene pool and breeding of legumes (lupins, vetch, soy, beans). St. Petersburg, 1995;9-116. (in Russian)

Markova O.M., Karpenko V.A., Sayshkina A.S., Likhota T.T. Use of physicochemical methods in the analysis of pharmaceuticals of a vegetative origin. *Bulletin of Voronezh State University. Series: Chemistry. Biology. Pharmacy*. 2003;1:99-100. (in Russian)

Mironenko A.V. Methods for the determination of alkaloids. Minsk, 1966;182 p.

Orekhov A.P. Chemistry of alkaloids. Moscow, 1955. (in Russian)

Pankina I.A., Borisova L.M. Alkaloidness investigation of Lupin seeds. *Scientific journal ITMO University. Series* "*Processes and Food Production Equipment*." 2015;4:80-87. (in Russian)

Popova N.V., Potoroko I.Yu. Increase of efficiency of biologically active substance extraction from vegetable raw material by ultrasonic treatment. *Bulletin of the South Ural State University. Series: Food and Biotechnology*. 2018;6(1):1-9. (in Russian)

Resta D., Boschin G., D'agostina A., Arnoldi A. Evaluation of total quinolizidine alkaloids content in lupin flours, lupin-based ingredients, and foods. *Mol. Nutr. Food Res.* 2008;52(4):490-495. DOI 10.1002/mnfr.200700206.

Roberts M.F., Wink M. Alkaloids: Biochemistry, Ecology, and Medicinal Applications. 2013.

Sengbusch R. Bitterstoffarme Lupinen. Zuchter. 1931;H.4:93-109.

Smirnova M.I. Biochemistry of lupine. In: Biochemistry of cultivated plants. Moscow-Leningrad: Selhozgis Publ., 1938;2:270-327. (in Russian)

Vilkhu K., Mawson R., Simons L., Bates D. Applications and opportunities for ultrasound assisted extraction in the food industry. A review. *Innov. Food Sci. Emerg. Technol.* 2008;9:161-169. DOI 10.1016/j.ifset.2007.04.014.

Vishnyakova M.A., Seferova I.V., Buravtseva T.V., Burlyaeva M.O., Semenova E.V., Filipenko G.I., Alexandrova T.G., Egorova G.P., Yankov I.I., Bulyntsev S.V., Gerasimova T.V., Drugova E.V. *Guidelines* "VIR Global Collection of Grain Legumes Crop Resources: Replenishment, Preservation and Study." St. Petersburg, 2018. (in Russian)

Williams W., Harrison J.E.M., Jayasekera S. Genetical control of alkaloid production in *Lupinus mutabilis* and the effect of a mutant allele Mutal isolated following chemical mutagenesis. *Euphytica*. 1984; 33(3):811-817. DOI 10.1007/BF00021907.

Wink M. Quinolizidine alkaloids: biochemistry, metabolism and function in plants and cell suspension cultures. *Planta Med.* 1987;53(6): 509-514. DOI 10.1055/s-2006-962797.

Zharylgasina G.T., Nurmaganbetov Zh.S., Turmukhambetov A.Zh., Adekenov S.M. Modern methods for extraction of alkaloids from plant materials. *Sci. Pract. J.* 2014;3-4:105-122. (in Russian)

ORCID ID

A.V. Kushnareva orcid.org/0000-0002-5709-7961
T.V. Shelenga orcid.org/0000-0003-3992-5353
I.N. Perchuk orcid.org/0000-0001-6568-5248
G.P. Egorova orcid.org/0000-0002-8645-3072
L.L. Malyshev orcid.org/0000-0002-8595-1336
Yu.A. Kerv orcid.org/0000-0002-3728-6968
A.L. Shavarda orcid.org/0000-0003-1778-2814
M.A. Vishnyakova orcid.org/0000-0003-2808-7745

Acknowledgements. The work was supported by the Russian Foundation for Basic Research, project No. 20-016-00072-A, and budgetary project No. 0662-2019-0002.

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

Received May 15, 2020. Revised October 23, 2020. Accepted October 26, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):836-843
DOI 10.18699/VJ20.681

Candidate genes for productivity identified by genome-wide association study with indicators of class in the Russian meat merino sheep breed

A.Y. Krivoruchko, O.A. Yatsyk ✉, E.Y. Safaryan

All-Russian Research Institute of Sheep and Goat Breeding – Branch of the North Caucasus Federal Agricultural Research Center, Stavropol, Russia
✉ e-mail: malteze@mail.ru

Abstract. Genome-wide association studies allow identification of loci and polymorphisms associated with the formation of relevant phenotypes. When conducting a full genome analysis of sheep, particularly promising is the study of individuals with outstanding productivity indicators – exhibition animals, representatives of the super-elite class. The aim of this study was to identify new candidate genes for economically valuable traits based on the search for single nucleotide polymorphisms (SNPs) associated with belonging to different evaluation classes in rams of the Russian meat merino breed. Animal genotyping was performed using Ovine Infinium HD BeadChip 600K DNA, association search was performed using PLINK v. 1.07 software. Highly reliable associations were found between animals belonging to different evaluation classes and the frequency of occurrence of individual SNPs on chromosomes 2, 6, 10, 13, and 20. Most of the substitutions with high association reliability are concentrated on chromosome 10 in the region 10: 30859297–31873769. To search for candidate genes, 15 polymorphisms with the highest association reliability were selected ($-\log_{10}(p) > 9$). Determining the location of the analyzed SNPs relative to the latest annotation Oar_rambouillet_v1.0 allowed to identify 11 candidate genes presumably associated with the formation of a complex of phenotypic traits of animals in the exhibition group: *RXFP2*, *ALOX5AP*, *MEDAG*, *OPN5*, *PRDM5*, *PTPRT*, *TRNAS-GGA*, *EEF1A1*, *FRY*, *ZBTB21-like*, and *B3GLCT-like*. The listed genes encode proteins involved in the control of the cell cycle and DNA replication, regulation of cell proliferation and apoptosis, lipid and carbohydrate metabolism, the development of the inflammatory process and the work of circadian rhythms. Thus, the candidate genes under consideration can influence the formation of exterior features and productive qualities of sheep. However, further research is needed to confirm the influence of genes and determine the exact mechanisms for implementing this influence on the phenotype.

Key words: sheep; SNP; genome-wide association study; GWAS; candidate gene; Russian meat merino.

For citation: Krivoruchko A.Y., Yatsyk O.A., Safaryan E.Y. Candidate genes for productivity identified by genome-wide association study with indicators of class in the Russian meat merino sheep breed. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):836-843. DOI 10.18699/VJ20.681

Гены-кандидаты продуктивности, выявленные при полногеномном поиске ассоциаций с показателями классности у овец породы российский мясной меринос

А.Ю. Криворучко, О.А. Яцык ✉, Е.Ю. Сафарян

Всероссийский научно-исследовательский институт овцеводства и козоводства – филиал Северо-Кавказского федерального научного аграрного центра, Ставрополь, Россия
✉ e-mail: malteze@mail.ru

Аннотация. Полногеномный поиск ассоциаций позволяет идентифицировать локусы и отдельные полиморфизмы, связанные с формированием интересующих фенотипов. При полногеномном анализе у овец особо перспективным представляется изучение особей, отличающихся выдающимися показателями продуктивности – выставочных животных, представителей класса «суперэлита». Целью настоящего исследования явилось выявление новых генов-кандидатов хозяйственно ценных признаков на основе поиска однонуклеотидных полиморфизмов, ассоциированных с принадлежностью к различным бонитировочным классам, у баранов породы российский мясной меринос. Генотипирование животных выполнено с использованием ДНК-биочипов Ovine Infinium HD BeadChip (600K), поиск ассоциаций – с использованием программного обеспечения PLINK v. 1.07. Выявлены высокодостоверные ассоциации между принадлежностью животных к различным бонитировочным классам и частотой встречаемости отдельных однонуклеотидных полиморфизмов на хромосомах 2, 6, 10, 13 и 20. Большая часть замен с высокой достоверностью ассоциаций сконцентрирована на хромосоме 10 в области 30859297–31873769. Для поиска генов-кандидатов отобрано 15 полиморфизмов с

наибольшей достоверностью ассоциаций ($-\log_{10}(p) > 9$). Определение местоположения анализируемых однонуклеотидных полиморфизмов относительно новейшей аннотации Oar_rambouillet_v1.0 позволило выявить 11 генов-кандидатов, предположительно, связанных с формированием комплекса фенотипических признаков животных выставочной группы: *RXFP2*, *ALOX5AP*, *MEDAG*, *OPN5*, *PRDM5*, *PTPRT*, *TRNAS-GGA*, *EEF1A1*, *FRY*, *ZBTB21-like* и *B3GLCT-like*. Перечисленные гены кодируют белки, вовлеченные в контроль клеточного цикла и репликации ДНК, регуляцию пролиферации и апоптоза клеток; участвующие в липидном и углеводном обменах, развитии воспалительного процесса и работе циркадных ритмов. Благодаря этому рассматриваемые гены-кандидаты могут влиять на формирование экстерьерных особенностей и продуктивные качества овец. Однако необходимы дальнейшие исследования, направленные на подтверждение влияния генов и определение точных механизмов этого воздействия на фенотип.
Ключевые слова: овца; однонуклеотидный полиморфизм; полногеномный поиск ассоциаций; полногеномный анализ ассоциаций; ген-кандидат; российский мясной меринос.

Introduction

Genome-wide association study (GWAS) is a modern and powerful tool for identifying loci and individual polymorphisms associated with economically important traits in various species of productive animals (Georges et al., 2019). Loci associated with reproductive qualities (Abdoli et al., 2019), resistance to parasitic diseases (Yan et al., 2017), indicators of wool (Wang Z. et al., 2014), milk (Garcia-Gámez et al., 2012) and meat productivity (Rovadoscki et al., 2018; Zhang T. et al., 2019) were identified in the sheep genome using GWAS tools.

Most of these studies identify associations with a specific performance trait characteristic of the breed under study. In our opinion, the search for loci associated not with individual parameters of productivity, but with a complex of phenotypic characteristics that determine the breeding value and class of sheep during grading is of particular interest. The division of sheep into classes is carried out according to the aggregate level of wool and meat productivity, constitutional characteristics and the degree of compliance with the breed standard. The most valuable is the study of rare genotypes of outstanding representatives of the breed – exhibition animals, according to the results of the appraisal assigned to the super-elite class. Identification of genetic markers of class opens up opportunities for genetic assessment, selection of highly productive animals and optimal selection of parental pairs capable of transferring their economically valuable characteristics to offspring.

The most common approach of GWAS is to search for associations with the analyzed quantitative trait (for example, live weight) (Gudmundsdottir, 2015). But in the case of a search for associations with belonging to the super-elite class associated with a relatively small sample size, it is advisable to use a non-quantitative analysis approach of the case-control type. In such an analysis, an individual carrying the phenotypic trait of interest gets into the case group, and the individual without the qualities of interest into the control group (Gudmundsdottir, 2015). Previously, non-quantitative analyzes have been successfully performed in sheep for white wool / non-white wool traits (Kijas et al., 2013), multiple pregnancy / non-multiple pregnancy (Xu et al., 2018), high muscle mass / low muscle mass (Gudmundsdottir, 2015). If associations with class are identified during GWAS, the phenotype of an animal of the super-elite class can be designated as "case", and the phenotype of the main herd as "control."

It seems promising to conduct a search for genome-wide associations in animals of the Russian meat merino breed, which combines high wool and meat productivity. Sheep of the Russian meat merino breed exceed the current minimum requirements for sheep of the meat-wool production type in terms of live weight and shearing of washed wool. The average live weight of stud rams is 107 kg, and the live weight of super-elite rams reaches 121 kg (Amerkhanov et al., 2018). Animals are characterized by a strong constitution, hornless rams and ewes, thick, thin and even hair, high vigor and pronounced meat forms (Selionova et al., 2017).

In this regard, the purpose of this study was to identify new candidate genes for economically valuable traits based on the search for single nucleotide polymorphisms (SNPs) associated with belonging to different grading classes in Russian meat merino breed.

Materials and methods

The studies were carried out on the basis of the laboratories of the All-Russian Research Institute of Sheep and Goat Breeding – branch of the North Caucasus Federal Scientific Agricultural Center (Stavropol, Russia), the Skolkovo Institute of Science and Technology "Skoltech" (Moscow, Russia), the Scientific Diagnostic and Veterinary Medicine Center of the Stavropol State Agrarian University (Stavropol, Russia), stud farm "Vtoraya Pyatiletka" of the Stavropol region (Russia).

The object of the study was the Russian meat merino sheep, 12 months old (n = 54), belonging to the breeding group. Based on the results of the assessment carried out, 49 rams were assigned the elite class, they made up the control group (Fig. 1, *a*). Five animals were characterized as super elite. The latter, as outstanding individuals, were selected into the group of exhibition animals and were characterized as animals with the "case" phenotype parameter (Fig. 1, *b*). All rams were clinically healthy.

Quality control of genotyping

Quality control of genotyping was carried out using the PLINK v. 1.07 software (Purcell et al., 2007). The data processing included samples with an indicator of the number of detected SNPs (call rate) greater than 0.95. SNPs with no chromosomal or physical localization, with the minor allele frequency less than 0.01, and the missing genotypes frequency (missing genotype) more than 0.1 were excluded from the analysis. The value p = 0.0001 was used as the threshold value according

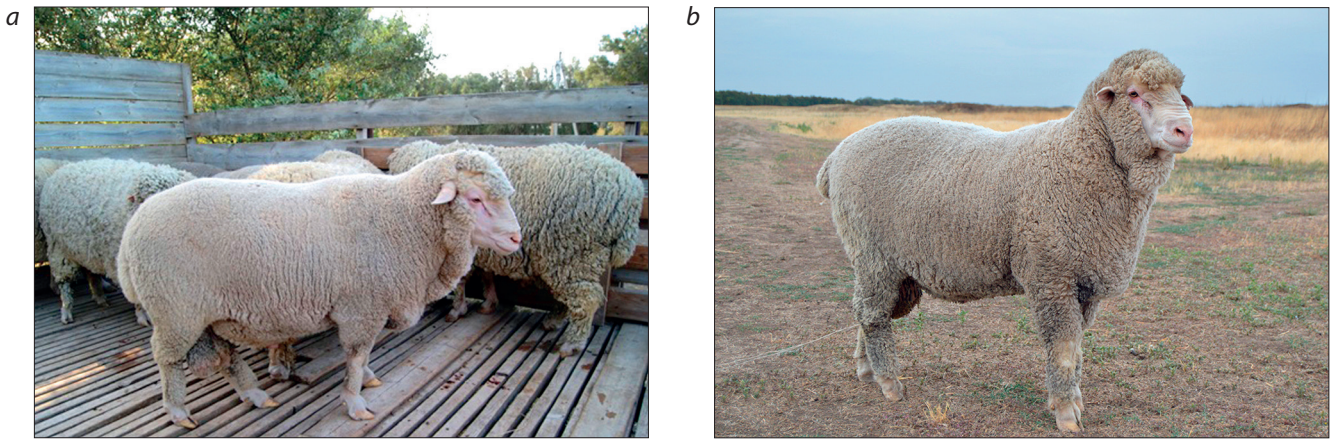


Fig. 1. Russian meat merino sheep breed: *a* – phenotype "case", *b* – phenotype "control."

to the Hardy–Weinberg equilibrium criterion by the Fisher method. With a positive result, 54 samples passed the quality control of genotyping (5 samples of the "case" phenotype, 49 samples of the "control" phenotype). From 606,006 SNPs, 521,829 polymorphisms were used for further analysis.

Genetic and statistical analysis

A genome-wide search for associations was performed using the PLINK v. 1.07 software, the assoc function (Purcell et al., 2007) based on the assessment of the significance of the SNP influence on the attribution class. To confirm the significance of differences in multiple comparisons, the p-score with Bonferroni's correction was used. Visualization and plotting were performed using the QQman package in the R programming language. The search for candidate genes was carried out among the nearest genes located at a distance not exceeding 200,000 bp from SNP, which showed significant differences in the occurrence among animals of the studied groups. In connection with the appearance of updated assemblies of the sheep genome containing updated information on the location and sequences of encoded genes, the location of analyzed SNPs was estimated using the current annotation Oar_rambouillet_v1.0. Gene annotation was performed using the tools of the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov).

Results

As a result of a genome-wide associations search between the frequency of occurrence of individual SNPs and the animals belonging to the exhibition group, more than 50 single nucleotide substitutions were identified that passed the confidence threshold, determined taking into account the Bonferroni correction. The threshold for $-\log_{10}(p)$ values was 0,95 * 7, the top line in the Manhattan plot (Fig. 2).

The results of the differences significance distribution assessment for 26 chromosomes are shown in the quantile-quantile plot. Beginning with $-\log_{10}(p) > 2$, a deviation from the theoretically expected distribution is observed if the null hypothesis is confirmed (Fig. 3).

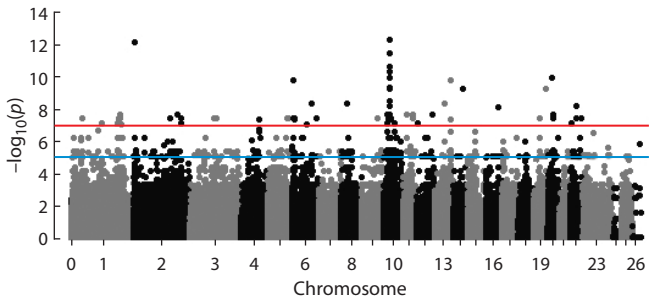


Fig. 2. Manhattan plot of the results of the GWAS with $-\log_{10}(p)$ values for investigated SNP.

Here and also in Fig. 4 the lower line indicates the threshold of the expected significance of differences at the value $-\log_{10}(p) = 5$, the upper line indicates the threshold of high significance of differences at the value of $-\log_{10}(p) = 7$.

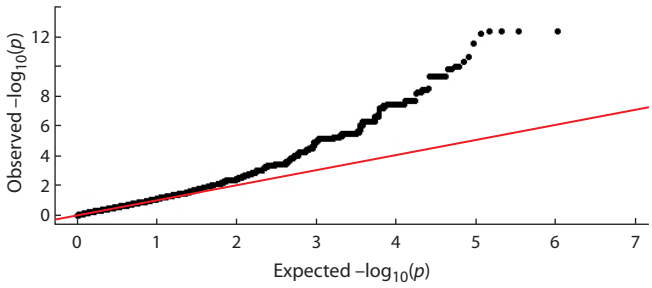


Fig. 3. Quantile-quantile plot for the probabilities of the distribution of the validity of SNP estimates throughout the genome.

The largest number of significant associations was found for polymorphisms located on chromosome 10 (Table). The Manhattan plot shows that the substitutions with the highest confidence value are located relatively close to each other (Fig. 4, *a*).

In a more detailed analysis of their localization, it was found that most of them are concentrated in the region with coordinates from 30859297 to 31873769 1 Mb in length, which

Characterization of the SNP with the highest reliability indicators of association with the exhibition group of animals during the GWAS

Polymorphism	Chromosome / position	Gene / distance (base pair)	A1	F_A	F_U	A2	*p*
rs427646265	10 / 30895552	*RXFP2* / 68303 *EEF1A1* / 66476 *FRY* / 71725	A	0.60	0.01	G	5.34e–13
rs420098635	10 / 30911879	*RXFP2* / 51976 *EEF1A1* / 50149 *FRY* / 88052	A	0.60	0.01	C	5.34e–13
rs426516358	10 / 30964378	*RXFP2* / in exon	G	0.60	0.01	A	5.34e–13
rs424203328	10 / 31020356	*RXFP2* / in intron	A	0.60	0.01	G	5.34e–13
rs417953503	2 / 4742955	*ZBTB21-like* / 145998 *TRNAS-GGA* / 150781	G	0.50	0.00	A	7.62e–13
rs425814243	10 / 31872355	*ALOX5AP* / 74071 *MEDAG* / 25922	G	0.70	0.03	A	3.49e–12
rs425771944	10 / 31867999	*ALOX5AP* / 78427 *MEDAG* / 21566	A	0.60	0.02	C	2.62e–11
rs398157763	10 / 30961940	*RXFP2* / 1915 *EEF1A1* / 88 *FRY* / 138113	A	0.70	0.04	G	5.20e–11
rs408317317	10 / 30859297	*RXFP2* / 104558 *EEF1A1* / 102731 *FRY* / 35470	A	0.50	0.01	G	1.19e–10
rs414101315	20 / 22506181	*OPN5* / in intron	A	0.50	0.01	C	1.19e–10
rs426567665	6 / 5759904	*PRDM 5* / 47457	G	0.40	0.00	A	1.77e–10
rs402834568	13 / 74521952	*PTPRT* / in intron	G	0.40	0.00	A	1.77e–10
rs400005597	10 / 31109147	*RXFP2* / 70714 *B3GLCT-like* / 139308	G	0.70	0.05	A	4.95e–10
rs402948485	10 / 31110090	*RXFP2* / 71657 *B3GLCT-like* / 146293	G	0.70	0.05	A	4.95e–10
rs425859016	10 / 31190471	*RXFP2* / 152038 *B3GLCT-like* / 57984	A	0.60	0.03	G	5.44e–10

Note. A1 – minor allele; A2 – major allele; F_A – frequency of minor allele in the exhibition group of animals; F_U – frequency of minor allele in the selection group.

includes the sequences of 9 different genes. Also, a high reliability of associations was revealed for SNPs located on chromosomes 2, 6, 13, 20. However, on these chromosomes it was not possible to identify areas with a high concentration of reliable associations, since the substitutions are located at a significant distance from each other (Fig. 4, *b*–*e*).

To search for candidate genes, 15 polymorphisms were selected with the highest reliability of associations ($-\log_{10}(p) > 9$), among them one missense mutation in the exon, two substitutions located in gene introns, and eleven substitutions located in intergenic areas (see Table).

High reliability of associations was found for the substitutions rs426516358 and rs424203328 located in exon 18 and intron 1–2 of the *RXFP2* gene, as well as for substitutions located in adjacent intergenic regions. So, the substitutions rs427646265, rs420098635, rs398157763, and rs408317317 are localized in the region between the *RXFP2* and *FRY* genes. Substitutions rs400005597, rs402948485 and rs425859016 – between genes *RXFP2* and *B3GLCT-like*. The rs425814243 and rs425771944 polymorphisms are located in the region between the *ALOX5AP* and *MEDAG* genes. The single nucleotide substitution rs417953503 is located in the intergenic region, practically at an equal distance from the *ZBTB21-like* pseudogen and the gene encoding tRNA TRNAS GGA. The rs426567665 polymorphism is located in the intergenic region, at a distance of 47 kbp from the *PRDM5* gene. The rs402834568 substitution is located in intron 5–6 of the *PTPRT* gene. The rs414101315 polymorphism, highly reliably associated with the super-elite group of animals, is located in intron 4–5 of the *OPN5* gene.

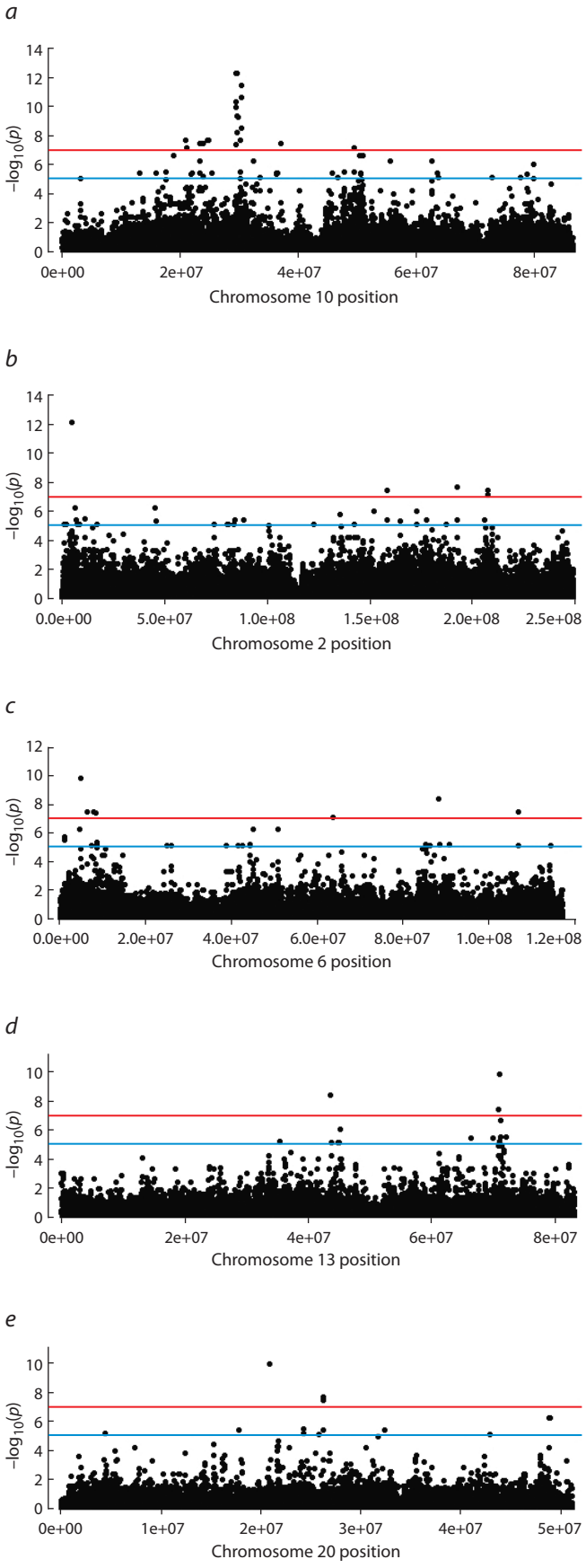


Fig. 4. Manhattan plot that shows the results of the GWAS with $-\log_{10}(p)$ values for investigated SNPs on chromosomes 10, 2, 6, 13 and 20.

Discussion

In the presented work, to identify SNPs associated with performance indicators, a non-quantitative analysis of the case-control type was used, based on comparing the frequency of SNP occurrence in rams of different grading classes, differing in breeding value, wool and meat productivity. A similar approach was previously used to analyze the frequency of SNP occurrence in rams with high and low muscle mass, differing in the level of meat productivity. At the same time, 13 candidate genes for muscle growth and meat productivity were identified on ten different chromosomes (Gudmundsdottir, 2015). As a result of our work, 11 candidate genes were identified on 5 chromosomes, presumably associated with the formation of a complex of phenotypic traits demonstrated by animals of the super-elite class.

Chromosome 10. According to the results of GWAS, one of the most promising candidate genes, probably associated with the belonging of animals to different grading classes in Russian meat merino sheep, is the gene *RXFP2* (*relaxin family peptide receptor 2*), the gene for the relaxin family peptide receptor. The RXFP2 receptor mediates the action of relaxin and insulin-like peptides, which play an important physiological role in the functioning of the reproductive and cardiovascular systems (Scott et al., 2012). The expression level of *RXFP2* positively correlates with the concentration of testosterone in the blood (Johnston et al., 2011). In sheep, *RXFP2* is a marker gene for predicting the type and length of horns (Dominik et al., 2012; Wiedemar, Drögemüller, 2015; Duijvesteijn et al., 2018). Thus, some substitutions associated, according to the results of our studies, with the phenotype of the exhibition animal, were previously proposed to predict the phenotype of the horn. Substitution of rs426516358 in exon 18 of the *RXFP2* gene leads to a change in the encoded amino acid (p.Leu687Phe). According to the results of studies by N. Duijvesteijn et al. (2018), male merino sheep with the GG genotype for the replacement rs426516358 will always be hornless. The substitution rs408317317 has been proposed as a marker of the hornless phenotype for Australian merino sheep (Dominik et al., 2012); its relationship with the type, length, and circumference of the horn base in wild sheep Soay has been revealed (Johnston et al., 2013). The rs398157763 substitution is also associated with horn characteristics in wild Soay sheep (Johnston et al., 2011). There is evidence that, by affecting the formation of horns, polymorphism of the *RXFP2* gene and adjacent regions also affects reproductive success and survival in wild sheep. Most interestingly, in our study, the polymorphism of the *RXFP2* gene and its flanking regions was associated with the conformation characteristics of hornless rams.

Promising candidate genes are also genes located in relative proximity to the *RXFP2* gene and polymorphisms with a high reliability of associations: genes *EEF1A1*, *FRY*, and *B3GLCT-like*. The *EEF1A1* gene (elongation factor 1-alpha 1, *LOC101110773*) is 58 bp away from the *RXFP2* gene in its 3′-flanking region. In humans, the *EEF1A1* gene is responsible for the enzymatic delivery of aminoacyl tRNAs to the ribosome and is involved in the maintenance of cell homeostasis

as a regulator of proliferation and apoptosis (Dapas et al., 2020). The substitution rs398157763 considered in sheep as a marker of polledness is located at a distance of 88 bp from the *EEF1A1* gene. The *FRY* gene (*protein furry homolog, LOC101110521*) encodes a protein that interacts with protein kinases in signaling pathways and induces changes in gene expression. The FRY protein activates the Hippo/Yap pathway, which controls the size of internal organs in animals by regulating cell proliferation and apoptosis (Liu et al., 2019). The *B3GLCT-like* gene (*beta-1,3-glucosyltransferase-like, LOC114116650*) is a homologue of the *B3GLCT* gene, which encodes an enzyme involved in protein metabolism and glycosylation (Weh et al., 2017).

The *ALOX5AP* and *MEDAG* genes are located in relative proximity to the substitutions with high confidence in the associations rs425814243 and rs425771944. The *ALOX5AP* gene (*arachidonate 5-lipoxygenase activating protein*) encodes a protein essential for the synthesis of leukotrienes. It belongs to the family of non-heme iron oxygenases involved in the production and metabolism of fatty acid hydroperoxidases. In sheep, an association of polymorphisms located in the flanking region of the *ALOX5AP* gene with the fat tail phenotype was revealed (Moioli et al., 2015). For fat tailed sheep, the gene was also considered to be associated with climate adaptation (Mastrangelo et al., 2019). *MEDAG* (*mesenteric estrogen dependent adipogenesis*) is an adipogenic gene capable of stimulating the differentiation of preadipocytes into adipocytes, increasing the lipid content and the rate of glucose uptake by cells. It is expressed predominantly in the cells of the visceral fat depot (Zhang H. et al., 2012).

Chromosome 2. The rs417953503 polymorphism identified in the super-elite class is located between the *ZBTB21-like* pseudogene (*LOC101117056*, *zinc finger and BTB domain-containing protein 21-like*) and the *TRNAS GGA transfer RNA* gene (*transfer RNA serine*, *anticodon GGA*). The product of the true gene *ZBTB21* is a negative regulator of transcription for genes that control cell division and DNA replication (Wang J. et al., 2005). In humans, a connection between the *ZBTB21* gene polymorphism and the indicator of physical performance was revealed. Interesting that the *ZBTB21* gene has been proposed as a candidate gene associated with tenderness in beef (Boudon et al., 2020). Transport RNA genes ensure the delivery of activated amino acid residues to the ribosome and their incorporation into the synthesized protein chain. The sheep genome contains 120 copies of the *TRNAS-GGA* gene. In merino sheep, a copy of the *TRNAS-GGA* gene located on chromosome 6 has been proposed as a candidate gene associated with body weight at birth (Dakhlan et al., 2018). In cattle, according to the results of GWAS, polymorphisms located in the flanking regions of the *TRNAS-GGA* genes on chromosomes 6 and 24 are associated with live weight at birth (Edea et al., 2018) and sperm viability (Kaminski et al., 2016).

Chromosome 6. The closest candidate gene with respect to the rs426567665 substitution found in the animals of the exhibition group is the *PRDM5* gene (*PR/SET domain 5*), which encodes a DNA-binding transcription factor that affects the functioning of hematopoietic and microRNA genes. The *PRDM5* gene regulates the intensity of synthesis of proteins involved in the development and maintenance of fibrillar collagens, connective tissue components, and molecules that regulate cell proliferation, differentiation, migration, and adhesion, including the transforming growth factor beta-2 (Burkitt Wright et al., 2011).

Chromosome 13. The rs402834568 substitution was found in the intron region of the *PTPRT* (*protein tyrosine phosphatase receptor type T*) gene, which encodes a protein from the tyrosine phosphatase family that regulates the mitotic cycle, as well as cell growth and differentiation. The *PTPRT* gene is expressed in the cells of the nervous system and regulates the development of neurons (Lee, 2015). In farm animals, a connection between the *PTPRT* gene polymorphism and resistance to some bacterial and parasitic infections was revealed. In goats, polymorphism is associated with resistance to brucellosis (Rossi et al., 2017), in cattle, with resistance to tuberculosis (Bermingham et al., 2014), in Romney sheep, with resistance to invasion by gastrointestinal nematodes (Yan et al., 2017).

Chromosome 20. The *OPN5* gene (*opsin 5*) is expressed in the retina, skin, brain and spinal cord. It encodes the UV-sensitive photopigment neuropsin, which is involved in the regulation of circadian rhythms (Buhr et al., 2019). We propose the *OPN5* gene as a candidate gene, since its intron contains the SNP rs414101315 with high reliability of associations.

Conclusion

In the course of the work done, highly reliable associations were revealed between the belonging of animals to different grading classes and the frequency of occurrence of individual SNPs on chromosomes 2, 6, 10, 13 and 20. Determination of the location of analyzed SNPs relative to the latest annotation Oar_rambouillet_v1.0. made it possible to identify 11 candidate genes, presumably associated with the formation of a complex of phenotypic traits of animals of the exhibition group: *RXFP2*, *ALOX5AP*, *MEDAG*, *OPN5*, *PRDM5*, *PTPRT*, *TRNAS-GGA*, *EEF1A1*, *FRY*, *ZBTB21-like*, *B3GLCT-like*. These genes encode proteins with a number of important biological functions involved in the control of the cell cycle and DNA replication, regulation of cell proliferation and apoptosis, involved in lipid and carbohydrate metabolism, the development of the inflammatory process, and the work of circadian rhythms. Due to this, the candidate genes under consideration can influence the formation of conformational characteristics and productive qualities of sheep. However, further research is needed to confirm the effect of genes and to determine the exact mechanisms of this effect on the phenotype.

References

Abdoli R., Mirhoseini S.Z., Ghavi Hossein-Zadeh N., Zamani P., Moradi M.H., Ferdosi M.H., Gondro C. Genome-wide association study of first lambing age and lambing interval in sheep. *Small Rumin. Res*. 2019;178:43-45. DOI 10.1016/j.smallrumres.2019.07.014.

Amerhanov H.A., Egorov M.V., Selionova M.I., Shumaenko S.N., Efimova N.I. A new breed of sheep: Russian meat Merino. *Sel'skokhozyajstvennyj Zhurnal = Agricultural Journal.* 2018;

1(11):50-56. DOI 10.25930/0372-3054-2018-1-11-65-71. (in Russian)

Bermingham M.L., Bishop S.C., Woolliams J.A., Pong-Wong R., Allen A.R., McBride S.H., Ryder J.J., Wright D.M., Skuce R.A., McDowell S.W., Glass E.J. Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. *Heredity.* 2014;112(5):543-551. DOI 10.1038/hdy.2013.137.

Boudon S., Henry-Berger J., Cassar-Malek I. Aggregation of omic data and secretome prediction enable the discovery of candidate plasma biomarkers for beef tenderness. In *Int. J. Mol. Sci.* 2020; 21(2):54-63. DOI 10.3390/ijms21020664.

Buhr E.D., Vemaraju S., Diaz N., Lang R.A., Van Gelder R.N. Neuropsin (OPN5) Mediates Local Light-Dependent Induction of Circadian Clock Genes and Circadian Photoentrainment in Exposed Murine Skin. *Curr. Biol.* 2019;29(20):3478-3487. DOI 10.1016/ j.cub.2019.08.063.

Burkitt Wright E.M.M., Spencer H.L., Daly S.B., Manson F.D.C., Zeef L.A.H., Urquhart J., Zoppi N., Bonshek R., Tosounidis I., Mohan M., Madden C., Dodds A., Chandler K. E., Banka S., Au L., Clayton-Smith J., Khan N., Biesecker L.G., Wilson M., Black G.C.M. Mutations in PRDM5 in Brittle Cornea Syndrome Identify a Pathway Regulating Extracellular Matrix Development and Maintenance. *Am. J. Hum. Genet.* 2011;88(6):767-777. DOI 10.1016/j.ajhg.2011.05.007.

Dakhlan A., Moghaddar N., Gondro C., Werf J.H.J. Gene by birth type interaction in merino lamb. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 2018;22:45-48.

Dapas B., Pozzato G., Zorzet S., Capolla S., Paolo M., Scaggiante B., Coan M., Guerra C., Gnan C., Gattei V., Zanconati F., Grassi G. Effects of eEF1A1 targeting by aptamer/siRNA in chronic lymphocytic leukaemia cells. *Int. J. Pharm.* 2020;57(4):48-59. DOI 10.1016/j.ijpharm.2019.118895.

Dominik S., Henshall J.M., Hayes B.J. A single nucleotide polymorphism on chromosome 10 is highly predictive for the polled phenotype in Australian Merino sheep. *Anim. Genet.* 2012; 43(4):468-470. DOI 10.1111/j.1365-2052.2011.02271.x.

Duijvesteijn N., Bolormaa S., Daetwyler H.D., Van Der Werf J.H.J. Genomic prediction of the polled and horned phenotypes in Merino sheep. *Genet. Sel. Evol.* 2018;50(1):1-11. DOI 10.1186/ s12711-018-0398-6.

Edea Z., Jeoung Y.H., Shin S.S., Ku J., Seo S., Kim I.H., Kim S.W., Kim K.S. Genome–wide association study of carcass weight in commercial Hanwoo cattle. *Asian-Australas. J. Anim. Sci.* 2018; 31(3): 327-334. DOI 10.5713/ajas.17.0276.

García-Gámez E., Gutiérrez-Gil B., Sahana G., Sánchez J.P., Bayón Y., Arranz J.J. GWA Analysis for Milk Production Traits in Dairy Sheep and Genetic Support for a QTN Influencing Milk Protein Percentage in the LALBA Gene. *PLoS ONE.* 2012; 7(10):1-9. DOI 10.1371/journal.pone.0047782.

Georges M., Charlier C., Hayes B. Harnessing genomic information for livestock improvement. *Nat. Rev. Genet*. 2019;20(3): 135-156. DOI 10.1038/s41576-018-0082-2.

Gudmundsdottir O.O. Genome-wide association study of muscle traits in Icelandic sheep. Agricultural University of Iceland, 2015.

Johnston S.E., Gratten J., Berenos C., Pilkington J.G., Clutton-Brock T.H., Pemberton J.M., Slate J. Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature.* 2013;502(7469):93-95. DOI 10.1038/nature12489.

Johnston S.E., McEwan J.C., Pickering N.K., Kijas J.W., Beraldi D., Pilkington J.G., Pemberton J. M., Slate J. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Mol. Ecol.* 2011;20(12):2555-2566. DOI 10.1111/ j.1365-294X.2011.05076.x.

Kamiński S., Hering D.M., Oleński K., Lecewicz M., Kordan W. Genome-wide association study for sperm membrane integrity in frozen-thawed semen of Holstein-Friesian bulls. *Anim. Reprod. Sci.* 2016;170:135-140. DOI 10.1016/j.anireprosci.2016.05.002.

Kijas J.W., Serrano M., Mcculloch R., Li Y., Salces Ortiz J., Calvo J.H., Pérez-Guzmán M.D. Genomewide association for a dominant pigmentation gene in sheep. *J. Anim. Breed. Genet.* 2013;130(6):468-475. DOI 10.1111/jbg.12048.

Lee J.R. Protein tyrosine phosphatase PTPRT as a regulator of synaptic formation and neuronal development. *BMB Rep*. 2015; 48(5):249-255. DOI 10.5483/BMBRep.2015.48.5.037.

Liu Y., Chen X., Gong Z., Zhang H., Fei F., Tang X., Wang J., Xu P., Zarbl H., Ren X. Fry Is Required for Mammary Gland Development During Pregnant Periods and Affects the Morphology and Growth of Breast Cancer Cells. *Front. Oncol.* 2019;4(2):1-12. DOI 10.3389/fonc.2019.01279.

Mastrangelo S., Moioli B., Ahbara A., Latairish S., Portolano B., Pilla F., Ciani E. Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. *Anim. Prod. Sci.* 2019;59(5):835-842. DOI 10.1071/AN17753.

Moioli B., Pilla F., Ciani E. Signatures of selection identify loci associated with fat tail in sheep. *J. Anim. Sci.* 2015;93(10):4660-4669. DOI 10.2527/jas.2015-9389.

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

Rossi U.A., Hasenauer F.C., Caffaro M.E., Neumann R., Salatin A., Poli M.A., Rossetti C.A. A haplotype at intron 8 of PTPRT gene is associated with resistance to Brucella infection in Argentinian creole goats. *Vet. Microbiol*. 2017;2(3):133-137. DOI 10.1016/ j.vetmic.2017.06.001.

Rovadoscki G.A., Pertile S.F.N., Alvarenga A.B., Cesar A.S.M., Pértille F., Petrini J., Franzo V., Soares W.V.B., Morota G., Spangler M.L., Pinto L.F.B., Carvalho G.G.P., Lanna D.P.D., Coutinho L.L., Mourão G.B. Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Ines sheep. *BMC Genom.* 2018;19(1):1-14. DOI 10.1186/ s12864-018-4777-8.

Scott D.J., Rosengren K.J., Bathgate R.A.D. The different ligand-binding modes of relaxin family peptide receptors RXFP1 and RXFP2. *Mol. Endocrinol*. 2012;26(11):1896-1906. DOI 10.1210/me.2012-1188.

Selionova M.I., Shumaenko S.N., Efimova N.I. Surov A.I. Bobrishov S.S. Target indicators and characteristics of the Russian Meat Merino breed: Proceedings of the Research Institute for Sheep and Goat Farming. *Sel'skokhozyajstvennyj Zhurnal = Agricultural Journal.* 2017;2(10):10-16. (in Russian)

Wang J., Kudoh J., Takayanagi A., Shimizu N. Novel human BTB/ POZ domain-containing zinc finger protein ZNF295 is directly associated with ZFP161. *Biochem. Biophys. Res. Commun.* 2005;327(2):615-627. DOI 10.1016/j.bbrc.2004.12.048.

Wang Z., Zhang H., Yang H., Wang S., Rong E., Pei W., Li H., Wang N. Genome-wide association study for wool production traits in a Chinese merino sheep population. *PLoS ONE.* 2014; 9(9):3-10. DOI 10.1371/journal.pone.0107101.

Weh E., Takeuchi H., Muheisen S., Haltiwanger R.S., Semina E.V. Functional characterization of zebrafish orthologs of the hu-

man Beta 3-Glucosyltransferase B3GLCT gene mutated in Peters Plus Syndrome. *PLoS ONE.* 2017;12(9):e0184903. DOI 10.1371/journal.pone.0184903.

Wiedemar N., Drögemüller C.A 1.8-kb insertion in the 3′-UTR of RXFP2 is associated with polledness in sheep. *Anim. Genet.* 2015;46(4):457-461. DOI 10.1111/age.12309.

Xu S.S., Gao L., Xie X.L., Ren Y.L., Shen Z.Q., Wang F., Shen M., Eypórsdóttir E., Hallsson J.H., Kiseleva T., Kantanen J., Li M.H. Genome-wide association analyses highlight the potential for different genetic mechanisms for litter size among sheep breeds. *Front. Genet*. 2018;6(2):1-14. DOI 10.3389/fgene.2018.00118.

Yan J., Dukkipati V., Blair H.T., Biggs P.J., Hamie J.C., Aw G. A genome-wide scan of positive selection signature using Ovine Infinium HD SNP BeadChip in two Romney lines, selected for resistance or resilience to nematodes. *Anim. Genet.* 2017;4: 87-94.

Zhang H., Chen X., Sairam M.R. Novel genes of visceral adiposity: Identification of mouse and human Mesenteric Estrogen-Dependent Adipose (MEDA)-4 gene and its adipogenic function. *Endocrinology.* 2012;153(6):2665-2676. DOI 10.1210/en.2011-2008.

Zhang T., Gao H., Sahana G., Zan Y., Fan H., Liu J., Shi L., Liu J., Du L., Wang L., Zhao F. Genome-wide association studies revealed candidate genes for tail fat deposition and body size in the Hulun Buir sheep. *J. Anim. Breed. Genet.* 2019;6(1):1-9. DOI 10.1111/jbg.12402.

ORCID ID

A.Y. Krivoruchko orcid.org/0000-0003-4536-1814
O.A. Yatsyk orcid.org/0000-0003-2730-2482
E.Y. Safaryan orcid.org/0000-0002-2049-842X

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

Received May 19, 2020. Revised September 15, 2020. Accepted October 29, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):844-852
DOI 10.18699/VJ20.682

Impact of sex on the adaptation of adult mice to long consumption of sweet-fat diet

N.M. Bazhan[1, 2] ✉, T.V. Iakovleva[1], A.D. Dubinina[1], E.N. Makarova[1]

[1] Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
✉ e-mail: bazhan-nm@yandex.ru

Abstract. In rodents, the most adequate model of human diet-induced obesity is obesity caused by the consumption of a sweet-fat diet (SFD), which causes more pronounced adiposity in females than in males. The aim of this work was to determine the sex-associated effect of SFD on the expression of genes related to carbohydrate-lipid metabolism in adult mice. For 10 weeks, male and female C57Bl mice were fed a standard laboratory chow (Control group) or a diet, which consisted of laboratory chow supplemented with sweet cookies, sunflower seeds and lard (SFD group). Weights of body, liver and fat depots, blood concentrations of hormones and metabolites, liver fat, and mRNA levels of genes involved in regulation of energy metabolism in the liver, perigonadal and subcutaneous white adipose tissue (pgWAT, scWAT) and brown adipose tissue (BAT) were measured. SFD increased body weight and insulin resistance in mice of both sexes. Female mice that consumed SFD (SFD females) had a greater increase in adiposity than SFD males. SFD females showed a decreased expression of genes related to lipogenesis (*Lpl*) and glucose metabolism (*G6pc*, *Pklr*) in liver, as well as lipogenesis (*Lpl*, *Slca4*) and lipolysis (*Lipe*) in pgWAT, suggesting reduced energy expenditure. In contrast, SFD males showed increased lean mass gain, plasma insulin and FGF21 levels, expressions of *Cpt1a* gene in pgWAT and scWAT and *Pklr* gene in liver, suggesting enhanced lipid and glucose oxidation in these organs. Thus, in mice, there are sex-dependent differences in adaptation to SFD at the transcriptional level, which can help to explain higher adiposity in females under SFD consumtion.
Key words: C57BL/6J mice; sweet-fat diet; adiposity; sex differences; liver; adipose tissue; FGF21, insulin; gene expression.

For citation: Bazhan N.M., Iakovleva T.V., Dubinina A.D., Makarova E.N. Impact of sex on the adaptation of adult mice to long consumption of sweet-fat diet. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):844-852. DOI 10.18699/VJ20.682

Влияние пола на адаптацию взрослых мышей к длительному потреблению сладко-жирной диеты

Н.М. Бажан[1, 2] ✉, Т.В. Яковлева[1], А.Д. Дубинина[1], Е.Н. Макарова[1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ e-mail: bazhan-nm@yandex.ru

Аннотация. Наиболее адекватной моделью диет-индуцированного ожирения у человека является ожирение грызунов, вызванное потреблением сладко-жирной диеты (СЖД), которая в большей степени увеличивает долю жира у женщин, чем у мужчин. Целью работы было определение обусловленного полом влияния СЖД на экспрессию генов, контролирующих углеводно-жировой обмен у взрослых мышей. Самцов и самок мышей линии C57BL/6J кормили в течение 10 недель стандартной лабораторной пищей (контрольная группа) или диетой, которая состояла из лабораторной пищи с добавлением сладкого печенья, семян подсолнечника и сала (группа СЖД). Были измерены вес тела, печени и жировых депо, концентрация гормонов и метаболитов в крови, содержание жира в печени и уровни мРНК генов, участвующих в регуляции энергетического обмена, в печени, окологонадном и подкожном белом жире и в буром жире. Потребление СЖД вызвало ожирение (у самок в большей степени, чем у самцов) и резистентность к инсулину у мышей обоих полов. У самок, получавших СЖД, была снижена относительно контроля экспрессия генов печени, связанных с липогенезом (*Lpl*), метаболизмом глюкозы (*G6pc*, *Pklr*), и генов белого жира, связанных с липогенезом (*Lpl*, *Slca4*) и липолизом (*Lipe*), что предполагает снижение расхода энергии в этих тканях. Потребление СЖД у самцов, в отличие от самок, резко повысило уровни в крови инсулина и FGF21, а также экспрессию гена *Cpt1a* в окологонадном и подкожном белом жире и гена *Pklr* в печени, что свидетельствует об усилении окисления липидов и глюкозы в этих тканях. Таким образом, у мышей

были выявлены половые различия в адаптации к СЖД как на уровне целого организма, так и на уровне транскрипции генов. Эти результаты могут способствовать развитию поло-специфических подходов к коррекции ожирения у человека.
Ключевые слова: мыши C57BL/6J; сладко-жирная диета; ожирение; половые различия; печень; жировая ткань; FGF21; инсулин; экспрессия генов.

Introduction

In the human population, there is a significant increase in the number of people suffering from obesity and associated metabolic diseases such as type 2 diabetes, cardiovascular diseases and non-alcoholic fatty liver. The mechanisms of obesity development are studied in laboratory animals with various models of diet-induced obesity. Among the high-calorie diets, the high-fat and the sweet-fat diet (SFD), or the cafeteria diet are the most popular. SFD is most consistent with the consumption of "pleasant" food, which provokes the development of obesity in the human population (Sampey et al., 2011). A special study carried out on male rats showed that SFD more effectively than a high-fat diet induced the development of obesity, hyperphagia, and increased blood cholesterol and leptin levels (Buyukdere et al., 2019).

It is known that most of the characteristics of energy metabolism differ in males and females (Mauvais-Jarvis, 2015). However, the question of the impact of sex on the adaptation of adult mice to long-term consumption of a SFD remains unexplored.

Fibroblast Growth Factor 21 (FGF21) is a protein hormone of the liver that helps the body adapt to metabolic stresses (hunger, cold, overeating and obesity) (Fisher et al., 2010). Exogenous FGF21 reduces body weight, normalizes the lipid profile, and increases insulin sensitivity in various models of obesity and insulin resistance (Zhang, Li, 2014). Earlier, we and others showed that, SFD dramatically increased blood FGF21 level and its hepatic gene expression in mature male, but not female mice (Chukijrungroat et al., 2017; Gasparin et al., 2018; Bazhan et al., 2019). Based on this, it can be assumed that adult males and females will differ in the ways of adaptation to the consumption of SFD. The effects of FGF21 are partially realized through the regulation of the expression of genes controlling carbohydrate-lipid metabolism in the liver, white and brown fat (Coskun et al., 2008; Camporez et al., 2013). The aim of this work was to study the ways of adaptation to the consumption of SFD at the level of the whole organism and at the level of expression of genes involved in lipid and carbohydrate metabolism in the liver and adipose tissue, in mature male and female mice.

Materials and methods

All experiments were performed according to the European Convention for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes (Council of Europe No. 123, Strasbourg, 1985) and Russian national instructions for the care and use of laboratory animals. The protocols were approved by the Independent Ethics Committee of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences.

Animals. Ten-week-old C57BL mice (the vivarium of the Institute of Cytology and Genetics) were used. Both male and female mice were housed in group (3 mice per cage) and were fed with standard laboratory chow (Assortiment Agro, Moscow region, Turacovo, Russia) (control diet, control) or with mixed diet, which consisted of standard laboratory chow supplemented with sweet cookies, sunflower seeds and lard (sweet-fat diet, SFD). There were 4 experimental groups (5–7 mice per group): control male, control female, SFD male and SFD female.

Mice were killed by decapitation after 10 weeks of diet, liver, white adipose tissue (WAT) of different localizations (perigonadal, pgWAT, subcutaneous, scWAT, and perirenal), and interscapular brown adipose tissue (BAT) were weighed. Lean body weight was determined by subtracting the total fat mass from the body weight. Gene expression was measured in the samples of these tissues, excluding perirenal WAT.

Assay of plasma biochemical parameters. Trunk blood was collected in test tubes with EDTA after decapitation, centrifuged and plasma was stored at –20 °C until the assay of hormones and metabolites. Concentrations of FGF21, insulin, adiponectin, and leptin were measured using the following ELISA Kits: Rat/Mouse Fibroblast Growth Factor-21 ELISA Kit, Rat/Mouse Insulin ELISA Kit, Mouse Adiponectin ELISA Kit и Mouse Leptin ELISA Kit (Millipore, St. Louis, MI, USA). Concentrations of glucose, free fatty acids (FFA), triglycerides (TG), and cholesterol were measured colorimetrically using Fluitest GLU, Fluitest TG, Fluitest CHOL (Analyticon Biotechnologies AG, Lichtenfels, Germany) and NEFA FS kits (non-esterified fatty acids) (DiaSys, Germany).

Glucose tolerance and insulin tolerance tests. On the day of testing, the animals were removed from the food at 10:00 am, and the water was left *ad libitum*. Insulin tolerance test (ITT) started at 2:00 pm, glucose tolerance test (GTT) – at 4:00 pm. In GTT, a glucose solution in water at a dose of 2 mg/1 g of body weight was administered orally. In the ITT, animals were injected intraperitoneally with protofan in physiological saline at a dose (0.5 IU/1 kg of body weight). The glucose level was determined in the blood from the tail vein using test strips and a OneTouch Select glucometer (Lifescan; Johnson and Johnson, USA) before drug administration and 15, 30, 60, and 120 minutes after administration in GTT and after 15, 30, 60 minutes in ITT.

The reaction of reverse transcription and real-time PCR. Total RNA was isolated from tissue samples with ExtractRNA (Evrogen, Moscow, Russia) according to the manufacturer's instructions. First-strand cDNA was synthesized with Moloney murine leukemia virus (MMLV) reverse transcriptase (Evrogen, Moscow, Russia) and oligo (dT) as a primer. Applied Biosystems TaqMan Gene Expression Assays, listed in Table 1, and qPCRmix-HS LowROX Master Mix (Evrogen, Moscow, Russia) were used for relative quantitation real-time PCR with β-actin as an endogenous control. Sequence amplification and fluorescence detection were performed with the Applied Biosystems ViiA™ 7 Real-Time PCR System (Life Technologies, 5791 Van Allen Way, Carlsbad, CA, USA).

Table 1. Taqman gene expression assays for mice (Applied Biosystems)

Name	Symbol	Catalogue No.
Carnitine palmitoyltransferase 1α	*Cpt1α*	Mm01231183_m1
Carnitine palmitoyltransferase 1β	*Cpt1β*	Mm00487191_g1
Deiodinase, iodothyronine, type II	*Dio2*	Mm00515664_m1
Fatty acid synthase	*Fasn*	Mm00662319_m1
Glucose-6-phosphatase, catalytic	*G6pc*	Mm00839363_m1
Glucokinase	*Gck*	Mm00439129_m1
Lipase, hormone sensitive	*Lipe*	Mm00495359_m1
Lipoprotein lipase	*Lpl*	Mm00434764_m1
Phosphoenolpyruvate carboxykinase 1, cytosolic	*Pck1*	Mm01247058_m1
Pyruvate kinase liver and red blood cell	*Pklr*	Mm00443090_m1
Solute carrier family 2 (facilitated glucose transporter), member1 (GLUT1)	*Slc2a1*	Mm00441480_m1
Solute carrier family 2 (facilitated glucose transporter), member 2 (GLUT2)	*Slc2a2*	Mm00446229_m1
Solute carrier family 2 (facilitated glucose transporter), member 4 (GLUT4)	*Slc2a4*	Mm00436615_m1
Uncoupling protein 1 (mitochondrial, proton carrier)	*Ucp1*	Mm01244861_m1
Beta-actin	*Actb*	Mm00607939_s1

Relative quantitation was performed by the comparative CT method, where CT is the cycle threshold.

Statistical analysis. The results are presented as means ± SE from the indicated number of mice. Two-way ANOVA with factors sex (male, female) and diet (standard diet, control group and sweet and fat diet, SFD group) was used to analyze effect of sex and SFD on blood parameters, gene expression and area under curves in GTT and ITT with multiple comparisons using the post hoc Tukey test. Three-way ANOVA with factors sex, diet, and time (minutes 0, 15, 30, 60, 120 for GTT and 0, 15, 30, 60 for ITT) was used to analyze the results of GTT and ITT. Where indicated, groups were also compared using Student's *t*-test. Significance was determined as $p < 0.05$. The STATISTICA 6 software package (StatSoft, USA) was used for analysis.

Results

Weight characteristics

In females, body weight was lower than in males in both groups ($P < 0.001$) (Table 2). Under the SFD, both male and female mice gained more weight than their respective control diet fed counterpart ($P < 0.001$). FD consumption increased body weight: in males – by 39 %, and in females – by 40 % and contributed to the maximum manifestation of sex differences.

In females, hepatic weight and index were lower than in males ($P < 0.001$ for both parameters). Consumption of SFD increased hepatic weight ($P < 0.001$), but did not affect its relative weight in males and females. Maximum sex differences in absolute and relative hepatic weight were manifested only under SFD-induced obesity. An increase in liver mass was as-

Table 2. Weight-related parameters in mice, fed standard chow (control) and sweet-fat diet

Parameter	Males		Females		ANOVA
	Control	SFD	Control	SFD	
Body weight, g	29.3 ± 0.6	40.8 ± 1.9*	24.4 ± 0.7	35.1 ± 1.3*#	D, S
Liver, g	1.48 ± 0.04	2.10 ± 0.13*	1.16 ± 0.06	1.47 ± 0.06#	D, S
Liver index, % BW	5.2 ± 0.3	5.1 ± 0.4	4.8 ± 0.4	4.2 ± 0.2#	S
pgWAT, g	0.7 ± 0.2	1.6 ± 0,2	0.8 ± 0.1	3.3 ± 0.4*#	D, S, S*D
pgWAT index, % BW	2.4 ± 1.3	4.0 ± 1.4	3.2 ± 0.9	9.1 ± 0.8*#	D, S, S*D
scWAT, g	0.6 ± 0.1	2.2 ± 0.3*	0.8 ± 0.1	3.6 ± 0.4*#	D, S, S*D
scWAT index, % BW	2.3 ± 1.0	5.1 ± 2.0*	3.4 ± 0.5	10.2 ± 0.1*#	D, S, S*D
BAT, mg	0.1 ± 0.01	0.2 ± 0.03*	0.1 ± 0.02	0.15 ± 0.03#	D
BAT index, % BW × 10	3.4 ± 0.4	5.7 ± 0.2	4.0 ± 2.0	4.4 ± 0.7	D
Lean mass, g	27.0 ± 0.5	36.2 ± 1.5*	22.5 ± 0.6	27.2 ± 0.6*#	D, S, S*D
TG content, mg/g of liver	73 ± 8	192 ± 70	148 ± 17	222 ± 25	D

BW – body weight, two-way ANOVA was used with the factors S, sex effect; D, diet effect; and D*S, interactive effect of sex and diet. * $p < 0.05$ versus control group, # $p < 0.05$ versus males in the same group by post-hoc Tukey test

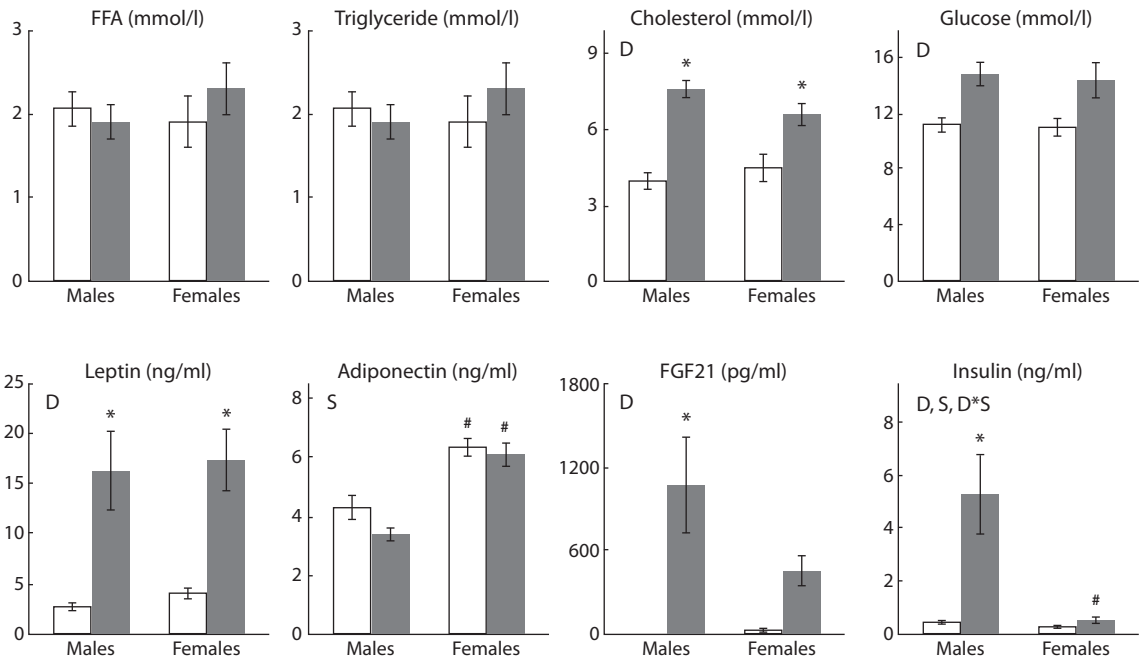


Fig. 1. Serum biochemical parameters in mice, fed standard chow (control, white columns) and sweet-fat diet (grey columns).

Two-way ANOVA was used with the factors S, sex effect; D, diet effect; and D*S, interactive effect of sex and diet. * $p < 0.05$ versus control group, # $p < 0.05$ versus males in the same group by post-hoc Tukey test.

sociated with an increase in hepatic fat deposition: the content of triglycerides (TG) in the liver, increased upon consumption of SFD ($P < 0.05$) in mice of both sexes.

In females, the mass and index of pgWAT were higher than in males ($P < 0.01$ for both parameters). SFD consumption increased them ($P < 0.001$ for both parameters) largely in females than in males (interaction of factors $P < 0.01$ for both parameters) and contributed to the maximum manifestation of sex differences.

The mass and proportion of scWAT in females were higher than in males ($P < 0.05$ and $P < 0.001$ respectively). Consumption of SFD increased the scWAT mass and index ($P < 0.001$ for both cases) largely in females than in males (interaction of factors $P < 0.07$ for weight and $P < 0.01$ for index) and contributed to the manifestation of significant sex differences.

In the control group, the BAT weight in males and females did not differ. SFD increased the BAT weight and index ($P < 0.01$ and $P < 0.05$ respectively), however, the increase, in contrast to the SFD effect on the pgWAT weight, was significantly more pronounced in males than in females and was statistically significant. As a result, the BAT weight in females was significantly lower than in males only under the SFD ($P < 0.05$).

In females, the lean mass was significantly lower than in males ($P < 0.001$). The consumption of SFD increased lean mass in mice of both sexes ($P < 0.05$), but in males largely (interaction of factors $P < 0.001$), thereby enhancing the expression of sex differences.

Plasma metabolite and hormone levels

In females, blood insulin levels were lower and adiponectin levels were higher than in males ($P < 0.05$ for insulin and $P < 0.001$ for adiponectin) in both groups (Fig. 1). SFD consumption increased blood levels of glucose, insulin, cholesterol, fibroblast growth factor (FGF21), and leptin ($P < 0.01$ for glucose, insulin, FGF21 and $P < 0.001$ for cholesterol and leptin) and did not alter the levels of free fatty acids (FFA), TG and adiponectin in mice of both sexes. Sex dimorphism was revealed only in the response of insulin and FGF21 to the SFD. Plasma insulin concentrations increased only in males and did not change in females, as evidenced by the significant interaction of factors sex and diet ($P < 0.05$). Plasma FGF21 concentration also significantly and reliably increased only in SFD males, while in SFD females the increase was less pronounced and not significant.

Glucose tolerance and insulin tolerance tests

In control males, insulin sensitivity was lower than that of control females. SFD consumption reduced glucose tolerance and insulin sensitivity in both males and females ($P < 0.001$ in all cases) (Fig. 2). However, the effect of the SFD was more pronounced in females: the fasting blood glucose level and the glucose excretion curve in the ITT in the SFD females were higher than in the control ($p < 0.05$ in both cases), while in the SFD males these parameters did not differ from the control.

Gene expression in metabolic tissues

Among the studied hepatic genes, only *Lpl* expression was dependent on sex ($P < 0.01$): it was lower in females than in males. The consumption of SFD down regulated the expression of this gene regardless of sex (Fig. 3). The consumption of SFD was accompanied by sex-dependent changes in the expression of the *Fasn* (fatty acid synthesis), *G6pc* (gluconeogenesis), and *Pklr* (glycolysis) genes: SFD males showed increased, while SFD females – decreased the mRNA levels of these genes in relation to control (interaction of factors

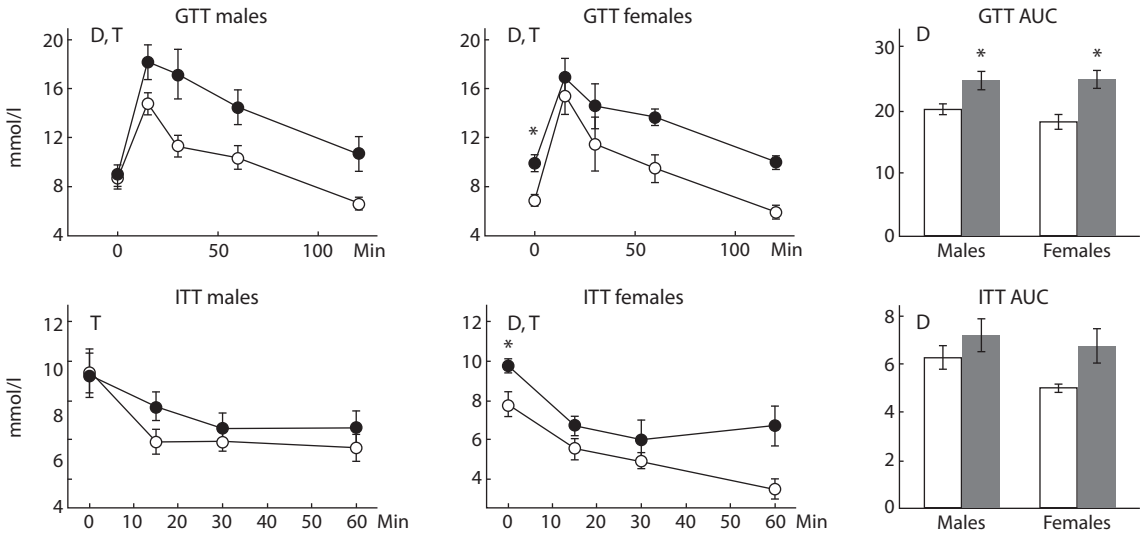


Fig. 2. Blood glucose level and area under the curve (AUC) in GTT and ITT in mice, fed standard chow (control, white symbols) and sweet-fat diet (black symbols).

Three-way ANOVA with factors sex, diet, and time (minutes 0, 15, 30, 60, 120 for GTT and 0, 15, 30, 60 for ITT) was used. T, time effect, and D, diet effect. * $p < 0.05$ versus control group by post-hoc Tukey test.

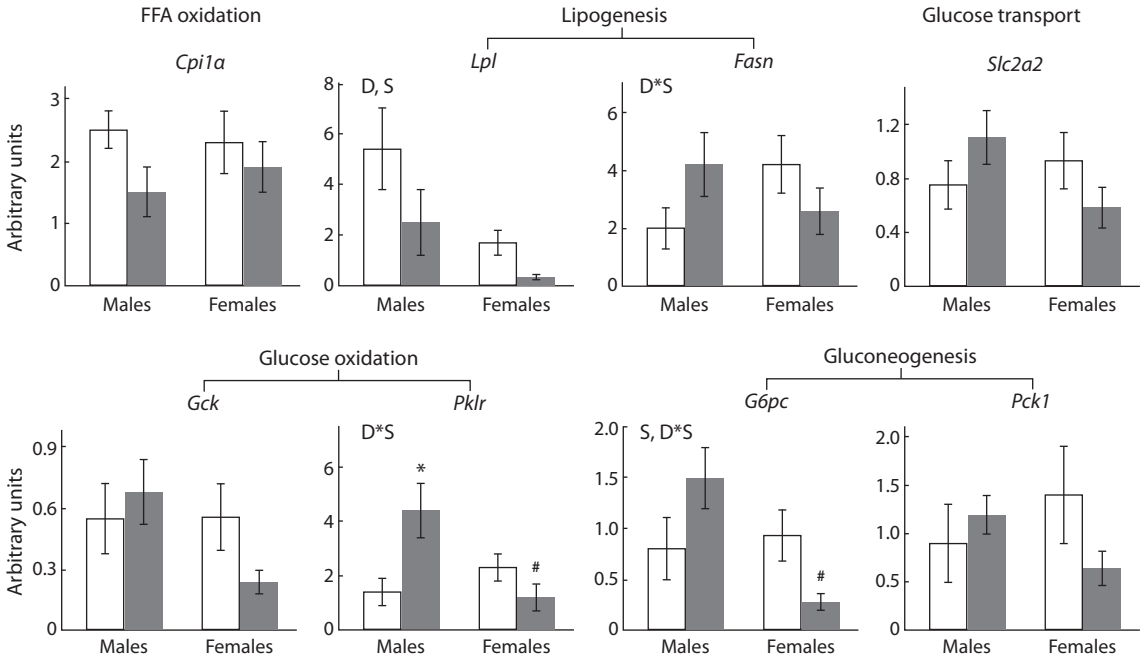


Fig. 3. The mRNA levels of hepatic genes involved in glucose and lipid metabolism in mice, fed standard chow (control, white columns) and sweet-fat diet (grey columns).

Two-way ANOVA was used with the factors S, sex effect; D, diet effect; and D*S, interactive effect of sex and diet. * $p < 0.05$ versus control group, # $p < 0.05$ versus males in the same group by post-hoc Tukey test.

$P < 0.05$ for all genes). As a result, in SFD females, *G6pc* gene expression was fivefold and *Pklr* gene expression was 2.4 times lower than in SFD males ($P < 0.05$).

There were no sex differences in the expression of the studied genes in pgWAT (Fig. 4, *a–f*). The consumption of SFD influenced the expression of *Cpt1α* (fatty acid oxidation), *Lipe* (lipolysis), and *Lpl* (lipogenesis) genes differently in males and females (interaction of factors $p < 0.05$ in all cases): only in males, *Cpt1α* mRNA level increased, only in females, *Lipe* and *Lpl* mRNA levels decreased. SFD down regulated *Slc2a4* gene expression regardless of sex ($P < 0.01$). However, the decrease was more pronounced in females (12 times) than in males (2.7 times).

In scWAT, in contrast to pgWAT, sex influenced the expression of the *Lipe* and *Slc2a4* genes ($P < 0.05$ for *Lipe*, $P < 0.01$ for *Slc2a4*): it was lower in females than in males

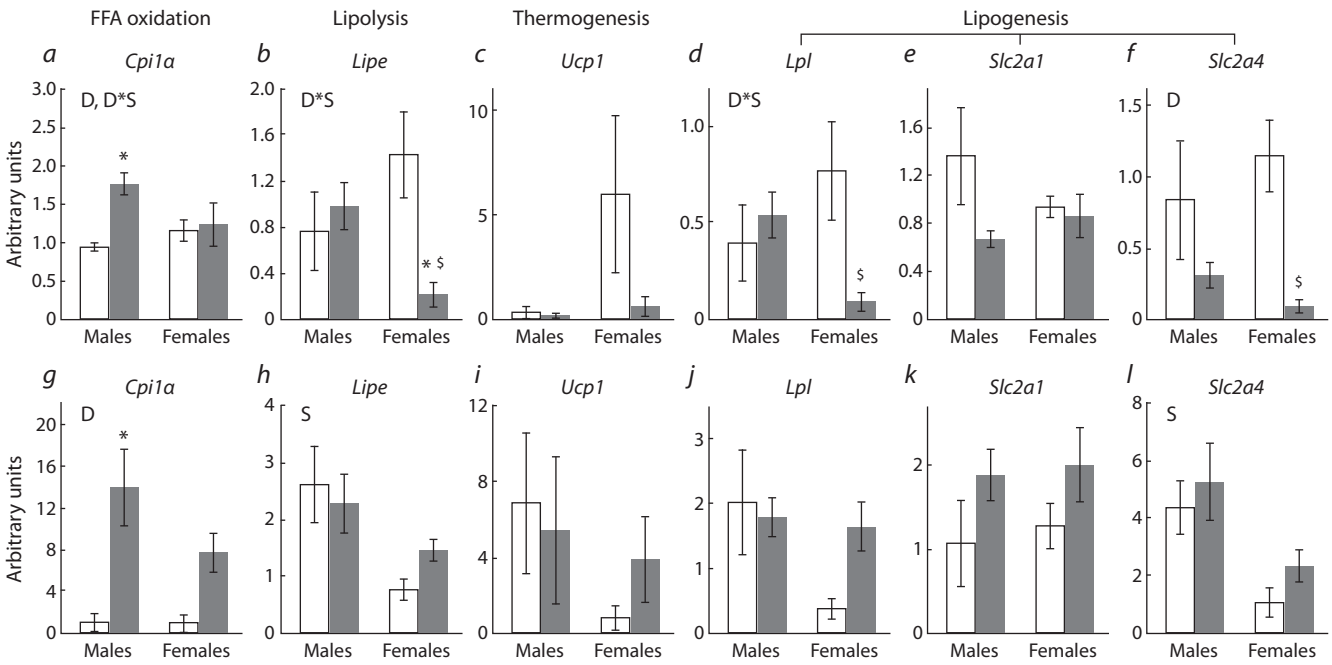


Fig. 4. The mRNA levels of pgWAT (*a–f*) and scWAT (*g–l*) genes involved in lipid metabolism in mice, fed standard chow (control, white columns) and sweet-fat diet (grey columns).

Two-way ANOVA was used with the factors S, sex effect; D, diet effect; and D*S, interactive effect of sex and diet. * $p < 0.05$ versus control group, $ $p < 0.05$ versus males in the SFD group by Student's test.

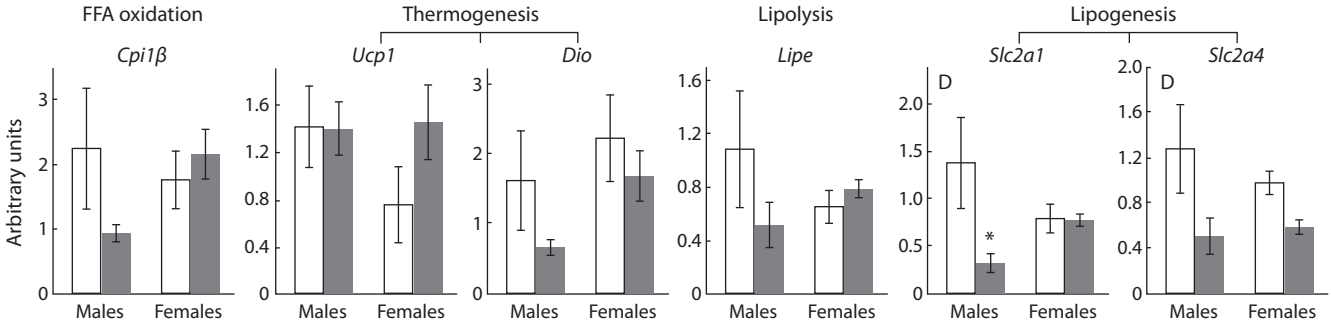


Fig. 5. The mRNA levels of BAT genes involved in lipid metabolism in mice, fed standard chow (control, white columns) and sweet-fat diet (grey columns).

Two-way ANOVA was used with the factors S, sex effect; D, diet effect; and D*S, interactive effect of sex and diet. * $p < 0.05$ versus control group by post-hoc Tukey test.

(see Fig. 4, *g–l*). SFD consumption did not affect the expression of most of the studied genes and only upregulated *Cpt1α* gene expression of ($P < 0.001$). In males, this increase was significant ($P < 0.01$ post-hoc Tukey test) and more pronounced (13 times) than in females (7 times).

There were no sex differences in the expression of the studied genes in BAT (Fig. 5). SFD did not affect the expression of genes involved in fat metabolism (*Cpt1β*, *Lipe*) and thermogenesis (*Ucp1*, *Dio2*); however, it down regulated the expression of genes that control glucose uptake into the cell – *Slc2a1* and *Slc2a4* ($P < 0.05$ for both genes).

Diet had a sex-independent effect on *Slc2a4* – and sex-dependent on *Slc2a* gene expression (interaction of factors $P = 0.06$, tendency). Diet reduced *Slc2a1* mRNA level by 4.5 times only in males, as a result, its expression in SFD males was 2.5 times less than in SFD females.

Discussion

A sweet-fat diet increases fat and carbohydrate proportion in food. To maintain a constant levels of blood lipids and carbohydrates, two ways of adaptation are possible: the deposition of fat excess and increased glucose and fatty acid oxidation in the liver, muscles and adipose tissues. Our results suggest that in male mice, both ways of adaptation were used and in female mice, the reservation of energy excess in the form of white fat prevailed. SFD males showed increased scWAT weight, although to a lesser extent than SFD females, and increased fatty acid oxidation in WAT and glucose in the liver. Only SFD males demonstrated increased expression of *Cpt1α* gene (a marker of fatty acid oxidation) in white adipose tissue and *Pklr* gene (a marker of glucose oxidation) in the liver. In addition, SFD males showed a more pronounced, than SFD females increase in "lean mass", which may indicate a

more intensive oxidation of metabolic substrates that occurs in the muscles, and, possibly, a greater infiltration of fat into muscle tissue.

The phenomenon of more intense fat accumulation in females than in males when fed high-energy diets was previously described in the literature (Priego et al., 2008; Medrikova et al., 2012; Chang et al., 2018). Several physiological mechanisms of this phenomenon have been proposed. First, estradiol is known to increase the number of adipocyte progenitor cells (Dieudonne et al., 2000); therefore, their number is higher in females than in males (Wu et al., 2017; Chang et al., 2018). Second, SFD increases the number of adipocyte progenitor cells only in females (Wu et al., 2017; Chang et al., 2018), but the reason for this is not known. Third, insulin sensitivity and lipogenesis are increased in white fat adipocytes in females compared to males (Macotela et al., 2009).

The data on pgWAT genes expressions obtained in our work complement the known mechanisms of intensive fat accumulation in females under sweet-fat diet consumption. In pgWAT, only in females, diet downregulated expression of genes, involved in lipid metabolism – *Lipe* (lipolysis) and *Lpl* (lipogenesis). Expression of the *Slc2a4* gene, which is also involved in lipogenesis, was reduced in SFD females to a much greater extent than in SFD males. Recently we demonstrated, that SFD reduced mRNA level of *Pparγ* (a transcription factor, the main regulator of adipocyte differentiation and function) in pgWAT, only in females (Bazhan et al., 2019). Together, these data suggest that a decrease in the expression of genes involved in the regulation of multidirectional processes in pgWAT, is an indicator of a decrease in the intensity of lipid metabolism, what can contribute to the conservation of energy in the form of white fat reserves in females.

SFD increased the *Cpt1α* gene expression in WAT of males, regardless of localization, which is consistent with the literature data (Warfel et al., 2017). The mechanism of selective activation of the *Cpt1α* gene expression in WAT of males fed high-energy diets is not known. In our work, increased expression of the *Cpt1α* gene in WAT of SFD males was associated with a multiple increase in the FGF21 blood level. Previously, we and other authors have shown that, selectively in males, FSD increased not only the blood FGF21 levels, but also its gene expression in the liver (Chukijrangroat et al., 2018; Gasparin et al., 2018; Bazhan et al., 2019). Apparently, the activation of the FGF21 system in males was much more pronounced than in females upon SFD consumption.

The liver is the main site of FGF21 synthesis, and adipose tissues are the main site of FGF21 action. In pharmacological and genetic studies, FGF21 has been shown to increase energy expenditure in WAT and BAT and insulin sensitivity at the whole body level (Xu et al., 2009; Zhang, Li, 2014). These effects may be due to FGF21 facilitates oxidative processes in WAT mitochondria (Chau et al., 2010), in particular by stimulating the expression of the *Cpt1α* gene (Coskun et al., 2008). It can be assumed that the increased *Cpt1α* gene expression in WAT of SFD males contributed to the increased fatty acid oxidation and prevented fat deposition. Therefore, pgWAT and scWAT weights in SFD males were significantly less than in SFD females.

The liver plays a crucial role in the regulation of energy homeostasis at the level of the whole body and is the main site of estradiol action in the regulation of insulin sensitivity. According to our results, it is also the central link in the implementation of various pathways of adaptation to SFD in male and female mice: the response to SFD of most studied hepatic genes was sex-dependent. The mRNA levels of the *Fasn*, *Pklr*, *G6pc*, and *Slc2a2* genes were increased or unchanged, relative to control, in SFD males, and were decreased in SFD females.

SFD males showed increased or unchanged expressions of *Fasn*, *Pklr*, *G6pc*, and *Slc2a2* genes, while SFD females showed decreased expressions of these hepatic genes. The same multidirectional dynamics of the transcriptional response to SFD were observed for other hepatic genes measured in our work, although the sex effect was not statistically significant. These results are in good agreement with the previously published data showing that only in male mice, SFD increases the hepatic expression of the peroxisome proliferator-activated receptor-α (PPARα), a transcription factor that enhances the expression of many hepatic genes involved in the regulation of carbohydrate-lipid metabolism (Gasparin et al., 2018; Bazhan et al., 2019; Sasaki et al., 2019). As a result, the expression of these genes (*Slc2a2*, *Gck*, *Pklr*, *G6P*, and *Pck1*) was lower in females than in males under SFD-induced obesity. Taken together, our data suggest that male mice respond to SFD with enhanced oxidation of glucose and fatty acids not only in WAT, but also in the liver.

It is possible that the mechanism of selective FGF21 activation in males with SFD-induced obesity was associated with hyperinsulinemia, which was revealed in our work and in the works of other authors, carried out on rodents consuming high-calorie diet (Rodríguez et al., 2003; Priego et al., 2008). An association was found between high plasma insulin and FGF21 levels in obese rodents and humans (Zhang et al., 2008; Chavez et al., 2009), the exact mechanism of which is unknown. It can be assumed that the increased blood FGF21 levels in SFD males counteracts the development of metabolic syndrome: FGF21 reduce body weight, normalize the lipid profile, and increase insulin sensitivity in various models of insulin resistance (Zhang, Li, 2014). In females, the SFD consumption caused a less pronounced than in males and insignificant increase in the blood insulin and FGF21 levels; apparently, FGF21 did not participate in adaptation to the SFD in females.

Our results showed that the SFD consumption stimulated the development of metabolic syndrome regardless of sex: obesity, increased blood glucose, insulin, cholesterol levels, hepatic TG content, and decreased glucose tolerance and insulin sensitivity. It should be noted that the SFD consumption disturbed different links in the regulation of blood glucose levels in males and females: satiated hyperinsulinemia was observed only in SFD males, and fasting hyperglycemia – only in SFD females. The mechanisms of sex-associated dysregulation of carbohydrate metabolism under obesity caused by a sweet-fat diet consumption are not known and need to be explored.

In BAT, in contrast to WAT, glucose enters the cells through Glut1 to the same extent as through Glut4 (Czech, 2020). The regulation of the expression of these genes and corresponding protein activity in BAT differs from that in WAT. *Slc2a4* gene expression is regulated by insulin (Burcelin et al., 1993), and gene expression and activity of the Glut1 protein are regulated

by norepinephrine through activation of beta 3 adrenoreceptors via a cAMP-dependent mechanism (Cannon, Needergaart, 2004). Our data demonstrated that in BAT, SFD consumption reduced the *Slc2a4* gene expression equally in males and females, and the *Slc2a1* gene expression only in males. The latter may be due to the effect of sex on the expression of beta3-adreno receptors under SFD consumption. The cafeteria diet has been shown to reduce the level of protein and the beta3-adreno receptor gene expression in BAT in male rats, but does not affect them in female rats (Rodríguez et al., 2001).

Glucose itself is not the dominant thermogenic substrate in BAT, it is converted into fatty acids, which oxidizing in the mitochondria, enhance thermogenesis (Cannon, Needergaart, 2004). It has been shown that obesity caused by long-term FSD consumption is associated with a decrease in energy consumption at the level of the whole body and with a decrease in thermogenesis at the level of BAT (Penna-de-Carvalho et al., 2014). It can be assumed that diet-induced decrease in the expression of glucose transporter genes in BAT will be accompanied by a decrease in thermogenesis, and this effect will be more pronounced in males than in females. This assumption is supported by data obtained earlier that in male mice, high-energy diets reduces in BAT, the expression of transcription factor Pparγ which stimulates the expression of target genes involved in the regulation of thermogenesis (Penna-de-Carvalho et al., 2014; Bazhan et al., 2019).

Conclusion

Thus, the results showed that in mice, adaptation to the consumption of SFD associated with the accumulation of excess white fat was observed both in males and females, but in females to a much greater extent than in males. In females, the diet down regulated the expression of hepatic and white adipose tissue genes involved in carbohydrate and fat metabolism, which could contribute to a decrease in energy expenditure and white fat accumulation. Only in males, adaptation to SFD, associated with enhanced oxidation of energy carriers in the liver and white fat, was observed, SFD males showed a significantly increased lean mass, blood insulin and FGF21 levels, and expressions of the *Cpt1α* genes in white fat tissues and *Pklr* in the liver. This suggests increased energy expenditure for fatty acid and glucose oxidation in WAT, muscle, and liver, and may inhibit the storage of energy in the form of white fat.

Adaptation ensure the maintenance of constant FFA and triglyceride blood levels, but led to the appearance of signs of insulin resistance (decreased insulin sensitivity, glucose tolerance, and increased TG levels in the liver) in males and females. Consumption of SFD disrupted different links in the regulation of insulin sensitivity in males and females: only in males, it caused satiated hyperinsulinemia and only in females – fasting hyperglycemia. The study of the sex characteristics of the molecular physiological mechanisms underlying adaptation to SFD in mice is a necessary step for the development of a gender-specific approach to the correction of metabolic disorders in humans.

References

Bazhan N., Jakovleva T., Balyibina N., Dubinina A., Denisova E., Feofanova N., Makarova E. Sex dimorphism in the Fgf21 gene expression in liver and adipose tissues is dependent on the metabolic condition. *Online J. Biol. Sci.* 2019;19(1):28-36. DOI 10.3844/ojbsci.2019.28.36.

Burcelin R., Kande J., Ricquier D., Girard J. Changes in uncoupling protein and GLUT4 glucose transporter expressions in interscapular brown adipose tissue of diabetic rats: relative roles of hyperglycaemia and hypoinsulinaemia. *Biochem. J.* 1993;291(Pt. 1(1)):109-113. DOI 10.1042/bj2910109.

Buyukdere Y., Gulec A., Akyol A. Cafeteria diet increased adiposity in comparison to high fat diet in young male rats. *PeerJ.* 2019;7(4):e6656. DOI 10.7717/peerj.6656.

Camporez J.P.G., Jornayvaz F.R., Petersen M.C., Pesta D., Guigni B.A., Serr J., Zhang D., Kahn M., Samuel V.T., Jurczak M.J., Shulman G.I. Cellular mechanisms by which FGF21 improves insulin sensitivity in male mice. *Endocrinol.* 2013;154(9):3099-3109. DOI 10.1210/en.2013-1191.

Cannon B., Nedergaard J. Brown adipose tissue: function and physiological significance. *Physiol. Rev.* 2004;84(1):277-359. DOI 10.1152/physrev.00015.2003.

Chang E., Varghese M., Singer K. Gender and sex differences in adipose tissue. *Curr. Diab. Rep.* 2018;18(9):69. DOI 10.1007/s11892-018-1031-3.

Chau M.D.L., Gao J., Yang Q., Wu Z., Gromada J. Fibroblast growth factor 21 regulates energy metabolism by activating the AMPK–SIRT1–PGC-1α pathway. *Proc. Natl. Acad. Sci. USA.* 2010;107(28):12553-12558. DOI 10.1073/pnas.1006962107.

Chavez A.O., Molina-Carrion M., Abdul-Ghani M.A., Folli F., Defronzo R.A., Tripathy D. Circulating fibroblast growth factor-21 is elevated in impaired glucose tolerance and type 2 diabetes and correlates with muscle and hepatic insulin resistance. *Diabetes Care.* 2009;32(8):1542-1546. DOI 10.2337/dc09-0684.

Chukijrungroat N., Khamphaya T., Weerachayaphorn J., Songserm T., Saengsirisuwan V. Hepatic FGF21 mediates sex differences in high-fat high-fructose diet-induced fatty liver. *Am. J. Physiol. Endocrinol. Metab.* 2017;313(2):E203-E212. DOI 10.1152/ajpendo.00076.2017.

Coskun T., Bina H.A., Schneider M.A., Dunbar J.D., Hu C.C., Chen Y., Moller D.E., Kharitonenkov A. Fibroblast growth factor 21 corrects obesity in mice. *Endocrinol.* 2008;149(12):6018-6027. DOI 10.1210/en.2008-0816.

Czech M.P. Mechanisms of insulin resistance related to white, beige, and brown adipocytes. *Mol. Metab.* 2020;3427-3442. DOI 10.1016/j.molmet.2019.12.014.

Dieudonne M.N., Pecquery R., Leneveu M.C., Giudicelli Y. Opposite effects of androgens and estrogens on adipogenesis in rat preadipocytes: Evidence for sex and site-related specificities and possible involvement of insulin-like growth factor 1 receptor and peroxisome proliferator-activated receptor γ2. *Endocrinol.* 2000;141(2):649-656. DOI 10.1210/endo.141.2.7293.

Fisher F.M., Chui P.C., Antonellis P.J., Bina H.A., Kharitonenkov A., Flier J.S., Maratos-Flier E. Obesity is a fibroblast growth factor 21 (FGF21)-resistant state. *Diabetes.* 2010;59(11):2781-2789. DOI 10.2337/db10-0193.

Gasparin F.R.S., Carreño F.O., Mewes J.M., Gilglioni E.H., Pagadigorria C.L.S., Natali M.R.M., Utsunomiya K.S., Constantin R.P., Ouchida A.T., Curti C., Gaemers I.C., Elferink R.P.J.O., Constantin J., Ishii-Iwamoto E.L. Sex differences in the development of hepatic steatosis in cafeteria diet-induced obesity in young mice. *Biochim. Biophys. Acta Mol. Basis Dis.* 2018;1864(7):2495-2509. DOI 10.1016/j.bbadis.2018.04.004.

Macotela Y., Boucher J., Tran T.T., Kahn C.R. Sex and depot differences in adipocyte insulin sensitivity and glucose metabolism. *Diabetes.* 2009;58(4):803-812. DOI 10.2337/db08-1054.

Mauvais-Jarvis F. Sex differences in metabolic homeostasis, diabetes, and obesity. *Biol. Sex Differ.* 2015;6(1):14. DOI 10.1186/s13293-015-0033-y.

Medrikova D., Jilkova Z.M., Bardova K., Janovska P., Rossmeisl M., Kopecky J. Sex differences during the course of diet-induced obesity in mice: adipose tissue expandability and glycemic control. *Int. J. Obes.* (*Lond.*). 2012;36(2):262-272. DOI 10.1038/ijo.2011.87.

Penna-de-Carvalho A., Graus-Nunes F., Rabelo-Andrade J., Mandarim-de-Lacerda C.A., Souza-Mello V. Enhanced pan-peroxisome proliferator-activated receptor gene and protein expression in adipose tissue of diet-induced obese mice treated with telmisartan. *Exp. Physiol.* 2014;99(12):1663-1678. DOI 10.1113/expphysiol.2014.081596.

Priego T., Sánchez J., Picó C., Palou A. Sex-differential expression of metabolism-related genes in response to a high-fat diet. *Obesity (Silver Spring).* 2008;16(4):819-826. DOI 10.1038/oby.2007.117.

Rodríguez E., Monjo M., Rodríguez-Cuenca, S., Pujol E., Amengual B., Roca P., Palou A. Sexual dimorphism in the adrenergic control of rat brown adipose tissue response to overfeeding. *Pflügers Arch*. 2001;442:396-403. DOI:10.1007/s004240100556.

Rodríguez A.M., Roca P., Bonet M.L., Picó C., Oliver P., Palou A. Positive correlation of skeletal muscle UCP3 mRNA levels with overweight in male, but not in female, rats. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 2003;285(4):R880-8. DOI 10.1152/ajpregu.00698.2002.

Sampey B.P., Vanhoose A.M., Winfield H.M., Freemerman A.J., Muehlbauer M.J., Fueger P.T., Newgard C.B., Makowski L. Cafeteria diet is a robust model of human metabolic syndrome with liver and adipose inflammation: comparison to high-fat diet. *Obesity.* 2011;19(6):1109–1117. DOI 10.1038/oby.2011.18.

Sasaki Y., Raza-Iqbal S., Tanaka T., Murakami K., Anai M., Osawa T., Matsumura Y., Sakai J., Kodama T. Gene expression profiles induced by a novel selective peroxisome proliferator-activated receptor α modulator (SPPARMα) pemafibrate. *Int. J. Mol. Sci.* 2019;20(22):5682. DOI 10.3390/ijms20225682.

Warfel J.D., Vandanmagsar B., Dubuisson O.S., Hodgeson S.M., Elks C.M., Ravussin E., Mynatt R.L. Examination of carnitine palmitoyltransferase 1 abundance in white adipose tissue: implications in obesity research. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 2017;312(5):R816-R820. DOI 10.1152/ajpregu.00520.2016.

Wu Y., Lee M.-J., Ido Y., Fried S.K. High-fat diet-induced obesity regulates MMP3 to modulate depot- and sex-dependent adipose expansion in C57BL/6J mice. *Am. J. Physiol. Endocrinol. Metab.* 2017;312(1):E58-E71. DOI 10.1152/ajpendo.00128.2016.

Xu J., Stanislaus S., Chinookoswong N., Lau Y.Y., Hager T., Patel J., Ge H., Weiszmann J., Lu S.-C., Graham M., Busby J., Hecht R., Li Y.-S., Li Y., Lindberg R., Véniant M.M. Acute glucose-lowering and insulin-sensitizing action of FGF21 in insulin-resistant mouse models-association with liver and adipose tissue effects. *Am. J. Physiol. Endocrinol. Metab.* 2009;297(5):E1105-E1114. DOI 10.1152/ajpendo.00348.2009.

Zhang J., Li Y. Fibroblast growth factor 21, the endocrine FGF pathway and novel treatments for metabolic syndrome. *Drug Discov. Today.* 2014;19(5):579-589. DOI 10.1016/j.drudis.2013.10.021.

Zhang X., Yeung D.C.Y., Karpisek M., Stejskal D., Zhou Z.-G., Liu F., Wong R.L.C., Chow W.-S., Tso A.W.K., Lam K.S.L., Xu A. Serum FGF21 levels are increased in obesity and are independently associated with the metabolic syndrome in humans. *Diabetes.* 2008;57(5):1246-1253. DOI 10.2337/db07-1476.

ORCID ID

N.M. Bazhan orcid.org/0000-0002-7246-4758
T.V. Iakovleva orcid.org/0000-0001-7628-5856
A.D. Dubinina orcid.org/0000-0003-3907-3488
E.N. Makarova orcid.org/0000-0002-6417-9893

Acknowledgements. This work was supported by the Russian Science Foundation, grant No. 17-15-01036-П. The studies are implemented using the equipment of the Center for Genetic Resources of Laboratory Animals at ICG SB RAS, supported by the Ministry of Education and Science of Russia (Unique identifier of the project RFMEFI62119X0023).

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

Received August 24, 2020. Revised September 22, 2020. Accepted October 2, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):853-860
DOI 10.18699/VJ20.683

Genetic markers for the resistance of honey bee to *Varroa destructor*

M.D. Kaskinova ✉, L.R. Gaifullina, E.S. Saltykova, A.V. Poskryakov, A.G. Nikolenko

Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Center of the Russian Academy of Sciences, Ufa, Russia
✉ e-mail: kaskinovamilyausha@mail.ru

Abstract. In the mid-20th century, the first case of infection of European bees *Apis mellifera* L. with the ectoparasite mite *Varroa destructor* was recorded. The original host of this mite is the Asian bee *Apis cerana*. The mite *V. destructor* was widespread throughout Europe, North and South America, and Australia remained the only continent free from this parasite. Without acaricide treatment any honeybee colony dies within 1–4 years. The use of synthetic acaricides has not justified itself – they make beekeeping products unsuitable and mites develop resistance to them, which forces the use of even greater concentrations that can be toxic to the bees. Therefore, the only safe measure to combat the mite is the use of biological control methods. One of these methods is the selection of bee colonies with natural mite resistance. In this article we summarize publications devoted to the search for genetic markers associated with resistance to *V. destructor*. The first part discusses the basic mechanisms of bee resistance (*Varroa* sensitive hygienic behavior and grooming) and methods for their assessment. The second part focuses on research aimed at searching for loci and candidate genes associated with resistance to varroosis by mapping quantitative traits loci and genome-wide association studies. The third part summarizes studies of the transcriptome profile of *Varroa* resistant bees. The last part discusses the most likely candidate genes – potential markers for breeding *Varroa* resistant bees. Resistance to the mite is manifested in a variety of phenotypes and is under polygenic control. The establishing of gene pathways involved in resistance to *Varroa* will help create a methodological basis for the selection of *Varroa* resistant honeybee colonies.

Key words: *Apis mellifera*; *Varroa destructor*; *Varroa* resistance; marker-assisted selection.

For citation: Kaskinova M.D., Gaifullina L.R., Saltykova E.S., Poskryakov A.V., Nikolenko A.G. Genetic markers for the resistance of honey bee to *Varroa destructor*. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):853-860. DOI 10.18699/VJ20.683

Генетические маркеры резистентности медоносной пчелы к *Varroa destructor*

М.Д. Каскинова ✉, Л.Р. Гайфуллина, Е.С. Салтыкова, А.В. Поскряков, А.Г. Николенко

Институт биохимии и генетики – обособленное структурное подразделение Уфимского федерального исследовательского центра Российской академии наук, Уфа, Россия
✉ e-mail: kaskinovamilyausha@mail.ru

Аннотация. В середине XX в. был зафиксирован первый случай заражения европейских пчел *Apis mellifera* L. клещом-эктопаразитом *Varroa destructor*, изначальным хозяином которого является азиатская пчела *Apis cerana*. Клещ распространился по всей Европе, Северной и Южной Америке, и единственным континентом, свободным от этого паразита, осталась Австралия. Без обработки акарицидами семья медоносной пчелы погибает в течение 1–4 лет. Использование синтетических акарицидов не оправдало себя – они делают непригодными продукты пчеловодства и у клещей возникает к ним резистентность, что заставляет использовать еще большие концентрации препаратов, которые могут быть токсичны для пчел. Единственная безопасная мера борьбы – использование методов биологического контроля. Одним из таких методов является селекция семей пчел, обладающих естественной резистентностью к клещу. В обзоре обобщены публикации, посвященные поиску генетических маркеров, ассоциированных с устойчивостью к *V. destructor*. Рассматриваются основные механизмы устойчивости пчел к клещу (Varroa-чувствительное гигиеническое поведение и груминг) и методы их оценки. Обсуждаются исследования, направленные на поиск локусов и генов-кандидатов, ассоциированных с устойчивостью к варроатозу, при помощи картирования локусов количественных признаков и полногеномного поиска ассоциаций. Обобщены исследования транскриптомного профиля Varroa-устойчивых пчел. Рассмотрены наиболее вероятные гены-кандидаты – потенциальные маркеры для селекции Varroa-резистентных

пчел. Резистентность к клещу проявляется в виде разнообразных фенотипов и находится под полигенным контролем. Установление генных путей, задействованных в механизме резистентности к *Varroa*, поможет создать методологическую базу для селекции устойчивых к варроатозу семей *A. mellifera*.
Ключевые слова: *Apis mellifera*; *Varroa destructor*; Varroa-резистентность; маркер-опосредованная селекция.

Introduction

The *Varroa destructor* Anderson & Trueman, 2000 is the most widespread and most harmful pest of bees (Anderson, Trueman, 2000; Martin et al., 2012). Review articles devoted to *V. destructor* deal with various aspects of its biology (Calderon et al., 2010; Rosenkranz et al., 2010; Nazzi et al., 2016; Evans, Cook, 2018), ways to mite control (Chandler et al., 2001; Dietemann et al., 2012; Kamler et al., 2016; Plettner et al., 2017), issues of bee resistance to mite and hygienic behavior (Zakar et al., 2014; Kurze et al., 2016; Locke, 2016a; Leclercq et al., 2017).

The invasion of *Varroa* has become a challenge for the European bee, since it has not developed the natural defense mechanisms that well developed in the original host of the mite – the Asian bee *Apis cerana*. The resistance of the Asian bee to the mite is due to the fact that it has well-developed behavioral defense mechanisms and the mite parasitizes mainly on drone brood (Pritchard, 2016). The currently known methods of fighting *V. destructor* are based on the use of synthetic acaricides and biological control methods (Dietemann et al., 2012; Kamler et al., 2016; Plettner et al., 2017). The problem of acaricides accumulation in beekeeping products and the development of acaricides resistance in the mite make beekeepers refuse to use them. Therefore, biological control methods are of great importance, one of which is the selection of bees that have resistance toward the *Varroa* mites.

The purpose of this review is to summarize the materials of experimental studies devoted to the establishment of the genetic basis of honey bee resistance to the *V. destructor*.

Mechanisms of resistance to varroatosis

There are two main phenotypes associated with resistance to mite: *Varroa* sensitive hygiene behavior and grooming, which includes auto-grooming (self-cleaning) and allogrooming (cleaning the body of another member of colony).

Before considering the concept of *Varroa* sensitive hygienic behavior, let's get acquainted with such a mechanism of protecting bees from brood diseases as hygienic behavior. In 1964 the brood removal behavior of bees infected with American foulbrood was described (Rothenbuhler, 1964). This behavior, called hygienic, consisted of the following actions – detecting, uncapping and removing the infected brood. About twenty years later, Gilliam et al. (1983) showed that hygienic behavior is also effective against ascospherosis. In 1993, the breeding program for honey bee colonies with a high level of hygienic behavior has been started in the University of Minnesota (Spivak, 1996). It was found that hygienic behavior is performed by 15–17 days old bees (Arathi et al., 2003). Bees remove fifth instar larvae infected with the bacterium *Paenibacillus larvae* (caused American foulbrood) and the fungus *Ascosphaera apis* (causative agent of ascospherosis) before the pathogens reach the sporulation stage (Spivak, Reuter, 2001; Albo et al., 2017).

In 1997, the Suppression of Mite Reproduction (SMR) phenomenon was described: bee colonies with this phenotype have a low number of reproductively successful female mites (Harbo, Harris, 1999). It soon became clear that SMR is a consequence of specific hygienic behavior aimed at removing a mite, which has offspring. It is known that the foundress mite, after penetrating into an unsealed cell with a bee larva, begins to lay eggs only 3 days after the cell is sealed (Spivak, 1996; Harbo, Harris, 2005; Harris, 2007; Harris et al., 2010; Rosenkranz et al., 2010). The detection and removal of the cells content with mite offspring leads to a reduction in the total number of mites in the bee colony. This type of behavior has been termed *Varroa* sensitive hygiene (VSH) (Harbo, Harris, 2005).

To assess hygienic behavior, two tests have been developed and are widely used, – freeze-killed brood assay, FKB (Spivak, 1996; Facchini et al., 2019) and pin-killed brood assay, PKB (Gramacho et al., 1999). These tests are often used in experimental studies to analyze resistance toward the *Varroa* mite, so we will consider a short protocol for their implementation. The brood combs are frozen (FKB) or killed with a pin (PKB) and introduced into the test colony for 24 hours. If colony removes more than 95 % of the killed brood it is considered highly hygienic. VSH assessment is more complex: a section of combs with sealed brood infested with mites is introduced into the test colony and after a week the percentage of uncapped and cleaned cells and other indicators are calculated (Villa et al., 2009). FKB assay was developed to assess hygiene behavior, however Danka et al. (2013) reported that colonies bred for VSH remove frozen brood faster (in 6–12 hours) than colonies bred for FKB assay. At the same time, colonies selected using FKB assay do not cope with the test developed to assess VSH phenotype. Therefore, FKB assay can be used to test VSH phenotype, but this fact requires additional verification.

Grooming behavior is another natural defense mechanism of bees, which consists in the ability of bees to clean themselves (auto-grooming) or other bees (allogrooming) from external parasites and pollution (Boecking, Spivak, 1999; Land, Seeley, 2004). It is strongly expressed in *A. cerana* (Fries et al., 1996). This is especially true for allogrooming: if an Asian bee cannot remove a mite by itself, it performs a special dance that provokes other bees to perform allogrooming (Land, Seeley, 2004). There are also a difference in grooming between *A. mellifera* subspecies. For example, Africanized bees remove mites more intensively than European subspecies (Invernizzi et al., 2015). Colonies are assessed for this feature both at the individual (Aumeier, 2001) and colony level (Bienefeld, 1999).

In addition, populations of *A. mellifera* were identified that survived and coexist with *V. destructor* for a long time. Evaluation of such colonies showed that they have a high level of *Varroa* sensitive and grooming behavior (Locke, 2016b). On the basis of genomic and transcriptome studies, loci and genes associated with *Varroa* resistance were identified.

Mapping of loci and genes associated with *Varroa* resistance

Oxley et al. (2010) identified the Hyg1 locus on chromosome 2 associated with hygienic behavior. The 95 % confidence interval of this locus (see Darvasi, Soller, 1997) included genes associated with behavior, smell, development and functioning of neurons, receptor and transcriptional activity. Harpur et al. (2019) based on genome-wide sequencing of drones from two apiaries selected for hygienic behavior and one non-selected apiary identified 73 candidate genes. 49 of them were located near previously identified loci (Oxley et al., 2010; Tsuruda et al., 2012). Of great interest are the *abscam*, *goosecoid* (*Hox*-gene) and *tropomysin-2-like* genes on chromosome 6, the ortholog of the Drosophila *dyschronic* gene (*GB45054*) on chromosome 11, and the insulin-like receptor (*GB53353*) on chromosome 9. *Abscam* is known to play an important role in axonal guidance, in particular of olfactory neurons. The *goosecoid* and *tropomysin-2-like* genes are also essential for the development of the nervous system. The *GB45054* gene is involved in biological processes such as sensory perception of sounds and light stimuli. *GB53353* is involved in protein phosphorylation and the transmembrane receptor protein tyrosine kinase signaling pathway. Kim et al. (2019) performed genome-wide sequencing of *A. m. caucasica* with high hygienic behavior and *A. m. carnica* with a low level of hygiene. They obtained 20 SNP markers associated with hygienic behavior, and candidate genes were identified for three of them. SNP1 is located in the *twitchin* (chromosome 2), in the previously identified locus Hyg1 (Oxley et al., 2010). SNP8 and SNP9 are located in the gene encoding a peroxidase-like protein (chromosome 4).

In studies (Oxley et al., 2010; Harpur et al., 2019; Kim et al., 2019) hygiene behavior was assessed using FKB assay, and, as it was said, colonies selected on FKB do not always successfully cope with a mite. However, given that *Varroa* sensitive and general hygienic behavior are based on the same mechanism (detecting and uncapping diseased brood), results obtained by these authors should not be excluded from further consideration.

Genome-wide analysis of VSH was carried out by research groups from the USA (Tsuruda et al., 2012) and Germany (Spotter et al., 2012, 2016). Tsuruda et al. (2012) identified a locus on chromosome 9 associated with VSH phenotype. This locus contains the *NorpA2* gene (homologue of the *D. melanogaster NorpA*) and the dopamine receptor *Dop3*. *NorpA2*, encoding phospholipase C, is associated with learning and memory formation in the honey bee (Suenami et al., 2018). Whereas dopamine plays a critical role in the formation of aversive memory in insects (Beggs, Mercer, 2009).

Spotter et al. (2012) analyzed three samples of bees with different levels of VSH and developed a differentiating panel of 44,000 SNPs. In next study (Spotter et al., 2016) they identified 6 SNPs associated with resistance towards the *V. destructor*. For four of them, candidate genes were proposed: *AdoR*, *Cdk5alpha*, *Octbeta2R*, and *Obp1*. The identified SNPs are not located in candidate genes themselves, but are localized near them. Therefore, their role in the formation of VSH phenotype has yet to be proven. The authors substantiated the choice of these candidate genes by their function. Adenosine receptors (encoded by the *AdoR* gene) belong to the family of G protein-coupled receptors and are involved in extracellular adenosine signaling. Adenosine is an important regulator of the nervous system; it is involved in the modulation of synaptic plasticity (Dolezelova et al., 2007). *Cdk5alpha* encodes an activator of the cyclin-dependent kinase gene *Cdk5*. *Cdk5* regulates many cellular processes (neuronal migration, axon guidance, ensuring the stability of microtubules and synapses, etc.), and it has been shown that in the Asian bee *A. cerana Cdk5*, together with its activator gene, is involved in the cell response to oxidative stress (Zhao et al., 2018). The biogenic amine octopamine is an important neurotransmitter, modulator and hormone in invertebrates. It was shown that the octopamine receptor gene *Octbeta2R* plays an important role in the formation of adaptations in the high-mountain population of *A. m. monticola* (Wallberg et al., 2017). *Obp1*, expressed in the antennae of worker bees, is responsible for the perception of queen pheromones (Lartigue et al., 2004), and probably for the perception of other olfactory signals.

In addition to the colonies selected for hygiene, there are populations that coexisted with *V. destructor* for a long time without acaricide treatment (in review Locke, 2016b). These populations have become the object of close scrutiny by geneticists. Behrens et al. (2011) analyzed offspring of two hybrid queens from a *Varroa* tolerant colony from the Gotland population. They uncapped the sealed drone brood and estimated the number of mites with and without offspring. Colonies with mites without offspring were considered as resistant. Using 488 SSR markers for mapping, they identified a locus on chromosome 7 associated with this phenotype. This locus contains two important candidate genes, orthologs of *D. melanogaster* genes, *foxo* (*GB11764*, a transcription factor in the insulin signaling pathway) and *futsch* (*GB11509*, induces synaptic plasticity in neurons). Lattorff et al. (2015) based on data from Behrens et al. (2011) also analyzed bee colonies from the Gotland population. They compared colonies before (2000) and after (2007) selection using 39 SSR markers on chromosomes 4 and 7. 11 candidate genes were identified on chromosome 7, including 10 protein-coding genes and one gene of long non-coding RNA, the target of which is unknown. The authors propose the oxidoreductase gene *GMCOX18* as a promising candidate gene. Oxidoreductases are involved in glucose metabolism and cuticle biosynthesis. Therefore, the authors hypothesized that the *GMCOX18* may play a role in altering substances secreted by bee larvae, which are required to trigger oogenesis in a mite.

Among the genetic markers found in *Varroa* tolerant colonies using SNP mapping (Conlon et al., 2018) candidate genes involved in the synthesis of ecdysone are distinguished. It is known that *V. destructor* cannot synthesize ecdysone itself

and receives it from bees. Ecdysone is necessary for mite to activate the reproductive cycle, while in insects it initiates molting and metamorphosis. Conlon et al. (2018) performed genome-wide sequencing of drones from *Varroa* tolerant colonies from Sweden and identified a locus on chromosome 15 associated with tolerance to the mite. This locus includes three genes involved in ecdysone synthesis: *Mblk-1*, *Cyp18a11* and *Phantom*. They continued their research by performing genome-wide sequencing of drones from another *Varroa* tolerant population, the Toulouse population from France (Conlon et al., 2019). As a result, 9 SNPs associated with *Varroa* olerance were identified, and three of them were located in the transcription factor *Mblk-1*.

A search was also carried out for genes associated with the grooming behavior of bees. Arechavaleta-Velasco et al. (2012) identified a locus on chromosome 5 and named it "groom-1". It includes 27 candidate genes, three of which (*Atlastin*, *Ataxin*, *AmNrx1*) are associated with the development of the nervous system and behavior.

Transcriptome analysis of *Varroa* resistance

After the decoding of the honeybee's genome, studies of its transcriptome were initiated. Differential gene expression analysis is often used to find candidate genes. It allows finding out how the activity of certain genes can affect the mechanisms of resistance.

A comparative analysis of the transcriptome profile of colonies with high and low levels of hygiene behavior (Boutin et al., 2015) revealed 28 genes with increased expression in the former. Most of them were located at previously identified loci (Oxley et al., 2010; Spotter et al., 2012; Tsuruda et al., 2012). Of great interest as markers are genes of cytochrome P450 gene superfamily (*Cyp4AZ1*, *Cyp4g11*, *Cyp6AS11*, *Cyp6AS8*), which are over-expressed in non-hygienic bees. Cytochrome P450 enzymes degrade odorant and pheromone molecules (Feyereisen, 1999), thereby reducing the ability of bees to detect infected brood.

Transcriptomic analysis of colonies with VSH phenotype was performed by two groups (Le Conte et al., 2011; Mondet et al., 2015). Le Conte et al. (2011) identified 39 differentially expressed transcripts in the brains of bees with VSH phenotype compared to control bees without VSH. Among the genes with increased expression in the brain of VSH bees, the authors emphasize *PRL-1*, which encodes tyrosine phosphatase, and *GB16747*. It was later shown that the expression of the *GB16747*, involved in the metabolism of ascorbate/aldarate, increases in response to infection with *V. destructor* (McDonnell et al., 2013). The *Cyp4g11* and *Obp3* genes and three exons of the *Dscam* were under-expressed.

Mondet et al. (2015) found 258 differentially expressed transcripts in the antennae of worker bees with and without VSH phenotype. Among genes involved in redox metabolism 12 genes were over-expressed and 3 genes were under-expressed in bees with VSH. Four genes that control the immune response, in particular the *Def1* and *Def2*, were under-expressed. Of particular interest are genes associated with olfaction (*Obp3*, *Trh*, *OR85b-like*, *CSP2*, *NT-7*, *Obp14*, et al.). Proteomic studies have also shown the involvement of the *Obp* genes (*Obp17* and *Obp18*) in the formation of VSH phenotype (Hu et al., 2016). Differential expression of the *Obp* genes indicates that the olfaction plays an important role in VSH.

Analysis of two susceptible and two tolerant colonies (Navajas et al., 2008) showed that mite-tolerant bees undergo changes in the expression of genes that regulate the neurons development and sensitivity, as well as the olfaction (orthologs of *D. melanogaster* genes, which are over-expressed: *poe*, *GluClα*, *para*, *Dhc64c*, and which are under-expressed: *futsch*, *scratch*, *fringe*, *Dscam*, etc.). Colonies were used as tolerant if they had not been treated with acaricides for 11 years and had a low level of mite infestation (the authors counted mites at the hive bottom 4 times a year for 5 years). In susceptible colonies the level of mite infestation was 10 times higher.

Jiang et al. (2016), comparing transcriptome profiles of a *V. destructor* tolerant colony that survived without acaricide treatment for 58 months and a susceptible colony that died from varroatosis within 17 months, identified 6 candidate genes. Of these, 4 encode proteins of cytochrome P450. The *Cyp6AS12* and *Cyp6BE1* genes were over-expressed in pupae of the tolerant mite-infested colony. *Cyp6BE1* and *Cyp9Q3* were over-expressed in adults from a tolerant mite-free colony relative to the same mite-infested colony, whereas at the pupal stage there were no significant differences in expression levels of the two genes.

Conlon et al. (2019) measured the expression of *Mblk-1*, *Cyp18a11*, and *Phantom* genes in workers and drones larvae from *Varroa* tolerant colonies to verify the results of genome wide analysis. The expression pattern of genes involved in ecdysone biosynthesis (in particular, the transcription factor *Mblk-1*) differed in drone larvae and worker larvae. If a mutation occurs in the genes responsible for the ecdysone synthesis, this can lead to a malfunction of the mite development cycle. It is possible that the preference of the drone brood by the mite and its more successful reproduction in it is a consequence of the differences in the level of ecdysone expression in the drone and bee brood.

Transcriptome analysis confirmed the contribution of the *neurexin I* gene (Arechavaleta-Velasco et al., 2012) to grooming behavior. In colonies with a high level of grooming behavior, the expression of this gene was increased (Hamiduzzaman et al., 2017).

Transcriptomic studies were also performed for the Asian bee *A. cerana*. Ji et al. (2014) compared the transcriptomes of nurse bees of *A. cerana* before and after infection with *V. destructor* (after 24 hours). Among genes whose expression increased in response to mite infection were genes associated with olfaction (*Obp4*, *Obp17*, *Obp18*, *Dscam)*, as well as transcription factors (*CREB-like 2-like* and *Mblk-1*). Diao et al. (2018) showed that *A. cerana* has more immune genes and genes encoding antimicrobial peptides than *A. mellifera*. However, *A. cerana* has fewer genes encoding odorant-binding proteins (Obp) and olfactory receptors. This suggests that after the divergence, the European bee lost some of its genes due to the lack of *Varroa* pressure, and when faced with it,

Overlapping candidate genes associated with *Varroa destructor* resistance

Gene candidate (chromosome)	Functional category	Phenotype	Reference
Cyp4g11 (16)	Electron carrier activity; steroid biosynthetic process	VSH	Spotter et al., 2012
			Le Conte et al., 2011
		FKB	Boutin et al., 2015
		VSH	Mondet et al., 2015
Obp3 (9)	Odorant binding; sensory perception of chemical stimulus	VSH	Le Conte et al., 2011
			Mondet et al., 2015
Obp4 (9)	Odorant binding; sensory perception of chemical stimulus	VSH	Tsuruda et al., 2012
		FKB	Boutin et al., 2015
Dscam (4)	Axon guidance; mushroom body development	*Varroa* tolerance	Navajas et al., 2008
		VSH	Le Conte et al., 2011
Neurexin I (5)	Growth, maintenance and maturation of synapses in the brain	Grooming	Arechavaleta-Velasco et al., 2012; Hamiduzzaman et al., 2017

A. mellifera activated other mechanisms. Differences in the methylation levels of genes responsible for learning and memory were also recorded. The formation of long-term memory and synaptic plasticity requires activation of neuronal signaling pathways. Transcriptome analysis showed that, in *A. cerana*, the expression of genes involved in signaling pathways (cAMP-PKA, MAPK, and CaMK IV) increases in response to mite infection.

Genetic markers of bee resistance to varroatosis

The above mentioned studies narrowed down the list of potential loci and candidate genes that determine the resistance of bees to the *Varroa* mite. Each of the studies identified its own candidate genes. Overlaps (coincidences of results) were obtained mainly for those studies in which the same methods for assessing the resistant phenotype were used (see the Table).

In studies of the transcriptome profile of the brain (Le Conte et al., 2011) and antennae (Mondet et al., 2015) of VSH bees, a common candidate gene *Cyp4g11* was identified. In the brain, *Cyp4g11* expression was decreased, and in the antennae it was increased. Decreased expression of the *Cyp4g11* gene in the brain was also shown for colonies with high hygiene behavior tested with FKB (Boutin et al., 2015). In addition, the *Cyp4g11* gene is located at one of the loci previously identified in VSH bees (Spotter et al., 2012). It is currently unknown what function *Cyp4g11* performs in the honey bee organism. Cytochrome P450 genes are involved in ecdysteroids metabolism, detoxification of xenobiotics and destruction of odorant molecules (Feyereisen, 1999).

A common candidate gene *Obp3* was identified for VSH colonies (in two studies independently). In VSH bees, the expression of this gene is increased in antennas (Mondet et al., 2015), while in the brain it is decreased (Le Conte et al., 2011). For one more gene from the *Obp* family, an overlap was found: the *Obp4*, which are under-expressed in the brain of bees selected for FKB (Boutin et al., 2015), is located at one of the loci on chromosome 9, identified earlier (Tsuruda et al., 2012).

The overlap was also shown for *Varroa* tolerant (Navajas et al., 2008) and VSH (Le Conte et al., 2011) colonies. *Dscam* expression (*GB15141*) was under-expressed in *Varroa* tolerant bees (Navajas et al., 2008). In a study (Le Conte et al., 2011), three exons of the *Dscam* gene also were under-expressed.

Common candidate genes with the Asian bee *A. cerana* were also identified. The resistance of the Asian bee to *Varroa* is the key to understanding the resistance of the European bee. The presence of overlapping genes such as *Mblk-1*, *Dscam*, and *Obp4* (Ji et al., 2014) confirms this. Further research is needed to establish role of these genes in the mechanism of *Varroa* resistance in bees.

Conclusion

Genomic and transcriptome studies have shown that genes associated with visual and olfactory perception, development and functioning of the nervous system (learning and memory formation) play the main role in *Varroa* sensitive hygiene behavior. Receptor genes are of great interest, most of which belong to the family of G protein-coupled receptors (dopamine, adenosine, and octopamine receptors). Some of the identified candidate genes can be successfully used as markers for the selection of specific subspecies or lines of bees for which they were obtained (Haddad et al., 2015; Kim et al., 2019), some of genes needs testing on other populations (Le Conte et al., 2011; Boutin et al., 2015; Mondet et al., 2015; Spotter et al., 2016; Hamiduzzaman et al., 2017). Some candidate genes are associated with a general immune response (Le Conte et al., 2011; Jiang et al., 2016). Further study of some genes (Ji et al., 2014; Lattorff et al., 2015; Conlon et al., 2019), for example, genes for ecdysone biosynthesis, will help to shed light on the nature of the parasite–host relationship, in particular the

question of why the mite in the original host reproduces more successfully on the drone brood. Do not forget that the mite is a parasite, and, like many parasites, some of its life support systems are reduced. Finding these pain points of the *Varroa* mite can also help fight varroatosis.

The resistance of the honey bee to the *V. destructor* mite is under polygenic control. The European bee was able to use other gene pathways to provide its defense against the *V. destructor*, despite the short period of time since the mite invasion. Establishing these pathways will help create a methodological basis for breeding *Varroa* resistant *A. mellifera* colonies.

References

Albo G.N., Cordoba S.B., Reynaldi F.J. Chalkbrood: pathogenesis and the interaction with honeybee defenses. *Int. J. Envir. Agric. Res.* 2017;3(1):71-80.

Anderson D.L., Trueman J.W.H. *Varroa jacobsoni* (Acari: Varroidae) is more than one species. *Exp. Appl. Acarol.* 2000;24:165-189. DOI 10.1023/A:1006456720416.

Arathi S., Burns I., Spivak M. Ethology of hygienic behaviour in the honey bee *Apis mellifera* L. (Hymenoptera: Apidae): behavioural repertoire of hygienic bees. *Ethology.* 2003;106:365-379. DOI 10.1046/j.1439-0310.2000.00556.x.

Arechavaleta-Velasco M.E., Alcala-Escamilla K., Robles-Rios C., Tsuruda J.M., Hunt G.J. Fine-scale lingkage mapping reveals a small set of candidate genes influencing honey bee grooming behavior in response to Varroa mites. *PLoS One.* 2012;7:e47269. DOI 10.1371/journal.pone.0047269.

Aumeier P. Bioassay for grooming effectiveness towards *Varroa destructor* mites in Africanized and Carniolan honey bees. *Apidologie.* 2001;32:81-90. DOI 10.1051/apido:2001113.

Beggs K.T., Mercer A.R. Dopamine receptor activation by honey bee queen pheromone. *Curr. Biol.* 2009;19:1206-1209. DOI 10.1016/j.cub.2009.05.051.

Behrens D., Huang Q., Geßner C., Rosenkranz P., Frey E., Locke B., Moritz R.F.A., Kraus F.B. Three QTL in the honey bee *Apis mellifera* L. suppress reproduction of the parasitic mite *Varroa destructor*. *Ecol. Evol.* 2011;1(4):451-458. DOI 10.1002/ece3.17.

Bienefeld K. Recording the proportion of damaged *Varroa jacobsoni* Oud. in the debris of honey bee colonies (*Apis mellifera*). *Apidologie.* 1999;30:249-256.

Boecking O., Spivak M. Behavioral defenses of honey bees against *Varroa jacobsoni* Oud. *Apidologie.* 1999;30:141-158. DOI 10.1051/apido:19990205.

Boutin S., Alburaki M., Mercier P.-L., Giovenazzo P., Derome N. Differential gene expression between hygienic and non-hygienic honeybee (*Apis mellifera* L.) hives. *BMC Genom.* 2015;16:500. DOI 10.1186/s12864-015-1714-y.

Calderon R.A., van Veen J.W., Sommeijer M.J., Sanchez L.A. Reproductive biology of *Varroa destructor* in Africanized honey bees (*Apis mellifera*). *Exp. Appl. Acarol.* 2010;50(4):281-297. DOI 10.1007/s10493-009-9325-4.

Chandler D., Sunderland K.D., Ball B.V., Davidson G. Prospective biological control agents of *Varroa destructor* n. sp., an important pest of the European honeybee, *Apis mellifera*. *Biocontrol Sci. Technol.* 2001;11(4):429-448. DOI 10.1080/09583150120067472.

Conlon B.H., Aurori A., Giurgiu A.I., Kefuss J., Dezmirean D.S., Moritz R.F.A., Routtu J. A gene for resistance to the *Varroa* mite (Acari) in honey bee (*Apis mellifera*) pupae. *Mol. Ecol.* 2019;28(12):2958-2966. DOI 10.1111/mec.15080.

Conlon B.H., Frey E., Rosenkranz P., Locke B., Moritz R.F.A., Routtu J. The role of epistatic interactions underpinning resistance to parasitic *Varroa* mites in haploid honey bee (*Apis mellifera*) drones. *J. Evol. Biol.* 2018;31:801-809. DOI 10.1111/jeb.13271.

Danka R.G., Harris J.W., Villa J.D., Dodds G. Varying congruence of hygienic responses to *Varroa destructor* and freeze-killed brood among different types of honeybees. *Apidologie.* 2013;44:447-457. DOI 10.1007/s13592-013-0195-8.

Darvasi A., Soller M. A simple method to calculate resolving power and confidence interval of QTL map location. *Behav. Genet.* 1997;27:125-132. DOI 10.1023/A:1025685324830.

Diao Q., Sun L., Zheng H., Zeng Z., Wang S., Xu S., Zheng H., Chen Y., Shi Y., Wang Y., Meng F., Sang Q., Cao L., Liu F., Zhu Y., Li W., Li Z., Dai C., Yang M., Chen S., Chen R., Zhang S., Evans J.D., Huang Q., Liu J., Hu F., Su S., Wu J. Genomic and transcriptomic analysis of the Asian honeybee *Apis cerana* provides novel insights into honeybee biology. *Sci. Rep.* 2018;8:822. DOI 10.1038/s41598-017-17338-6.

Dietemann V., Pflugfelder J., Anderson D., Charriere J.D., Chejanovsky N., Dainat B., de Miranda J.R., Delaplane K.S., Dillier F.-X., Fuch S., Gallmann P., Gauthier L., Imdorf A., Koeniger N., Kralj J., Meikle W.G., Pettis J.S., Rosenkranz P., Sammataro D., Smith D.R., Yañez O., Neumann P.J. *Varroa destructor*: research avenues towards sustainable control. *J. Apic. Res.* 2012;51(1):125-132. DOI 10.3896/IBRA.1.51.1.15.

Dolezelova E., Nothacker H.-P., Civelli O., Bryant P.J., Zurovec M. Drosophila adenosine receptor activates cAMP and calcium signaling. *Insect Biochem. Mol. Biol.* 2007;37(4):318-329. DOI 10.1016/j.ibmb.2006.12.003.

Evans J.D., Cook S.C. Genetics and physiology of *Varroa* mites. *Curr. Opin. Insect Sci.* 2018;26:130-135. DOI 10.1016/j.cois.2018.02.005.

Facchini E., Bijma P., Pagnacco G., Rizzi R., Brascamp E.W. Hygienic behaviour in honeybees: a comparison of two recording methods and estimation of genetic parameters. *Apidologie.* 2019;50:163-172. DOI 10.1007/s13592-018-0627-6.

Feyereisen R. Insect p450 enzymes. *Annu. Rev. Entomol.* 1999;44:507-533. DOI 10.1146/annurev.ento.44.1.507.

Fries I., Huazhen W., Jin C.S., Wei S. Grooming behavior and damaged mites (*Varroa jacobsoni*) in *Apis cerana cerana* and *Apis mellifera ligustica*. *Apidologie.* 1996;27:3-11. DOI 10.1051/apido:19960101.

Gilliam M., Taber S., Richardson G.V. Hygienic behavior of honey bees in relation to chalkbrood disease. *Apidologie.* 1983;14:29-39.

Gramacho K.P., Gonçalves L.S., Rosenkranz P., Jong D.D. Influence of body fluid from pin-killed honey bee pupae on hygienic behavior. *Apidologie.* 1999;30:367-374. DOI 10.1051/apido:19990502.

Haddad N., Batainh A.M., Migdadi O.S., Saini D., Krishnamurthy V., Parameswaran S., Alhamuri Z. Next generation sequencing of *Apis mellifera* syriaca identifies genes for *Varroa* resistance and beneficial bee keeping traits. *Insect Sci.* 2015;23:1-12. DOI 10.1111/1744-7917.12205.

Hamiduzzaman M.Md., Emsen B., Hunt G.J., Subramanyam S., Williams C.E., Tsuruda J.M., Guzman-Novoa E. Differential gene expression associated with honey bee grooming behavior in response to *Varroa* mites. *Behav. Genet.* 2017;47:335-344. DOI 10.1007/s10519-017-9834-6.

Harbo J.R., Harris J.W. Heritability in honey bees (Hymenoptera: Apidae) of characteristics associated with resistance to *Varroa jacobsoni* (Mesostigmata: Varroidae). *J. Econ. Entomol.* 1999;92:261-265.

Harbo J.R., Harris J.W. Suppressed mite reproduction explained by the behaviour of adult bees. *J. Apic. Res.* 2005;44(1):21-23. DOI 10.1080/00218839.2005.11101141.

Harpur B.A., Guarna M.M., Huxter E., Higo H., Moon K.-M., Hoover S.E., Ibrahim A., Melathopoulos A.P., Desai S., Currie R.W., Pernal S.F., Foster L.J., Zayed A. Integrative genomics reveals the genetics and evolution of the honey bee's social immune system. *Genome Biol. Evol*. 2019;11(3):937-948. DOI 10.1093/gbe/evz018.

Harris J. Bees with *Varroa* sensitive hygiene preferentially remove mite infested pupae aged ≤ five days post capping. *J. Apic. Res.* 2007;46:134-139. DOI 10.3896/IBRA.1.46.3.02.

Harris J.W., Danka R.G., Villa J.D. Honey bees (Hymenoptera: Apidae) with the trait of varroa sensitive hygiene remove brood with all reproductive stages of varroa mites (Mesostigmata: Varroidae). *Ann. Entomol. Soc. Am.* 2010;103:146-152. DOI 10.1603/AN09138.

Hu H., Bienefeld K., Wegener J., Zautke F., Hao Y., Feng M., Han B., Fang Y., Wubie A.J., Li J. Proteome analysis of the hemolymph, mushroom body, and antenna provides novel insight into honeybee resistance against varroa infestation. *J. Proteome Res.* 2016;15(8):2841-2854. DOI 10.1021/acs.jproteome.6b00423.

Invernizzi C., Zefferinoa I., Santosa E., Sanchez L., Mendoza Y. Multilevel assessment of grooming behavior against *Varroa destructor* in Italian and Africanized honey bees. *J. Apic. Res.* 2015;54(4):321-327. DOI 10.1080/00218839.2016.1159055.

Ji T., Yin L., Liu Z., Liang Q., Luo Y., Shen J., Shen F. Transcriptional responses in eastern honeybees (*Apis cerana*) infected with mites, *Varroa destructor*. *Genet. Mol. Res.* 2014;13(4):8888-8900. DOI 10.4238/2014.October.31.4.

Jiang S., Robertson T., Mostajeran M., Robertson A.J., Qiu X. Differential gene expression of two extreme honey bee (*Apis mellifera*) colonies showing *Varroa* tolerance and susceptibility. *Insect Mol. Biol.* 2016;25(3):272-282. DOI 10.1111/imb.12217.

Kamler M., Nesvorna M., Stara J., Erban T., Hubert J. Comparison of tau-fluvalinate, acrinathrin, and amitraz effects on susceptible and resistant populations of *Varroa destructor* in a vial test. *Exp. Appl. Acarol.* 2016;69(1):1-9. DOI 10.1007/s10493-016-0023-8.

Kim J.S., Kim M.J., Kim H.-K., Vung N.N., Kim I. Development of single nucleotide polymorphism markers specific to *Apis mellifera* (Hymenoptera: Apidae) line displaying high hygienic behavior against *Varroa destructor*, an ectoparasitic mite. *J. Asia Pac. Entomol.* 2019;22:1031-1039. DOI 10.1016/j.aspen.2019.08.005.

Kurze C., Routtu J., Moritz R.F.A. Parasite resistance and tolerance in honeybees at the individual and social level. *Zoology.* 2016;119(4):290-297. DOI 10.1016/j.zool.2016.03.007.

Land B., Seeley T. The grooming invitation dance of the honey bee. *Ethology.* 2004;110:1-10. DOI 10.1046/j.1439-0310.2003.00947.x.

Lartigue A., Gruez A., Briand L., Blon F., Bezirard V., Walsh M., Pernollet J.-C., Tegoni M., Cambillau C. Sulfur single-wavelength anomalous diffraction crystal structure of a pheromone-binding protein from the honeybee *Apis mellifera* L. *J. Biol. Chem.* 2004;279(6):4459-4464. DOI 10.1074/jbc.M311212200.

Lattorff H.M.G., Buchholz J., Fries I., Moritz R.F.A. A selective sweep in a *Varroa destructor* resistant honeybee (*Apis mellifera*) population. *Infect. Genet. Evol.* 2015;31:169-176. DOI 10.1016/j.meegid.2015.01.025.

Le Conte Y., Alaux C., Martin J.-F., Harbo J.R., Harris J.W., Dantec C., Severac D., Cros-Arteil S., Navajas M. Social immunity in honeybees (*Apis mellifera*): transcriptome analysis of varroa-hygienic behaviour. *Insect Mol. Biol.* 2011;20(3):399-408. DOI 10.1111/j.1365-2583.2011.01074.x.

Leclercq G., Pannebakker B., Gengler N., Nguyen B.K., Francis F. Drawbacks and benefits of hygienic behavior in honey bees (*Apis mellifera* L.): a review. *J. Apic. Res.* 2017;56(4):366-375. DOI 10.1080/00218839.2017.1327938.

Locke B. Inheritance of reduced *Varroa* mite reproductive success in reciprocal crosses of mite-resistant and mite-susceptible honey bees (*Apis mellifera*). *Apidologie.* 2016a;47(4):583-588. DOI 10.1007/s13592-015-0403-9.

Locke B. Natural *Varroa* mite-surviving *Apis mellifera* honeybee populations. *Apidologie.* 2016b;47(3):467-482. DOI 10.1007/s13592-015-0412-8.

Martin S.J., Highfield A.C., Brettell L., Villalobos E.M., Budge G.E., Powell M. Global honey bee viral landscape altered by a parasitic mite. *Science.* 2012;336(6086):1304-1306. DOI 10.1126/science.1220941.

McDonnell C.M., Alaux C., Parrinello H., Desvignes J.-P., Crauser D., Durbesson E., Beslay D., Le Conte Y. Ecto- and endoparasite induce similar chemical and brain neurogenomic responses in the honey bee (*Apis mellifera*). *BMC Ecol*. 2013;13:25. DOI 10.1186/1472-6785-13-25.

Mondet F., Alaux C., Severac D., Rohmer M., Mercer A.R., Conte Y.L. Antennae hold a key to *Varroa* sensitive hygiene behaviour in honey bees. *Sci. Rep.* 2015;5:10454. DOI 10.1038/srep10454.

Navajas M., Migeon A., Alaux C., Martin-Magniette M.L., Robinson G.E., Evans J.D., Cros-Arteil S., Crauser D., Le Conte Y. Differential gene expression of the honey bee *Apis mellifera* associated with *Varroa destructor* infection. *BMC Genom.* 2008;9:301. DOI 10.1186/1471-2164-9-301.

Nazzi F., Le Conte Y., Berenbaum M.R. Ecology of *Varroa destructor*, the major ectoparasite of the western honey bee, *Apis mellifera*. *Annu. Rev. Entomol.* 2016;61:417-432. DOI 10.1146/annurev-ento-010715-023731.

Oxley P.R., Spivak M., Oldroyd B.P. Six quantitative trait loci influence task thresholds for hygienic behaviour in honeybees (*Apis mellifera*). *Mol. Ecol.* 2010;19:1452-1461. DOI 10.1111/j.1365-294X.

Plettner E., Eliash N., Singh N.K., Pinnelli G.R., Soroker V. The chemical ecology of host-parasite interaction as a target of *Varroa destructor* control agents. *Apidologie.* 2017;48(1):78-92. DOI 10.1007/s13592-016-0452-8.

Pritchard D.J. Grooming by honey bees as a component of varroa resistant behavior. *J. Apic. Res.* 2016;55(1):38-48. DOI 10.1080/00218839.2016.1196016.

Rosenkranz P., Aumeier P., Ziegelmann B. Biology and control of *Varroa destructor. J. Invertebr. Pathol.* 2010;103:S96-119. DOI 10.1016/j.jip.2009.07.016.

Rothenbuhler W.C. Behavior genetics of nest cleaning in honey bees. IV. Responses of F1 and backcross generations to disease-killed brood. *Am. Zool.* 1964;4:111-123.

Spivak M. Honey bee hygienic behavior and defense against *Varroa jacobsoni*. *Apidologie.* 1996;27:245-260. DOI 10.1051/apido:19960407.

Spivak M., Reuter G.S. Resistance to American foulbrood disease by honey bee colonies *Apis mellifera* bred for hygienic behavior. *Apidologie.* 2001;32:555-565.

Spotter A., Gupta P., Mayer M., Reinsch N., Bienefeld K. Genome-wide association study of a *Varroa*-specific defense behavior in honeybees (*Apis mellifera*). *J. Hered.* 2016;107(3):220-227. DOI 10.1093/jhered/esw005.

Spotter A., Gupta P., Nurnberg G., Reinsch N., Bienefeld K. Development of a 44K SNP assay focussing on the analysis of a varroa-specific defence behaviour in honey bees (*Apis mellifera*

carnica). *Mol. Ecol. Resour*. 2012;12:323-332. DOI 10.1111/j.1755-0998.2011.03106.x.

Suenami S., Iino S., Kubo T. Pharmacologic inhibition of phospholipase C in the brain attenuates early memory formation in the honeybee (*Apis mellifera* L.). *Biol. Open.* 2018;7:bio028191. DOI 10.1242/bio.028191.

Tsuruda J.M., Harris J.W., Bourgeois L., Danka R.G., Hunt G.J. High-resolution linkage analyses to identify genes that influence Varroa sensitive hygiene behavior in honey bees. *PLoS One.* 2012;7(11):e48276. DOI 10.1371/journal.pone.0048276.

Villa J.D., Danka R.G., Harris J.W. Simplified methods of evaluating colonies for levels of Varroa Sensitive Hygiene (VSH). *J. Apic. Res.* 2009;48(3):162-167. DOI 10.3896/IBRA.1.48.3.03.

Wallberg A., Schoning C., Webster M.T., Hasselmann M. Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. *PLoS Genet.* 2017;13(5):e1006792. DOI 10.1371/journal.pgen.1006792.

Zakar E., Javor A., Kusza S. Genetic bases of tolerance to *Varroa destructor* in honey bees (*Apis mellifera* L.). *Insect. Soc*. 2014;61:207-215. DOI 10.1007/s00040-014-0347-5.

Zhao G., Wang C., Wang H., Gao L., Liu Z., Xu B., Guo X. Characterization of the *CDK5* gene in *Apis cerana cerana* (*AccCDK5*) and a preliminary identification of its activator gene, *AccCDK5r1*. *Cell Stress Chaperones.* 2018;23:13-28. DOI 10.1007/s12192-017-0820-y.

ORCID ID

M.D. Kaskinova orcid.org/0000-0003-4960-6559
L.R. Gaifullina orcid.org/0000-0002-3285-118X
E.S. Saltykova orcid.org/0000-0003-0123-7037
A.V. Poskryakov orcid.org/0000-0001-6189-4472
A.G. Nikolenko orcid.org/0000-0002-9235-680X

Acknowledgements. The work of 1 and 2 authors was carried out within the framework of state assignment No. AAAA-A16-116020350026-0; the work of 3–5 authors was funded by RFBR, project number 19-54-70002.

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

Received April 30, 2020. Revised July 10, 2020. Accepted July 23, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):861-867
DOI 10.18699/VJ20.684

Screening of West Siberian patients with primary congenital glaucoma for *CYP1B1* gene mutations

D.E. Ivanoshchuk[1] ✉, S.V. Mikhailova[1], O.G. Fenkova[2], E.V. Shakhtshneider[1], A.Z. Fursova[1], I.Y. Bychkov[3], M.I. Voevoda[1]

[1] Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State Regional Hospital, Novosibirsk, Russia
[3] S.N. Fyodorov FSBI IRTC Eye Microsurgery, Novosibirsk branch, Novosibirsk, Russia
✉ e-mail: dinara2084@mail.ru

Abstract. Primary congenital glaucoma (PCG) is a visual organ pathology that leads to progressive blindness and poor vision in children. Its main cause is an anomaly of the anterior chamber angle. Most cases of PCG are sporadic, but familial cases with an autosomal recessive (predominantly) and autosomal dominant (rare) type of inheritance have been described. Congenital glaucoma is a rare condition (1 case per 10,000–20,000 newborns), but its prevalence is substantially higher (up to 1 case per 250 newborns) in countries where consanguineous marriages are common. Mutations in the *CYP1B1* gene, which encodes cytochrome P450 1B1, are the most common cause of autosomal recessive primary congenital glaucoma. This enzyme is known to be involved in retinoic acid metabolism and is necessary for normal eye development. The aim of this work was to assess the polymorphism of the *CYP1B1* gene among West Siberian patients with primary congenital glaucoma. Direct automatic Sanger sequencing of exons and adjacent splicing sites of the *CYP1B1* gene was carried out in 28 people with the PCG phenotype from a West Siberian region. As a result, in the sample of the white population we examined, pathogenic variants previously described in other ethnic groups were revealed: E387K (rs55989760), R444* (rs377049098), R444Q (rs72549376), and P437L (rs56175199), as well as novel single-nucleotide deletion p.F114Lfs*38 in the *CYP1B1* gene. The latter can cause a frame shift, changed amino acid composition, and a formation of truncated in the protein. None of the detected mutations were found in the control sample of ophthalmologically examined individuals without PCG (100 people). Variants R444* (rs377049098) and R444Q (rs72549376) were not found in the general population sample either (576 randomly selected West Siberia residents). All the detected mutations caused the development of the autosomal recessive form of primary congenital glaucoma. The most severe clinical phenotype was observed in carriers of mutations in codon 444 of the gene. Consequently, in children with signs of increased intraocular pressure, molecular genetic analysis of the *CYP1B1* gene is advisable for early diagnosis and timely initiation of PCG therapy.
Key words: human; congenital glaucoma; *CYP1B1*; genetic analysis; cytochrome P450 1B1.

For citation: Ivanoshchuk D.E., Mikhailova S.V., Fenkova O.G., Shakhtshneider E.V., Fursova A.Z., Bychkov I.Y., Voevoda M.I. Screening of West Siberian patients with primary congenital glaucoma for *CYP1B1* gene mutations. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):861-867. DOI 10.18699/VJ20.684

Скрининг мутаций гена *CYP1B1* у пациентов Западной Сибири с первичной врожденной глаукомой

Д.Е. Иванощук[1] ✉, С.В. Михайлова[1], О.Г. Фенькова[2], Е.В. Шахтшнейдер[1], А.Ж. Фурсова[1], И.Ю. Бычков[3], М.И. Воевода[1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Государственная Новосибирская областная клиническая больница, Новосибирск, Россия
[3] Новосибирский филиал ФГАУ НМИЦ «МНТК Микрохирургия глаза» им. академика С.Н. Федорова, Новосибирск, Россия
✉ e-mail: dinara2084@mail.ru

Аннотация. Первичная врожденная глаукома – это патология органа зрения, которая приводит к прогрессирующей слепоте и слабовидению у детей. Основной ее причиной являются аномалии/нарушения развития угла передней камеры глаза. Большинство случаев возникновения первичной врожденной глаукомы спорадические, но описаны и семейные варианты с аутосомно-рецессивным (преимущественно) и аутосомно-доминантным (редко) типом наследования. Врожденная глаукома – редкое заболевание (один случай на 10000–20000 новорожденных), но частота ее возникновения существенно возрастает (до одного случая на 250 новорожденных) в странах, где распространены близкородственные браки. Мутации в гене *CYP1B1*, который кодирует цитохром P450 1B1, становятся наиболее частой причиной аутосомно-рецессивных форм первичной врожденной глаукомы. Известно, что этот фермент участвует в метаболизме ретиноевой кислоты и необходим для нормального развития глаза. Целью нашей работы была оценка полиморфизма гена *CYP1B1* у пациентов Западной Сибири с первичной врожденной глаукомой. Было проведено прямое автоматическое секвенирование по Сэнгеру экзонов и прилегающих сайтов сплайсинга гена *CYP1B1* у 28 человек из Запад-

но-Сибирского региона с фенотипом первичной врожденной глаукомы. В результате в обследованной нами выборке европеоидного населения выявлены ранее описанные в других этнических группах патогенные варианты этого гена: E387K (rs55989760), R444 * (rs377049098), R444Q (rs72549376) и P437L (rs56175199), а также новая однонуклеотидная делеция, приводящая к сдвигу рамки считывания p.F114Lfs*38 в гене *CYP1B1*. Последняя может вызывать образование белка с измененным аминокислотным составом и укороченного белка. Ни одна из выявленных мутаций не была выявлена в контрольной выборке офтальмологически обследованных лиц без первичной врожденной глаукомы (100 человек). Варианты R444* (rs377049098) и R444Q (rs72549376) не были найдены также в популяционной выборке (576 лиц, отобранных случайным образом) жителей Западной Сибири. Все обнаруженные варианты вызывали развитие аутосомно-рецессивной формы первичной врожденной глаукомы. При этом наиболее тяжелая клиника наблюдалась у носителей мутаций в 444 кодоне гена. Следовательно, у детей с признаками повышения внутриглазного давления оправдано проведение молекулярно-генетического анализа гена *CYP1B1* для ранней диагностики и своевременного начала терапии первичной врожденной глаукомы.
Ключевые слова: человек; врожденная глаукома; *CYP1B1*; генетический анализ; цитохром P450 1B1.

Introduction

Primary congenital glaucoma (PCG, OMIM 231300) is a visual organ pathology that leads to irreversible blindness and poor vision in children. The main cause of PCG is a malformation of the aqueous outflow system and disruption of its filtering ability followed by an increase in intraocular pressure, death of retinal ganglion cells, and as a consequence, blindness or a reduction in visual function (Thau et al., 2018; Badawi et al., 2019). PCG is a rare disease; its prevalence is within the range 1 per 10000–20000 live births in the USA, UK, and Ireland and is more frequent (1 per 1250 newborns) in populations where consanguineous marriages are common (Badawi et al., 2019). Most cases of PCG are sporadic, i.e., patients do not have a family history; however, familial cases with autosomal recessive (mainly) and autosomal dominant inheritance have been described as well (Fan, Wiggs, 2010; Souma et al., 2016; Hadrami et al., 2019). Numerous molecular genetic studies have shown genetic heterogeneity of PCG; various genes and combinations of their alleles can be involved in the formation of the pathological phenotype (Liu, Allingham, 2011; de Melo et al., 2015). The predominant cause of the autosomal recessive form of PCG is mutations in the *CYP1B1* gene (OMIM 601771): they account for up to 50 % of familial and up to 20 % of sporadic cases (Sarfarazi, Stoilov, 2000). Cytochrome P450 1B1, encoded by the *CYP1B1* gene, belongs to the superfamily of enzymes that oxidize steroids, fatty acids, and xenobiotics as well as carry out the biosynthesis of various endogenous compounds (Klingenberg, 1958). The role of this gene in the disease development is still not clear, but it is known that the CYP1B1 monooxygenase participates in the metabolism of retinoic acid, which is essential for normal eye development (Cvekl, Wang, 2009).

The *CYP1B1* gene is located on the short arm of chromosome 2 (2p22.2) and consists of three exons, the first of which is noncoding. A 543-amino-acid protein is encoded by exons 2 and 3 (Vasiliou, Gonzalez, 2008). The polypeptide contains a few functionally significant regions: proline-rich "hinge" and I-helix regions and a cytosolic globular domain, including highly conserved J-helix, K-helix, β-sheets, meander, and heme-binding regions (Stoilov et al., 1998; Zhao et al., 2015).

The enzyme is expressed in many tissues and organs (parenchymal and stromal tissue of the brain, kidneys, prostate, breasts, cervix, uterus, ovaries, and lymph nodes) and in intraocular structures. *CYP1B1* mRNA is detectable in the ciliary body, iris, and retina but is absent in trabecular meshwork (Muskhelishvili et al., 2001; Doshi et al., 2006). A study on homozygous mouse knockouts of this gene showed an ocular drainage structure malformation, confirming the involvement of this gene in the development of ocular aberrations (Libby et al., 2003). In humans, *CYP1B1* gene expression was revealed throughout all embryonic eye development and during the postnatal period, but its level is higher in fetal eyes than in adult ones. It can be assumed that the product of this gene metabolizes some important substrate that plays a key role in the development and maturation of eye tissues (Doshi et al., 2006).

More than 120 distinct mutations in this gene have been associated with the autosomal recessive form of PCG (http://www.hgmd.cf.ac.uk/ac/all.php). There are single-nucleotide substitutions in exons and regulatory regions (missense or frameshift mutations or premature stop codons) and larger rearrangements (insertions/deletions) in the *CYP1B1* gene, which alter its transcription and translation (http://www.hgmd.cf.ac.uk/ac/all.php). Of note, only one splice site pathogenic variant has been described until now (Afzal et al., 2019). Pathogenic mutations in the *CYP1B1* gene associated with PCG are most often localized in the hinge region or the cytosolic globular domain, where they change protein folding, heme binding, and the electron transfer ability (Sarfarazi, Stoilov, 2000).

Thus, the aim of this study was identification of the spectrum of mutations in coding and adjacent noncoding parts of the *CYP1B1* gene in patients with PCG in West Siberia by Sanger sequencing.

Materials and methods

The study protocol was approved by the local Ethics Committee of the Institute of Internal and Preventive Medicine (a branch of the Institute of Cytology and Genetics, the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia). Written informed consent to be examined and to participate in the study was obtained from each patient. For individuals younger than 18 years, the informed consent was signed by a parent or legal guardian.

Twenty-eight white patients with PCG were examined. The study population consisted of three families, two patients each (in two families, a pair of monozygous twins and one family with two sisters), and 22 patients who were not related to each other. There were 26 children (10 males and 16 females, from 1 to 12 years old, mean age 8.0 ± 5.5) and two adult sisters aged

45 and 49. All the patients were examined at the Novosibirsk Regional Hospital. Age of onset and a family history were recorded after an interview with the patients or their parents and were based on medical records. All study participants underwent ophthalmic examination: visual-acuity measurement, slit lamp biomicroscopic examination, indirect gonioscopy, tonometry, corneal pachymetry, fundoscopy, and optical coherence tomography of the optic nerve head. Other ocular aberrations and systemic disease were exclusion criteria.

Two control groups were used: healthy and population cohorts. One hundred healthy people (26–83 years old, mean age 67.6±6.9, males 36 %, whites 100 %), who did not have a family history of episodes of glaucoma and other systemic diseases, were enrolled during routine examination at FSBI IRTC Eye Microsurgery (Novosibirsk, Russia). The general-population group (576 subjects total) was randomly selected from two surveys: the population interviewed within the framework of the HAPIEE project (Pajak et al., 2013), Novosibirsk, Russia (376 people, mean age 53.96±6.4 years) and adolescents (188 subjects, mean age 14.83±0.88) from the same region (Zavyalova et al., 2011).

Genomic DNA for Sanger sequencing was isolated from leukocytes of venous blood by phenol-chloroform extraction (Sambrook, Russell, 2006). The primers were reported previously (Gong et al., 2015). PCRs were carried out using BioMaster LR HS-PCR (2x) (BiolabMix, Russia). The cycling program consisted of denaturing at 94 °C for 3 minutes and then 35 cycles of 94 °C for 30 seconds, 68 °C for 30 seconds, and 72 °C for 50 seconds. The PCR products were evaluated by electrophoresis in a 5 % polyacrylamide gel after visualization with ethidium bromide. A 100 bp DNA ladder (SibEnzyme, Russia) was added into each gel as a control. The amplicons were purified on Agencourt AMPure Xp beads (Beckman Coulter, USA), and the sequencing reactions were carried out on an automated ABI 3500 DNA sequencer (Thermo Fisher Scientific, USA) with the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, USA). The sequences were analyzed in the Vector NTI® Advance software (Thermo Fisher Scientific). We chose a wild-type sequence of the human *CYP1B1* gene from the Ensembl Genome Browser (https://www.ensembl.org/index.html) as a reference for alignment.

Rs72549376 and rs377049098 were genotyped by the restriction fragment length polymorphism analysis. Forward and reverse primers were designed by means of the Primer-Blast software (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The following primers were selected for both single-nucleotide variants (SNVs): 5′-CCTTTATGAAGCCATGCGC-3′ and 5′-TGGTCAGGTCCTTGTTGATGAG-3′. PCRs were set up using BioMaster HS-Taq PCR (2×) (BiolabMix, Russia), 1 μl of each primer, and 1 μl of DNA with a total final volume of 25 μl. The PCR program consisted of initial denaturation at 94 °C for 3 minutes and then 30 cycles of 94 °C for 20 seconds, 59 °C for 20 seconds, and 72 °C for 30 seconds. For genotyping of rs377049098 and rs72549376, 5 U of each restriction enzyme, Tag I and HinfI (SibEnzyme, Russia), were added to the PCR product and incubated for 12–16 h at 37 and 65 °C, respectively.

To detect the C/C genotype of rs377049098, 178 and 43 bp fragments were examined; to detect the T/T genotype, a 221 bp fragment was expected. All the fragments had to be present to detect the C/T genotype. If both gene copies are mutated, then the restriction site is disrupted, and product length should look like that for the T/T genotype: 221 bp. To detect the A/A genotype of rs72549376, 130, 50, and 41 bp fragments were examined; to detect the G/G genotype, 130 and 91 bp fragments were expected. The A/G genotype necessitated the presence of all the fragments. The PCR products were evaluated by electrophoresis in a 5% polyacrylamide gel after visualization with an ethidium bromide solution. A 100-bp DNA Ladder (SibEnzyme, Russia) served as molecular size markers on each gel.

Results

We analyzed exons and adjacent splice sites of the CYP1B1 gene in 28 patients with PCG. In all the patients examined, the diagnosis of PCG had been made before 2 years of age (in 27 patients before age 1 year); in one male and one female, the damage was unilateral (Table 1). After the diagnosis had been made, 25 patients underwent surgical treatment, and three had a nonsurgical intervention: eye drops. Conservative treatment (selective β-blockers and/or carbonic anhydrase inhibitors) was administered either only as a preoperative preparatory procedure or to prolong the hypotensive effect of antiglaucoma surgery in the late postoperative period (observation period from 1 year to 12 years). In probands P4 and P5, enucleation had been performed due to total retinal detachment, loss of vision, and severe pain syndrome (see Table 1). An analysis of the family history of the patients indicated autosomal recessive PCG inheritance. Genetic analysis of the patients revealed four previously described missense variants R444Q (rs72549376), E387K (rs55989760), R444* (rs377049098), and P437L (rs56175199) and a novel single-nucleotide deletion of cytosine: p.F114Lfs*38 (see Table 1). We did not identify homozygous carriers of a mutation in the studied gene. Compound heterozygous pathogenic variants of CYP1B1 were identified in three PCG cases. In one proband (P1) without a family history of PCG, we identified a single CYP1B1 mutation, R444Q, without any additional genetic variants in this gene. In three families, one patient with PCG was registered (P1, P2, and P3), and in one family, two sisters (P4 and P5) turned out to be compound heterozygous carriers of the CYP1B1 substitutions.

We found no carriers of R444*, R444Q, p.F114Lfs*38, P437L, or E387K among 100 healthy controls. The first two were not found in any of the 576 members of the general-population group. In addition to rare pathogenic variants, we found 6 previously described common SNVs, 5 of which are located in the coding part of *CYP1B1* and one in the gene promoter (Table 2).

It was previously shown that the detected substitutions (rs10012, rs1056827, rs1056836, rs1056837, and rs1800440) are in linkage disequilibrium in different populations (Chavarria-Soley et al., 2006). For some pathogenic mutations (including E387K, P437L, and R444Q), linkage to certain intragenic haplotypes of the *CYP1B1* gene has been shown (Plásilová et al., 1999; Sena et al., 2004; Chavarria-Soley et al., 2006). Accordingly, we analyzed the haplotypes of the *CYP1B1* gene in West Siberian patients with PCG that carried rare mutations (Table 3).

Table 1. Mutations identified in the *CYP1B1* gene among PCG probands

Proband, sex	*CYP1b1* variant (official name)	Minor allele frequency, gnomAD	Age, years/age of diagnosis	Clinical findings	Eye condition	Surgical operation, times	Family history of glaucoma
P1, male	R444Q (rs72549376)	0.00001	10/3 months	Bilateral buphthalmos, epiphora, photophobia, Haabs striae opacification of the cornea, high intraocular pressure, anxiety	Bilateral PCG, binocular blindness	3	No
	Not detected						
P2, male	E387K (rs55989760)	0.00027	12/6 months	Bilateral buphthalmos, epiphora, photophobia, Haabs striae opacification of the cornea, high intraocular pressure, anxiety	Bilateral PCG, stable	1	No
	R444* (rs377049098)	0.00003					
P3, male	P437L (rs56175199)	0.00002	9/8 months	Bilateral buphthalmos, epiphora, photophobia, Haabs striae opacification of the cornea, high intraocular pressure	Bilateral PCG, stable	1	No
	p.F114Lfs*38 (new)	–					
P4, female	R444* (rs377049098)	0.00003	49/2 years	Bilateral buphthalmos, epiphora, photophobia, Haabs striae opacification of the cornea, high intraocular pressure, anxiety	PCG (right), stable; left anophthalmos	12	Sister of P5 with PCG
	R444Q (rs72549376)	0.00001					
P5, female	R444* (rs377049098)	0.00003	45/4 months	Bilateral megalocornea without opacification of the cornea, high intraocular pressure	PCG (right), progressive; left anophthalmos	10	Sister of P4 with PCG
	R444Q (rs72549376)	0.00001					

No – absent of glaucoma family history.

Table 2. Common variants of the *CYP1B1* gene found in West Siberian patients with PCG, and the SNVs' minor allele frequency according to the gnomAD database

Name of an SNV	Location	Substitution (NP_000095.2)	Minor allele frequency (gnomAD)
rs2617266	Promoter	c.-1-12C>T	T = 0.29
rs10012	Exon 2	R48G c.142C>G	C = 0.31
rs1056827	Exon 2	A119S c.355G>T	A = 0.32
rs1056836	Exon 3	L432V c.1294G>C	C = 0.38
rs1056837	Exon 3	D449D c.1347T>C	T = 0.38
rs1800440	Exon 3	N453S c.1358A>G	G = 0.15

Discussion

As a rule, the indications for tonometric examination of children with suspected PCG are symptoms associated with or caused by increased intraocular pressure: pronounced hydrophthalmos, photophobia, tearing eyes, corneal whitening, and anxiety. Because the symptoms may be more or less pronounced, a genetic analysis can confirm the diagnosis, especially in families with previously detected PCG cases. Genetic heterogeneity of PCG makes it difficult to identify causative variants, thereby complicating assessments of disease risk and severity in probands and their relatives.

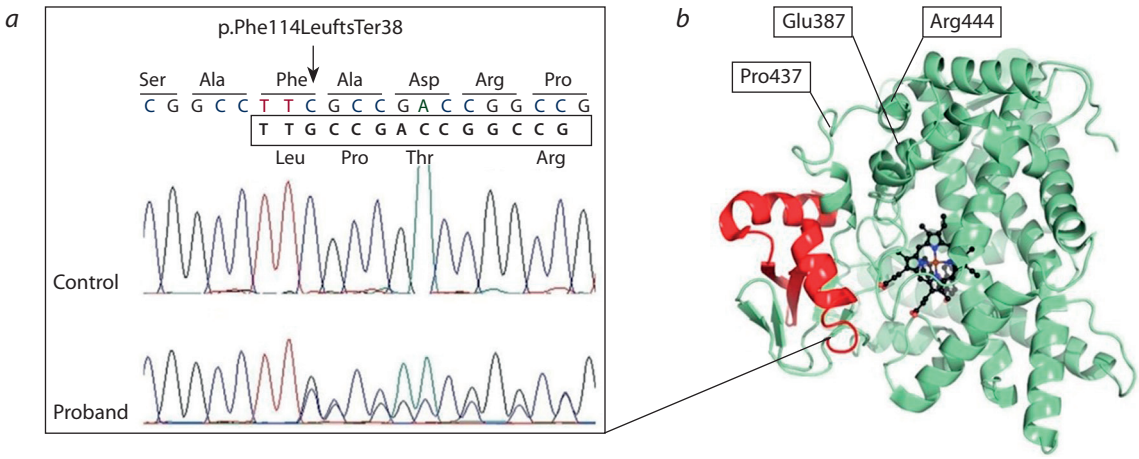
Among the analyzed West Siberian patients with PCG, 16 % (four out of 25 unrelated people) were carriers of *CYP1B1* mutation(s). Three cases were found to be compound heterozygotes in terms of previously described variants [E387K (rs55989760), R444* (rs377049098), and R444Q (rs72549376)]. In one case, the mutation combination consisted of P437L (rs56175199) and novel frameshift truncating mutation p.F114Lfs*38 (see Figure).

Amino acid residue Arg444 located in the "meander" region was predicted to be important for structural stabilization of the protein. Replacing Arg with Gln disrupts the contact of the amino acid residue at this position with the oxygen atoms of Trp434 and Asn439 inside the molecule. It can destabilize this structural component and affects the heme-binding and redox functions of CYP1B1 (Mashima et al., 2001). In the case of termination/stop codon Arg444* formation, a truncated protein is synthesized that does not contain the conserved L-helix and heme-binding regions.

In patient P1's genome, we found only one Arg444Gln substitution in the *CYP1B1* gene. Because the proband's parents did not have glaucoma, his pathological phenotype is most likely due to an additional genetic factor or factors. The same variant in compound heterozygotes was detected in two sisters, P4 and P5. Clinical presentation of congenital glaucoma associated with Arg444 substitution carriage was characterized by faster progression of the pathological process with subsequent loss of vision and an eye as a whole, despite active treatment (12 and 10 operations with subsequent unilateral enucleation in P4 and P5, respectively, and poor vision in the remaining eye; three operations in P1 with binocular glaucoma blindness as a result). Detailed examination of

Table 3. Probable intragenic haplotypes for *CYP1B1* in Russian patients with PCG

Proband	Mutation	R48G (rs10012)	A119S (rs1056827)	V432L (rs1056836)	D449D (rs1056837)	N453S (rs1800440)	Haplotype (amino acids)
P1	R444Q (rs72549376)	CC	GG	GG	TT	AA	CGGTA (RAVDN)
	ND						CGGTA (RAVDN)
P2	Glu387Lys (rs55989760)	GG	TT	CC	CC	AA	GTCCA (GSLDN)
	R444* (rs377049098)						GTCCA (GSLDN)
P3	P437L (rs56175199)	CC	GG	GC	TC	AA	CGGTA (RAVDN)
	p.F114Lfs*38 (new)						CGCCA (RALDN)
P4	R444* (rs377049098)	GC	GT	GC	TC	AA	GTCCA (GSLDN)
	R444Q (rs72549376)						CGGTA (RAVDN)
P5	R444* (rs377049098)	GC	GT	GC	TC	AA	GTCCA (GSLDN)
	R444Q (rs72549376)						CGGTA (RAVDN)
Healthy daughter of P4	R444* (rs377049098)	GC	GT	GC	TC	AA	GTCCA (GSLDN)
	R444R (rs377049098)						CGGTA (RAVDN)
Healthy mother of P4 and P5	R444* (rs377049098)	GC	GT	GC	TC	AA	GTCCA (GSLDN)
	R444R (rs377049098)						CGGTA (RAVDN)



Novel mutation in *CYP1B1* gene.

(*a*) An electropherogram of DNA sequence with the new single-nucleotide deletion, p.F114Lfs*38, in the *CYP1B1* gene. (*b*) 3D-structure of cytochrome P450 1B1 [Protein Data Bank ID 3PM0 (Wang et al., 2011)]. Red means the protein portion remaining unchanged in the case of deletion p.F114Lfs*38; localization of identified amino acid substitutions is indicated.

available relatives uncovered no signs of PCG, either in the mother of patients P4 and P5 or in the daughter of P4 (data not shown). Analysis of the general-population and healthy control groups revealed low prevalence of R444Q and R444* among Russians (less than 0.001).

It is likely that the mutations affecting the "meander" region can aggravate the course of the disease, as compared with other mutations. Mutations in this region have been described previously in Korean, Japanese, Lebanese, and Pakistani patients (Mashima et al., 2001; Chouiter, Nadifi, 2017; Micheal et al., 2017).

Missense mutation P437L of *CYP1B1* has been described previously too. Proline at position 437 is located on the protein surface and determines conformational rigidity of the structure by strongly bending the polypeptide chain. The P437L substitution may alter this special conformation and disrupt interactions of CYP1B1 with other molecules (Rashid et al., 2019). This mutation is found in populations of India, Pakistan, Brazil, Saudi Arabia, and Turkey (Kaur et al., 2011; Chouiter, Nadifi, 2017; Rashid et al., 2019).

The novel single-nucleotide deletion of cytosine in codon 114 of the *CYP1B1* gene changes the amino acid sequence and creates a premature stop codon at amino acid position 152. The resulting truncated protein with altered amino acid composition does not contain highly conserved regions, such as the cytosolic globular domain. This variant was not found in gnomAD, HGMD, or ClinVar. Various deletions and duplications in the *CYP1B1* gene are described in different populations but are more common in white patients with PCG (Sarfarazi, 2018).

The E387K substitution is most frequent in Europe (Chouiter, Nadifi, 2017). E387 is an invariant amino acid residue for all CYP450 family members (Stoilov et al., 1998; Sorenson et al., 2015). Lysine in this codon disrupts K-helix orientation and prevents formation of a stable hemoprotein complex (Stoilov et al., 1998; Sorenson et al., 2015). Among gypsies in Slovakia, this substitution was found in 100% of PCG cases; this result was explained by the founder effect (Plásilová et al., 1999). Glu387Lys has been identified in French, Brazilian, Canadian, Hungarian, US, and Spanish patients (Melki et al., 2004; Sena et al., 2004).

It was found that among whites, 4 haplotypes are most common for 5 CYP1B1 variants R48G, A119S, V432L, D449D, and N453S. They determine the formation of proteins with amino acids RALDN, RAVDN, RALDS, and GSLDN at the respective positions. Different enzymatic activities were experimentally established for these protein variants (Chavarria-Soley et al., 2008). The enzyme with RAVDN amino acids is 4-fold more active than the variant containing GSLDN. In combination with pathogenic variants, the enzymatic activity is even lower: because of a decrease in the enzymatic activity in the case of G61E and N203S, due to a decrease in the amount of protein in the case of Y81N and E229K, or both in the case of L343del (Chavarria-Soley et al., 2008). Combinations of the polymorphisms and rare mutations can cause additional differences in the phenotypic manifestation of PCG or the severity of the disease. Detection of substitutions R444Q, P437L, and E387K among Russians in the same intragenic haplotypes as in populations from Brazil, USA, Japan, and Romania, and Roma from Slovakia (Chavarria-Soley et al., 2006) allowed us to assume the monophyletic origin of these mutations in Asian, European, Roma, and Brazilian ethnic groups.

In Russia, genetic screening of PCG patients has been performed in the Republic of Bashkortostan and in St. Petersburg (14 and 45 patients, respectively) (Motushchuk et al., 2009; Lobov, 2017). No pathogenic substitutions in the *CYP1B1* gene were found in the Bashkortostan PCG patients. In one of the St. Petersburg patients, a heterozygous insertion of the CTC trinucleotide in codon 369 (c.1508insCTC, p.P369ins) was detected (Motushchuk et al., 2009); because the family history of the patient was not described in that study, it is impossible to determine the type of PCG inheritance.

Patients with PCG carrying pathogenic variants in the *CYP1B1* gene require more surgical operations to correct intraocular pressure and more thorough postoperative maintenance (as compared with patients without mutations in this gene) (Abu-Amero et al., 2011). Therefore, screening for mutations in *CYP1B1* gene of children with early-onset glaucoma is advisable for early detection of PCG; such patients subsequently require special attention.

Conclusions

In white West Siberian patients with PCG, previously described variants E387K (rs55989760), R444* (rs377049098), R444Q (rs72549376), and P437L (rs56175199) and novel frameshift mutation p.F114Lfs*38 were identified. In our study, the most serious clinical phenotype was noted in carriers of mutations R444Q and R444*. Identification of pathogenic variants in patients will contribute to their vision loss minimization owing to early disease detection and regular medical examinations of the substitution carriers. For the patients' family members, this analysis is also recommended because for individuals who are not carriers of pathogenic variants, the risk of PCG is comparable to that in the general population, and they do not require thorough ophthalmic monitoring.

For verification of PCG diagnoses in West Siberian patients, sequencing of exons and adjacent splice sites of the *CYP1B1* gene (rather than searching for point mutations) is recommended due to the absence of major causative mutations.

References

Abu-Amero K.K., Osman E.A., Mousa A., Wheeler J., Whigham B., Allingham R.R., Hauser M.A., Al-Obeidan S.A. Screening of CYP1B1 and LTBP2 genes in Saudi families with primary congenital glaucoma: genotype-phenotype correlation. *Mol. Vis*. 2011;17: 2911-2919.

Afzal R., Firasat S., Kaul H., Ahmed B., Siddiqui S.N., Zafar S.N., Shahzadi M., Afshan K. Mutational analysis of the CYP1B1gene in Pakistani primary congenital glaucoma patients: identification of four known and a novel causative variant at the 30 splice acceptor site of intron 2. *Congen. Anom.* 2019;59(5):152-161. DOI 10.1111/cga.12312.

Badawi A.H., Al-Muhaylib A.A., Al Owaifeer A.M., Al-Essa R.S., Al-Shahwan S.A. Primary congenital glaucoma: An updated review. *Saudi J. Ophthalmol*. 2019;33(4):382-388. DOI 10.1016/j.sjopt.2019.10.002.

Chavarria-Soley G., Michels-Rautenstrauss K., Pasutto F., Flikier D., Flikier P., Cirak S., Bejjani B., Winters D.L., Lewis R.A., Mardin C., Reis A., Rautenstrauss B. Primary congenital glaucoma and Rieger's anomaly: extended haplotypes reveal founder effects for eight distinct CYP1B1 mutations. *Mol. Vis.* 2006;12:523-531.

Chavarria-Soley G., Sticht H., Aklillu E., Ingelman-Sundberg M., Pasutto F., Reis A., Rautenstrauss B. Mutations in CYP1B1 cause primary congenital glaucoma by reduction of either activity or abundance of the enzyme. *Hum. Mut.* 2008;29(9):1147-1153.

Chouiter L., Nadifi S. Analysis of CYP1B1 gene mutations in patients with primary congenital glaucoma. *J. Pediatr. Genet.* 2017;6:205-214. https://doi.org/10.1055/s-0037-1602695.

Cvekl A., Wang W.L. Retinoic acid signaling in mammalian eye development. *Exp. Eye Res.* 2009;89(3):280-291. DOI 10.1016/j.exer.2009.04.012.

de Melo M.B., Mandal A.K., Tavares I.M., Ali M.H., Kabra M., de Vasconcellos J.P., Senthil S., Sallum J.M., Kaur I., Betinjane A.J., Moura C.R., Paula J.S., Costa K.A., Sarfarazi M., Paolera M.D., Finzi S., Ferraz V.E., Costa V.P., Belfort R. Jr., Chakrabarti S. Genotype-phenotype correlations in CYP1B1-associated primary congenital glaucoma patients representing two large cohorts from India and Brazil. *PLoS One.* 2015;10(5):e0127147. DOI 10.1371/journal.pone.0127147.

Doshi M., Marcus C., Bejjani B.A., Edward D.P. Immunolocalization of CYP1B1 in normal, human, fetal and adult eyes. *Exp. Eye Res.* 2006;82(1):24-32. DOI 10.1016/j.exer.2005.04.016.

Fan B.J., Wiggs J.L. Glaucoma: genes, phenotypes, and new directions for therapy. *J. Clin. Investig*. 2010;120(9):3064-3072. DOI 10.1172/JCI43085.

Gong B., Qu C., Li X., Shi Y., Lin Y., Zhou Y., Shuai P., Yang Y., Liu X., Zhang D., Yang Z. Mutation spectrum of CYP1B1 in Chinese patients with primary open-angle glaucoma. *Br. J. Ophthalmol*. 2015;99(3):425-430. DOI 10.1136/bjophthalmol-2014-306054.

Hadrami M., Bonnet C., Zeitz C., Veten F., Biya M., Hamed C.T., Condroyer C., Wang P., Sidi M.M., Cheikh S., Zhang Q., Audo I., Petit C., Houmeid A. Mutation profile of glaucoma candidate genes in Mauritanian families with primary congenital glaucoma. *Mol. Vis.* 2019;25:373-381. Published online 2019 Jul 13.

Kaur K., Mandal A.K., Chakrabarti S. Primary congenital glaucoma and the involvement of CYP1B1. *Middle East Afr. J. Ophthalmol.* 2011;18(1):7-16. DOI 10.4103/0974-9233.75878.

Klingenberg M. Pigments of rat liver microsomes. *Arch. Biochem. Biophys*. 1958;75:376-386. DOI 10.1016/0003-9861(58)90436-3.

Libby R.T., Smith R.S., Savinova O.V., Zabaleta A., Martin J.E., Gonzalez F.J., John S.W. Modification of ocular defects in mouse developmental glaucoma models by tyrosinase. *Science.* 2003;299(5612): 1578-1581. DOI 10.1126/science.1080095.

Liu Y., Allingham R.R. Molecular genetics in glaucoma. *Exp. Eye Res.* 2011;93(4):331-339. DOI 10.1016/j.exer.2011.08.007.

Lobov S.L., Khasanova R.R., Zagidullina A.S., Zaydullin I.S., Dzhemileva L.U., Khusnutdinova E.K. Analysis mutations CYP1B1 gene in patients of hereditary forms of glaucoma. *Medical Genetics.* 2017; 16(6):29-35. (in Russian)

Mashima Y., Suzuki Y., Sergeev Y., Ohtake Y., Tanino T., Kimura I., Miyata H., Aihara M., Tanihara H., Inatani M., Azuma N., Iwata T., Araie M. Novel cytochrome P4501B1 (CYP1B1) gene mutations in Japanese patients with primary congenital glaucoma. *Investig. Ophthalmol. Vis. Sci.* 2001;42(12):2211-2216.

Melki R., Colomb E., Lefort N., Brézin A.P., Garchon H.J. CYP1B1 mutations in French patients with early-onset primary open-angle glaucoma. *J. Med. Genet.* 2004;41:647-651. DOI 10.1136/jmg.2004.020024.

Micheal S., Siddiqui S.N., Zafar S.N., Florijn R.J., Bikker H., Boon C.J.F., Khan M., Hollander A.I.D., Bergen A. Identification of novel variants in CYP1B1, PITX2, FOXC1, and PAX6 in congenital glaucoma and anterior segment dysgenesis. *Investig. Ophthalmol. Vis. Sci*. 2017;58:2124.

Motushchuk A.E., Grudinina N.A., Rakhmanov V.V., Mandelstam M.Yu., Astakhov Yu.S., Vasiliev V.B. New P369INS mutation in the CYP1B1 gene in a patient with primary congenital glaucoma from St. Petersburg. *Scientific Notes of the Pavlov University.* 2009; 16(2):88-89. (in Russian)

Muskhelishvili L., Thompson P.A., Kusewitt D.F., Wang C., Kadlubar F.F. In situ hybridization and immunohistochemical analysis of cytochrome P450 1B1 expression in human normal tissues. *J. Histochem. Cytochem.* 2001;49:229-236. DOI 10.1177/002215540104900210.

Pajak A., Szafraniec K., Kubinova R., Malyutina S., Peasey A., Pikhart H., Nikitin Y., Marmot M., Bobak M. Binge drinking and blood pressure: cross-sectional results of the HAPIEE study. *PLoS ONE.* 2013;8(6):e65856. DOI 10.1371/journal.pone.0065856.

Plásilová M., Stoilov I., Sarfarazi M., Kádasi L., Feráková E., Ferák V. Identification of a single ancestral CYP1B1 mutation in Slovak Gypsies (Roms) affected with primary congenital glaucoma. *J. Med. Genet.* 1999;36(4):290-294.

Rashid M., Yousaf S., Sheikh S.A., Sajid Z., Shabbir A.S., Kausar T., Tariq N., Usman M., Shaikh R.S., Ali M., Bukhari S.A., Waryah A.M., Qasim M., Riazuddin S., Ahmed Z.M. Identities and frequencies of variants in CYP1B1 causing primary congenital glaucoma in Pakistan. *Mol. Vis.* 2019;25:144-154.

Sambrook J., Russell D.W. Purification of nucleic acids by extraction with phenol:chloroform. Cold Spring Harbor Protocols: Cold Spring Harbor, 2006.

Sarfarazi M. Targeted screening for predominant CYP1B1 mutations in primary congenital glaucoma. *J. Ophthalmic. Vis. Res*. 2018;3(4): 373-375. DOI 10.4103/jovr.jovr_232_18.

Sarfarazi M., Stoilov I. Molecular genetics of primary congenital glaucoma. *Eye* (*Lond*). 2000;(Pt 3B):422-428. DOI 10.1038/eye.2000.126.

Sena D.F., Finzi S., Rodgers K., Del Bono E., Haines J.L., Wiggs J.L. Founder mutations of CYP1B1 gene in patients with congenital glaucoma from the United States and Brazil. *J. Med. Genet.* 2004; 41(1):e6. DOI 10.1136/jmg.2003.010777.

Souma T., Tompson S.W., Thomson B.R., Siggs O.M., Kizhatil K., Yamaguchi S., Feng L., Limviphuvadh V., Whisenhunt K.N., Maurer-Stroh S., Yanovitch T.L., Kalaydjieva L., Azmanov D.N., Finzi S., Mauri L., Javadiyan S., Souzeau E., Zhou T., Hewitt A.W., Kloss B., Burdon K.P., Mackey D.A., Allen K.F., Ruddle J.B., Lim S.H., Rozen S., Tran-Viet K.N., Liu X., John S., Wiggs J.L., Pasutto F., Craig J.E., Jin J., Quaggin S.E., Young T.L. Angiopoietin receptor TEK mutations underlie primary congenital glaucoma with variable expressivity. *J. Clin. Investig*. 2016;126(7):2575-2587. DOI 10.1172/JCI85830.

Stoilov I., Akarsu A.N., Alozie I., Child A., Barsoum-Homsy M., Turacli M.E., Or M., Lewis R.A., Ozdemir N., Brice G., Aktan S.G., Chevrette L., Coca-Prados M., Sarfarazi M. Sequence analysis and homology modeling suggest that primary congenital glaucoma on 2p21 results from mutations disrupting either the hinge region or the conserved core structures of cytochrome P4501B1. *Am. J. Hum. Genet*. 1998;62(3):573-584. DOI 10.1086/301764.

Thau A., Lloyd M., Freedman S., Beck A., Grajewski A., Levin A.V. New classification system for pediatric glaucoma: Implications for clinical care and a research registry. *Curr. Opin. Ophthalmol.* 2018; 29(5):385-394. DOI 10.1097/icu.0000000000000516.

Vasiliou V., Gonzalez F.J. Role of CYP1B1 in glaucoma. *Annu. Rev. Pharmacol. Toxicol.* 2008;48:333-358. DOI 10.1146/annurev.pharmtox.48.061807.154729.

Wang A., Savas U., Stout C.D., Johnson E.F. Structural characterization of the complex between α-naphthoflavone and human cytochrome P450 1B1. *J. Biol. Chem.* 2011;18;286(7):5736-5743. DOI 10.1074/jbc.M110.204420.

Zavyalova L.G., Denisova D.V., Simonova G.I., Orlov P.S., Voevoda M.I. Association of polymorphisms of genes FTO and TCF7L2 with cadiometabolic parameters of the adolescents in Siberia. *Bulleten SB RAMS.* 2011;31(5):1-13. (in Russian)

Zhao Y., Sorenson C.M., Sheibani N., Cytochrome P450 1B1 and primary congenital glaucoma. *J. Ophthalmic Vis. Res.* 2015;10(1):60-67. DOI 10.4103/2008-322X.156116.

ORCID ID

D.E. Ivanoshchuk orcid.org/0000-0002-0403-545X
S.V. Mikhailova orcid.org/0000-0002-0897-5473
O.G. Fenkova orcid.org/0000-0001-8048-943X
E.V. Shakhtshneider orcid.org/0000-0001-6108-1025
A.Z. Fursova orcid.org/0000-0001-6311-5452
M.I. Voevoda orcid.org/0000-0001-9425-413X

Acknowledgements. This study was conducted within the framework of Russian Foundation for Basic Research project No. 18-315-00297 and as part of the main topic in state assignment No. AAAA-A17-117072710029-7. The English language was corrected and certified by shevchuk-editing.com. The authors thank the patients for participation in this study. The authors are grateful to Nikita V. Ivanisenko for help with the prediction of CYP1B1 protein structure. The authors declare that they have no conflicts of interest associated with publication of this article.

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

Received March 17, 2020. Revised October 18, 2020. Accepted November 3, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):868-875
DOI 10.18699/VJ20.685

Полиморфизм гена маннозосвязывающего лектина у коренных популяций территорий Арктической зоны Российской Федерации

С.Ю. Терещенко ✉, М.В. Смольникова

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

Аннотация. Маннозосвязывающий лектин (mannose-binding lectin, MBL) – паттерн-распознающий острофазовый белок, относящийся к системе врожденного иммунитета и активно участвующий в элиминации широкого круга патогенных микроорганизмов посредством активации лектинового пути системы комплемента. Значительная часть человеческой популяции имеет врожденно низкий уровень продукции и/или низкую функциональную активность MBL вследствие носительства различных вариантов гена *MBL2*, что может модифицировать течение самых разнообразных инфекционных заболеваний. Частота генотипов и гаплотипов полиморфизмов в гене *MBL2* имеет значительные популяционные различия. К настоящему времени данные относительно распределения генотипов гена *MBL2* в коренных популяциях территорий Арктической зоны Российской Федерации отсутствуют. Цель исследования – изучение частоты и этнической специфики распределения аллельных вариантов полиморфизмов гена *MBL2* rs11003125, rs7096206, rs7095891, rs5030737, rs1800450 и rs1800451 и их гаплотипов в популяциях Таймырского Долгано-Ненецкого района Красноярского края (ненцы, долганы-нганасаны, русские). В настоящем исследовании нами впервые получены данные о частотах генотипов и гаплотипов гена *MBL2* у коренных народностей, проживающих на территориях Арктической зоны Российской Федерации. Частота встречаемости гаплотипа *HYPA*, ассоциированного с высокой концентрацией MBL, составила 35.4 % для русских новорожденных Восточной Сибири, что соответствует частотам европейских популяций (27–33 %). У новорожденных арктических популяций частота гаплотипа *HYPA* была статистически значимо выше, чем у русских, и составила 64 % для ненцев и 56 % для долган-нганасан, что приближается к значениям частот, выявленных для эскимосов и североамериканских индейцев (64–81 %). Популяции ненцев и долган-нганасан демонстрируют существенно более низкие частоты MBL-дефицитных гаплотипов в сравнении с европеоидами Восточной Сибири (3.9, 6.4 и 21.3 % соответственно). Мы предполагаем, что изолированные арктические популяции исторически позже столкнулись с некоторыми внутриклеточными инфекциями (туберкулезом, лепрой) и, в отличие от европеоидных популяций, сохранили сформированную на ранних этапах эволюции человека высокую активность лектинового пути активации комплемента.
Ключевые слова: *MBL2*; полиморфизм генов; новорожденные; Россия; арктические популяции.

Для цитирования: Терещенко С.Ю., Смольникова М.В. Полиморфизм гена маннозосвязывающего лектина у коренных популяций территорий Арктической зоны Российской Федерации. *Вавиловский журнал генетики и селекции*. 2020;24(8):868-875. DOI 10.18699/VJ20.685

Polymorphism of the mannose-binding lectin gene in the Arctic indigenous populations of the Russian Federation

S.Yu. Tereshchenko ✉, M.V. Smolnikova

Scientific Research Institute of Medical Problems of the North, Federal Research Center "Krasnoyarsk Science Center" of the Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk, Russia
✉ e-mail: legise@mail.ru

Abstract. Mannose-binding lectin (MBL) is a pattern recognizing acute-phase protein of the innate immunity system actively involved in the elimination of a wide range of pathogenic microorganisms by activating the lectin pathway of the complement system. A significant part of the human population has a congenitally low production level and/or low MBL activity due to the carriage of various *MBL2* variants, which can modify the course of a wide range of infectious diseases. The genotype and haplotype frequencies of the *MBL2* polymorphisms have significant population differences. So far, data on the prevalence of the *MBL2* genotypes in indigenous populations of the Russian Arctic regions have not been available. The aim of the study was to analyze the frequency and ethnic specificity of the distribution of allelic variants of the *MBL2* polymorphisms rs11003125, rs7096206, rs7095891, rs5030737, rs1800450 and rs1800451 and their haplotypes in the populations of the Taimyr Dolgans-Nenets region of the Krasnoyarsk territory (Nenets, Dolgans-Nganasans, Russians). Data on the genotype and haplotype frequencies of the *MBL2* gene among

indigenous peoples of the Russian Arctic territories was first obtained in the study. The *HYPA* haplotype prevalence associated with a high concentration of MBL amounted to 35.4 % for Russian newborns in Eastern Siberia, corresponding to the one for European populations (27–33 %). In newborns of the Arctic populations, the prevalence of *HYPA* haplotype was significantly higher than in Russians and amounted to 64 % for Nenets and 56 % for the Dolgans-Nganasans, which is close to the one detected for the Eskimos and North American Indians (64–81 %). Populations of Nenets and Dolgans-Nganasans demonstrated a significantly lower prevalence of MBL-deficient haplotypes compared with Caucasians of Eastern Siberia (3.9, 6.4 and 21.3 % respectively). Isolated Arctic populations were suggested to experience some intracellular infections (tuberculosis, leprosy) historically later and, unlike Caucasoid populations, to retain the high activity of the lectin complement activation pathway formed in the early stages of human evolution.
Key words: *MBL2*; gene polymorphism; newborns; Russia; Arctic populations.

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

Введение

Система комплемента – древнейший компонент врожденного иммунитета, основной функцией которого является преимущественно интраваскулярная элиминация бактериальных агентов. Кроме того, протеины комплемента играют роль своеобразного моста между системами врожденного и адаптивного иммунитета, обеспечивая адекватные условия для созревания и дифференциации B- и T-лимфоцитов. Система комплемента состоит из плазменных протеинов и мембранных рецепторов. Плазменные протеины взаимодействуют между собой тремя известными каскадными путями – лектиновым (наиболее филогенетически древним), альтернативным и классическим.

Лектины – общий термин протеинов, формирующих отдельное суперсемейство паттерн-распознающих рецепторов, способных к распознаванию и агрегации молекул олиго- и полисахаридной природы. Среди всех лектинов уникальными функциями формирования комплексов с углеводными компонентами микробной стенки обладают фиколины (общий домен – фибриноген) и коллектины (общий домен – коллаген) – маннозосвязывающий лектин (mannose-binding lectin, MBL), печеночный и почечный коллектины (Kilpatrick, 2002; Zelensky, Gready, 2005; Bjarnadottir et al., 2016; Troldborg et al., 2017). Образование сложного комплекса полисахариды микробной стенки + коллектин/фиколин + специфические протеазы приводит в итоге к активации системы комплемента, воспалительной реакции и элиминации бактерии. Такой путь активации называется лектиновым.

MBL – классический лектин C-типа (C-type lectin), состоящий из нескольких субъединиц и склонный к олигомеризации до димеров, тримеров и тетрамеров. Способность к олигомеризации генетически детерминирована и критически повышает активность MBL в отношении связывания полисахаридов бактерий и активации комплемента (Kilpatrick, 2002). В настоящее время известно, что доминантные мутации в экзоне 1 гена *MBL*, расположенного на хромосоме 10 (10q21.1), приводят к снижению способности MBL к олигомеризации и, следовательно, к снижению его плазменной концентрации и функциональной активности. К таким однотипным последствиям приводят мутации в кодонах 52 (rs5030737; A/D), 54 (rs1800450; A/B) и 57 (rs1800451; A/C). Аллели, содержащие мутации в кодонах 52, 54 и 57, обозначают как D, B и C соответственно, в отличие от дикого аллеля A. В связи с однотипными физиологическими последствиями мутации D, B и C принято объединять и обозначать O.

Кроме кодирующих мутаций в экзоне 1, на иммунологическую функцию MBL также влияют мутации в промоторе гена: диморфизмы в локусах rs11003125 (H/L) и rs7096206 (Y/X) модулируют транскрипционную активность, значительно влияя на концентрацию MBL в плазме крови (H > L и Y > X) (Kilpatrick, 2002). Установлено, что *HY* диплотип ассоциирован с наиболее высокой плазменной концентрацией MBL, *LY* диплотип – со средней, а *LX* – с низкой (Madsen et al., 1995). Кроме того, был выявлен диморфизм в некодирующем регионе экзона 1 (rs7095891; *P/Q*).

В связи с выраженным неравновесным сцеплением все описанные мутации могут комбинироваться в ограниченное число гаплотипов из 64 возможных (*HYPA*, *LXPA*, *LYQA*, *LYPA*, *HYPD*, *LYPB*, *LYPD* и *LYQC*) (Madsen et al., 1995; Sullivan et al., 1996). Распределение частот гаплотипов гена *MBL* имеет крайне выраженные популяционные различия (Madsen et al., 1995; Boldt et al., 2006). Так, частота встречаемости гаплотипа *HYPA*, ассоциированного с высокой концентрацией MBL, варьирует от 6–8 % в африканских популяциях – Мозамбик, Кения (Madsen et al., 1995, 1998) – до 64–81 % в северных коренных популяциях – североамериканские индейцы и инуиты (Hegele et al., 1999; Best et al., 2004; Monsey et al., 2019). Европеоиды в этой градации занимают промежуточное положение с 27–30 % частотой гаплотипа *HYPA* (Skalnikova et al., 2004; Bernig et al., 2005; Steffensen et al., 2000).

Дополнительно, для оценки клинических последствий генетически детерминированных различий в экспрессии MBL было предложено выделять MBL-дефицитные (*YO*/*YO* или *XA*/*YO*), MBL-промежуточные (*YA*/*YO* или *XA*/*XA*) и MBL-высокоэкспрессирующие (*YA*/*YA* или *XA*/*YA*) диплотипы (Garred et al., 2009; Monsey et al., 2019). Принято считать, что 20–25 % всей человеческой популяции являются носителями MBL-дефицитных гаплотипов, а у 8–10 % MBL в плазме крови отсутствует или крайне низок (Madsen et al., 1995; Chalmers et al., 2013; Eisen, Osthoff, 2014).

Большинство MBL-дефицитных индивидов в целом здоровы. Явные клинические последствия MBL-дефицит имеет только в отдельных клинических ситуациях: у пациентов с нейтропенией, после трансплантации органов и тканей, у новорожденных, особенно у недоношенных

(Luo et al., 2014; Czerewaty et al., 2019). В то же время значительное количество исследований показывает, что генетически детерминированный уровень MBL может модифицировать риск возникновения и клинические характеристики многих инфекционных заболеваний. Такое влияние имеет плюрипотентный характер.

Высокий уровень MBL является протективным фактором в отношении возникновения и тяжести инфекций, вызванных инкапсулированными бактериями (*Streptococcus pneumoniae*, *Haemophilus influenzae* и *Neisseria meningitidis*), прежде всего у детей раннего возраста (Eisen et al., 2008; Tereshchenko et al., 2016). В то же время была высказана гипотеза, что нормальные/высокие уровни MBL могут повышать риск инфицирования и избыточной воспалительной реакции при инфекциях, вызванных некоторыми внутриклеточными возбудителями (*Mycobacterium tuberculosis*, *Leishmania*) (Verdu et al., 2006; Eisen, Osthoff, 2014). Следовательно, носители некоторых MBL-дефицитных гаплотипов могут иметь определенное клиническое преимущество при этих внутриклеточных инфекциях. Последние проведенные метаанализы показывают, что связь *MBL* генотипов с туберкулезом неоднозначна: некоторые генетические вариации повышают риск заболевания (rs1800450, rs5030737), а некоторые могут его снижать (rs1800451, rs7095891) (Areeshi et al., 2016; Cao et al., 2018; Tong et al., 2019). Анализ осложняет большая гетерогенность клинических форм туберкулеза в проведенных исследованиях. К тому же оценка риска в значительной мере может зависеть от этнического и возрастного состава исследованных популяций (Areeshi et al., 2016; Cao et al., 2018; Zhang et al., 2020). Насколько нам известно, к настоящему времени не опубликованы данные относительно распределения генотипов и гаплотипов гена *MBL2* в русской популяции Восточной Сибири и у коренных жителей, проживающих на территориях Арктической зоны Российской Федерации.

Материалы и методы

Для изучения однонуклеотидных полиморфизмов гена *MBL2* в Красноярском краевом консультативно-диагностическом центре медицинской генетики было получено в общей сложности 880 образцов высохших пятен крови от новорожденных из Таймырского Долгано-Ненецкого района Красноярского края. Материалом исследования послужила ДНК, выделенная из периферической крови с использованием набора DIAtom DNAPrep100 (ООО «Изоген», Россия). Новорожденные были разделены на четыре группы для изучения этнической специфики полиморфизмов *MBL2*:

- 1 – 260 человек из деревень с преимущественно ненецким населением (ненцы составляют 85 % населения);
- 2 – 110 человек из деревень с преимущественно долган-нганасанским населением (долганы-нганасаны составляют 91 % населения);
- 3 – 210 человек из деревень со смешанным населением с различной комбинацией коренных и смешанных популяций;
- 4 – 300 новорожденных из города Красноярска, имеющих европейские корни (русские составляют 91 % населения).

Table 1. Nucleotide sequences of allele-specific probes for genotyping

Polymorphism	Allele-specific probe nucleotide sequence	Fluorophore-allele
rs11003125	F - GGGCCAACGTAGTAAGAA R - GGAGTTTGCTTCCCCTTG	VIC-C/FAM-G
rs7096206	F - GCGTTGCTGCTGGAAGAC R - CAATGCACGGTCCCATTTG	VIC-G/FAM-C
rs7095891	F - GGGAAGGTTAATCTCAGTTAA R - CCAGGGATGGGTCATCTATT	VIC-A/FAM-G
rs5030737	F - CTCCAGGCATCAACGGC R - CCAACACGTACCTGGTTC	VIC-T/FAM-C

Исследование было одобрено этическим комитетом Научно-исследовательского института медицинских проблем Севера (№ 9 от 8.09.2014). Получено письменное информированное согласие на проведение исследования от родителей.

Генотипирование двух полиморфизмов rs1800450 и rs1800451 произведено с помощью рестрикционного анализа продуктов амплификации (ПДРФ-анализ) специфических участков генома. Фрагмент из 349 bp был амплифицирован с использованием пары праймеров: forward 5′-TAGGACAGAGGGCATGCTC-3′ и reverse 5′-CAGGCAGTTTCCTCTGGAAGG-3′ (температура отжига 60 °C). Эндонуклеазы рестрикции *AccB1 I* (rs1800450) и *Mbo II* (rs1800451) применяли для гидролиза амплификатов и далее фрагменты разделяли в 2 % агарозном геле с этидиумом бромидом для визуализации результатов. В случае rs1800450 полиморфизма использовали рестриктазу *AccB1 I*: фрагмент 349 bp соответствовал B аллелю, а два фрагмента 260 и 89 bp – A аллелю. В случае rs1800451 использовали *Mbo II* эндонуклеазу: фрагмент 349 bp соответствовал A аллелю, а два фрагмента 270 и 79 bp – C аллелю.

Генотипирование однонуклеотидных полиморфизмов *MBL2* rs11003125, rs7096206, rs7095891 и rs5030737 осуществлено при помощи метода ПЦР в режиме реального времени с использованием специфических олигонуклеотидных праймеров и флуоресцентно-меченных зондов (TagMan) (ООО «ДНК-синтез», Россия) по протоколу производителя (табл. 1).

Соответствие частот генотипов равновесию Харди – Вайнберга проверено с использованием χ^2. Сравнения частот генотипов проводили с использованием точного двустороннего теста Фишера. Гаплотипы оценивали и сравнивали между популяциями с использованием пакета *haplo.stats* для R среды. Для множественного тестирования применена коррекция Бонферрони. Статистически значимые различия были приняты при $p < 0.05$ после коррекции для множественного тестирования.

Результаты и обсуждение

Частоты генотипов всех включенных в исследование полиморфных участков гена *MBL2*, за исключением rs1800451, представлены в табл. 2.

Table 2. *MBL2* genotype frequencies among newborns from different ethnic populations of the Taymyr Dolgan-Nenets raion of the Krasnoyarsk oblast and of the city of Krasnoyarsk, *n* (%)

MBL2 genotype		Nenets (1), *n* = 260	Dolgans-Nganasans (2), *n* = 110	Mixed Arctic population (3), *n* = 210	Russians (4), *n* = 300
rs11003125 promoter	HH	114 (0.44)	32 (0.29)	71 (0.34)	54 (0.18)
	HL	121 (0.47)	61 (0.55)	103 (0.49)	134 (0.45)
	LL	25 (0.10)	17 (0.15)	36 (0.17)	112 (0.37)
	HWE *p*-value	0.381	0.172	0.896	0.212
rs7096206 promoter	XX	4 (0.02)	3 (0.03)	4 (0.02)	11 (0.04)
	XY	60 (0.23)	33 (0.30)	47 (0.22)	115 (0.38)
	YY	196 (0.75)	74 (0.67)	159 (0.76)	174 (0.58)
	HWE *p*-value	0.808	0.765	0.809	0.128
rs7095891 5′UTR	PP	207 (0.80)	85 (0.77)	149 (0.71)	210 (0.70)
	PQ	50 (0.19)	22 (0.20)	58 (0.28)	83 (0.28)
	QQ	3 (0.01)	3 (0.03)	3 (0.01)	7 (0.02)
	HWE *p*-value	0.992	0.296	0.316	0.720
rs5030737 exon 1	AA	252 (0.97)	110 (1.00)	201 (0.96)	265 (0.88)
	AD	8 (0.03)	0 (0.00)	9 (0.04)	32 (0.11)
	DD	0 (0.00)	0 (0.00)	0 (0.00)	3 (0.01)
	HWE *p*-value	0.801	–	0.751	0.080
rs1800450 exon 1	AA	218 (0.84)	85 (0.77)	164 (0.78)	221 (0.74)
	AB	37 (0.14)	24 (0.22)	35 (0.17)	54 (0.18)
	BB	5 (0.02)	1 (0.01)	11 (0.05)	25 (0.08)
	HWE *p*-value	0.030	0.624	0.000027	0.0000
Combined coding genotype (rs5030737, rs1800450, rs1800451)	AA	221 (0.81)	85 (0.77)	155 (0.74)	189 (0.63)
	AO	44 (0.17)	24 (0.22)	44 (0.21)	82 (0.27)
	OO	5 (0.02)	1 (0.01)	11 (0.05)	29 (0.10)
	HWE *p*-value	0.120	0.624	0.002	0.000043

Вариантный аллель C в участке rs1800451 обнаружен только в одном случае из 880 протестированных новорожденных – в гомозиготном состоянии (CC) у европеоида, проживающего в Красноярске. Среди гомозиготных вариаций изученных полиморфизмов гена *MBL2* наиболее заметные популяционные различия выявлены в промоторном регионе для участка rs11003125, где частота генотипа LL, ассоциированного с низкой продукцией MBL, в русской популяции превышала частоты в коренных популяциях Арктики в 2–3 раза: русские – 37 %, ненцы – 10 %, долганы-нганасаны – 15 % ($p_{1-2,\,3} < 0.001$).

Комбинированный аллель O был рассчитан на основании анализа мутаций в кодонах 52 (rs5030737, A/D), 54 (rs1800450, A/B) и 57 (rs1800451, A/C). Как указано выше, аллели, содержащие мутации в кодонах 52, 54 и 57, обозначены как D, B и C соответственно, в отличие от дикого аллеля A. Мутации D, B и C были закодированы и обозначены O. Частота комбинированного редкого аллеля O, он сформирован из кодирующих участков rs5030737, rs1800450 и rs1800451, в гомозиготном состоянии также была значительно выше в популяции русских новорожденных: русские – 10 %, ненцы – 2 %, долганы-нганасаны – 1 % ($p_{1-2,\,3} < 0.001$).

Наши данные о частоте гаплотипов гена *MBL2* показывают, что частота высокопродуцирующего гаплотипа *HYPA* составляет 35.4 % у русских новорожденных Восточной Сибири (табл. 3). Это соответствует частотам европейских популяций: Голландии – 27 % (Bernig et al., 2005), Дании – 30 % (Steffensen et al., 2000), Чехии – 33 % (Skalnikova et al., 2004), а также европеоидов Бразилии – 28–34 % (Boldt et al., 2006; Ferraroni et al., 2012). В то же время у новорожденных арктических популяций частота гаплотипа *HYPA* была статистически значимо выше, чем у русских, и составила 64 % для ненцев и 56 % для долган-нганасан, что приближается к значениям частот распространения, выявленных для эскимосов, – 81 % (Madsen et al., 1995; Hegele et al., 1999) и североамериканских индейцев – 64 % (Best et al., 2004). Одновременно

Table 3. *MBL2* haplotype frequencies among newborns from different ethnic populations of the Taymyr Dolgan-Nenets raion of the Krasnoyarsk oblast and of the city of Krasnoyarsk

Population	*MBL2* haplotype							
	HYPA	*LXPA*	*LYQA*	*LYPA*	*LYPB*	*LYQB*	HYPD	LYPD
Nenets (1), $n = 260$	0.638	0.127	0.100	0.026	0.070	0.007	0.015	0
Dolgans-Nganasans (2), $n = 110$	0.556	0.154	0.118	0.033	0.116	0	0	0
Mixed Arctic population (3), $n = 210$	0.551	0.118	0.120	0.044	0.100	0.025	0.015	0
Russians (4), $n = 300$	0.354	0.221	0.133	0.048	0.145	0.025	0.045	0.018
p	$p_{1-4} < 0.001$ $p_{2-4} = 0.001$ $p_{3-4} < 0.001$	$p_{1-4} = 0.022$ $p_{3-4} = 0.018$					$p_{2-4} = 0.011$	

Note. Only $p < 0.05$ values are shown in the table; when calculating the p-value, corrections were made for multiple comparisons.

Table 4. The prevalence of MBL-deficient haplotypes among newborns from different ethnic populations of the Taymyr Dolgan-Nenets raion of the Krasnoyarsk oblast and of the city of Krasnoyarsk, *n* (%)

MBL2 genotype	Nenets (1), $n = 260$	Dolgans-Nganasans (2), $n = 110$	Mixed Arctic population (3), $n = 210$	Russians, Krasnoyarsk (4), $n = 300$	p
Deficient	10 (3.9)	7 (6.4)	19 (9.1)	64 (21.3)	$p_{1, 2, 3-4} < 0.001$ $p_{1-3} = 0.02$
Intermediate	43 (16.5)	21 (19.1)	39 (18.6)	58 (19.3)	–
Sufficient	207 (79.6)	82 (74.6)	152 (72.4)	178 (59.3)	$p_{1-4} < 0.001$ $p_{2-4} = 0.005$ $p_{3-4} = 0.002$

у новорожденных российских арктических популяций закономерно зарегистрированы низкие частоты MBL-дефицитного гаплотипа *LXPA* (см. табл. 3). Наибольшие различия в частотах указанных гаплотипов были характерны для ненецкой популяции.

В табл. 4 суммированы данные о частотах MBL-дефицитных гаплотипов в изученных популяциях. Выделены MBL-дефицитные (*YO*/*YO* или *XA*/*YO*), MBL-промежуточные (*YA*/*YO* или *XA*/*XA*) и MBL-высокоэкспрессирующие (*YA*/*YA* или *XA*/*YA*) гаплотипы.

Популяции ненцев и долган-нганасан демонстрируют существенно более низкие частоты MBL-дефицитных гаплотипов в сравнении с европеоидами Восточной Сибири (3.9, 6.4 и 21.3 % соответственно, $p < 0.001$). Смешанная арктическая популяция демонстрирует промежуточное значение частоты – 9.1 %. На популяционном уровне клинические последствия врожденно высокой способности продукции функционально активных форм MBL у представителей арктических популяций заключаются в низком риске тяжелых бактериальных инфекций в раннем возрасте и, вероятно, более высоком риске туберкулеза в старшем возрасте, что предполагается многими исследователями (Eisen, Osthoff, 2014; Tong et al., 2019). Кроме того, низкая частота атеросклероза и кардиоваскулярных заболеваний среди коренных жителей Арктики, наряду с такими факторами, как высокое употребление омега-3 жирных кислот и особенности стиля жизни, может быть обусловлена и генетическими особенностями продукции и активности MBL. Вероятность такой связи показана в целом ряде публикаций (Hegele et al., 1999; Best et al., 2004; Fumagalli et al., 2017; Monsey et al., 2019).

В настоящем исследовании нами впервые получены данные о частотах генотипов и гаплотипов гена *MBL2* среди коренных народностей, проживающих на территориях Арктической зоны Российской Федерации. Ранее нами была показана большая частота распространенности генотипов, ассоциированных с высокой активностью L-фиколина, в арктических популяциях ненцев и долган-нганасан, в сравнении с европеоидами Восточной Сибири (Smolnikova et al., 2017). Таким образом, популяции коренных народов Арктики генетически характеризуются большей активностью как минимум двух различающихся компонентов лектинового пути активации комплемента – MBL и L-фиколина. Определенное преимущество нашего подхода к популяционной оценке распространенности MBL- и L-фиколин генотипов состоит в исследовании популяций новорожденных, когда еще не произошло вероятное выбывание неблагоприятных генетических вариаций, возможное в более старшем возрасте.

В настоящее время существуют две конкурирующие гипотезы, пытающиеся объяснить высокий уровень популяционного разнообразия генотипов *MBL2* с высоким накоплением MBL-дефицитных вариантов (Eisen, Osthoff, 2014).

Первая из них предполагает протективную роль низкопродуцирующих генотипов в отношении некоторых

внутриклеточных возбудителей: туберкулеза и лепры (микобактерии), висцерального лейшманиоза (род внутриклеточных паразитов *Leishmania*), атипичной пневмонии (внутриклеточные бактерии *Mycoplasma pneumoniae, Chlamydophila pneumoniae*, *Legionella pneumophila*, *Coxiella burnetii*). Впервые предположение о роли позитивной селекции в накоплении низкопродуцирующих генотипов высказано P. Garred и его сотрудниками в 1994 г. Они установили, что у пациентов с лепрой (возбудитель *Mycobacterium leprae*) уровень MBL в сыворотке крови был выше, чем у здоровых доноров той же популяции (Garred et al., 1994). В 1999 г. E.G. Hoal-Van Helden с коллегами показали протективную роль MBL-низкопродуцирующего аллеля B полиморфного участка rs1800450 гена *MBL* в отношении туберкулезного менингита (Hoal-Van Helden et al., 1999). Последние проведенные метаанализы также подтверждают роль полиморфизма *MBL2* при формировании туберкулезной инфекции (Areeshi et al., 2016; Cao et al., 2018; Tong et al., 2019). В 2001 г. I.K. Santos с сотрудниками показали, что MBL-низкопродуцирующий вариантный генотип OO встречался реже у пациентов с висцеральным лейшманиозом (Santos et al., 2001). В дальнейшем эти данные были подтверждены: риск висцерального лейшманиоза был значительно повышен у лиц с генетическими вариантами, ассоциированными с высокой продукцией MBL (Alonso et al., 2007). Наконец, недавнее проспективное исследование датской когорты пациентов с внебольничной пневмонией (n = 505) показало большую предрасположенность лиц с генетической детерминированной высокой продукцией основных факторов лектинового пути активации комплемента MBL и L-фиколина к внутриклеточным респираторным инфекциям: *Mycoplasma pneumoniae*, *Chlamydophila pneumoniae*, *Legionella pneumophila*, *Coxiella burnetii* (Van Kempen et al., 2017). Большинство исследователей считают, что высокий уровень лектин-опосредованного фагоцитоза может предрасполагать к более успешному проникновению внутриклеточных возбудителей в цитоплазму клеток хозяина, экранированию патогенов от факторов адаптивного иммунитета и, следовательно, большему риску формирования активного инфекционного процесса.

Кроме того, в ряде работ показано, что MBL-дефицит может быть протективным фактором в отношении атеросклероза и ассоциированных кардиоваскулярных заболеваний (Hegele et al., 1999; Best et al., 2004; Fumagalli et al., 2017; Monsey et al., 2019). Секвенирование генома 102 жителей США, представителей четырех основных этнических групп, показало наличие признаков селективного отбора в сторону большего накопления гетерозигот гена *MBL2* (Bernig et al., 2004).

Имеющиеся к настоящему времени фактические данные позволяют говорить о «двойной патофизиологической» роли лектинового пути активации комплемента: протективной – в отношении внеклеточных возбудителей, особенно у детей раннего возраста, и провокативной – в отношении некоторых внутриклеточных возбудителей и атеросклероза. Популяционно-генетические последствия такой «двойной» роли могут лежать в основе этнического разнообразия соответствующих генотипов, что представляет собой суть первой упомянутой нами гипотезы, основанной на предположении селекционной выгоды MBL-дефицита для некоторых популяций (Seyfarth et al., 2005; Eisen, Osthoff, 2014). В русле указанной гипотезы было высказано предположение, объясняющее низкую частоту MBL-дефицита среди арктических народностей и генетически близких к ним коренных североамериканских индейцев. Принято считать, что эти популяции исторически позже встретились с туберкулезом и лепрой, не сталкивались с возбудителями лейшманиоза и реже имели классические факторы риска атеросклероза: диабет, гиперлипидемию, а также, возможно, хроническое инфицирование *Chlamydophila pneumoniae* (Hegele et al., 1999; Best et al., 2004; Monsey et al., 2019). Следовательно, именно в этих популяциях не происходила характерная, согласно этой гипотезе, позитивная селекция MBL-дефицитных генотипов.

Вторая гипотеза отрицает наличие селекционного давления в отношении *MBL2* генотипов, объясняя генетическое разнообразие исключительно миграционными процессами и генетическим дрейфом. Так, исследование 1116 индивидов из различных географических регионов не выявило статистических признаков селективного отбора (Verdu et al., 2006). Такие же результаты были получены при статистической обработке данных различных популяций Бразилии и сравнительном изучении жителей Габона и Европы (Boldt et al., 2006, 2010). Исследование *MBL2* полиморфизма у детей Мозамбика показало отсутствие статистических признаков позитивной или балансирующей селекции (Valles et al., 2009). Впрочем, некоторые авторы делают при обсуждении собственных результатов оговорку: «Возможно, стохастические эволюционные факторы стерли большую часть древнего отпечатка, оставленного естественным селекционным отбором; для подтверждения данных требуются статистически более мощные исследования с включением большего числа популяций» (Boldt et al., 2006).

Заключение

Таким образом, по результатам настоящего исследования нами показана большая частота встречаемости MBL-высокопродуцирующих генетических вариаций в популяциях коренных арктических народностей, проживающих в Таймырском Долгано-Ненецком районе Красноярского края. Рассматривая представленные данные в совокупности с ранее опубликованными результатами полиморфизма гена L-фиколина в тех же популяциях (Smolnikova et al., 2017), можно говорить не только о накоплении отдельных генотипов *MBL2* и *FCN2* в арктических популяциях, но и о большем тонусе лектинового пути активации комплемента в целом. Указанные факты позволяют нам осторожно высказаться о гипотезе селективного популяционного давления в отношении лектинового пути активации комплемента как общего патофизиологического механизма, опосредованного генами *MBL2* и *FCN2*, и, вероятно, ассоциированного с предрасположенностью к некоторым инфекциям. Мы считаем, что изолированные арктические популяции исторически позже столкнулись с некоторыми внутриклеточными инфекциями (микобактериями, возможно *Chlamydophila pneumoniae*) и вследствие этого сохранили сформированную на ранних этапах

эволюции человека высокую активность лектинового пути активации комплемента. Безусловно, эта гипотеза требует дополнительной верификации в специально организованных исследованиях большей статистической мощности с использованием всего арсенала методов популяционной генетики.

Список литературы / References

Alonso D.P., Ferreira A.F., Ribolla P.E., Santos I.M., Cruz M.P., Carvalho F., Abatepaulo A.R., Costa D., Werneck G.L., Farias T., Soares M.J., Costa C.H. Genotypes of the mannan-binding lectin gene and susceptibility to visceral leishmaniasis and clinical complications. *J. Infect. Dis.* 2007;195(8):1212-1217. DOI 10.1086/512683.

Areeshi M.Y., Mandal R.K., Akhter N., Dar S.A., Jawed A., Wahid M., Mahto H., Panda A.K., Lohani M., Haque S. A meta-analysis of *MBL2* polymorphisms and tuberculosis risk. *Sci. Rep.* 2016;6: 35728. DOI 10.1038/srep35728.

Bernig T., Breunis W., Brouwer N., Hutchinson A., Welch R., Roos D., Kuijpers T., Chanock S. An analysis of genetic variation across the *MBL2* locus in Dutch Caucasians indicates that 3′ haplotypes could modify circulating levels of mannose-binding lectin. *Hum. Genet.* 2005;118(3-4):404-415. DOI 10.1007/s00439-005-0053-5.

Bernig T., Taylor J.G., Foster C.B., Staats B., Yeager M., Chanock S.J. Sequence analysis of the mannose-binding lectin (*MBL2*) gene reveals a high degree of heterozygosity with evidence of selection. *Genes Immun.* 2004;5(6):461-476. DOI 10.1038/sj.gene.6364116.

Best L.G., Davidson M., North K.E., Maccluer J.W., Zhang Y., Lee E.T., Howard B.V., Decroo S., Ferrell R.E. Prospective analysis of mannose-binding lectin genotypes and coronary artery disease in American Indians: the Strong Heart Study. *Circulation.* 2004;109(4): 471-475. DOI 10.1161/01.CIR.0000109757.95461.10.

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

Boldt A.B., Culpi L., Tsuneto L.T., De Souza I.R., Kun J.F., Petzl-Erler M.L. Diversity of the *MBL2* gene in various Brazilian populations and the case of selection at the mannose-binding lectin locus. *Hum. Immunol.* 2006;67(9):722-734. DOI 10.1016/j.humimm.2006.05.009.

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

Cao Y., Wang X., Cao Z., Wu C., Wu D., Cheng X. Genetic polymorphisms of *MBL2* and tuberculosis susceptibility: a meta-analysis of 22 case-control studies. *Arch. Med. Sci.* 2018;14(6):1212-1232. DOI 10.5114/aoms.2017.65319.

Chalmers J.D., Mchugh B.J., Doherty C., Smith M.P., Govan J.R., Kilpatrick D.C., Hill A.T. Mannose-binding lectin deficiency and disease severity in non-cystic fibrosis bronchiectasis: a prospective study. *Lancet Respir. Med.* 2013;1(3):224-232. DOI 10.1016/S2213-2600(13)70001-8.

Czerewaty M., Tarnowski M., Safranow K., Domanski L., Pawlik A. Mannose binding lectin 2 gene polymorphisms in patients after renal transplantation with acute graft rejection. *Transpl. Immunol.* 2019; 54:29-37. DOI 10.1016/j.trim.2019.01.004.

Eisen D.P., Dean M.M., Boermeester M.A., Fidler K.J., Gordon A.C., Kronborg G., Kun J.F., Lau Y.L., Payeras A., Valdimarsson H., Brett S.J., Ip W.K., Mila J., Peters M.J., Saevarsdottir S., Van Till J.W., Hinds C.J., Mcbryde E.S. Low serum mannose-binding lectin level increases the risk of death due to pneumococcal infection. *Clin. Infect. Dis.* 2008;47(4):510-516. DOI 10.1086/590006.

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

Ferraroni N.R., Segat L., Guimaraes R.L., Brandao L.A., Crovella S., Constantino-Silva R.N., Loja C., Da Silva Duarte A.J., Grumach A.S. Mannose-binding lectin and MBL-associated serine protease-2 gene polymorphisms in a Brazilian population from Rio de Janeiro. *Int. J. Immunogenet.* 2012;39(1):32-38. DOI 10.1111/j.1744-313X. 2011. 01052.x.

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

Garred P., Harboe M., Oettinger T., Koch C., Svejgaard A. Dual role of mannan-binding protein in infections: another case of heterosis? *Eur. J. Immunogenet.* 1994;21(2):125-131. DOI 10.1111/j.1744-313x.1994.tb00183.x.

Garred P., Honore C., Ma Y.J., Munthe-Fog L., Hummelshoj T. *MBL2*, *FCN1*, *FCN2* and *FCN3* – The genes behind the initiation of the lectin pathway of complement. *Mol. Immunol.* 2009;46(14):2737-2744. DOI 10.1016/j.molimm.2009.05.005.

Hegele R.A., Busch C.P., Young T.K., Connelly P.W., Cao H. Mannose-binding lectin gene variation and cardiovascular disease in Canadian Inuit. *Clin. Chem.* 1999;45(8 Pt 1):1283-1285.

Hoal-Van Helden E.G., Epstein J., Victor T.C., Hon D., Lewis L.-A., Beyers N., Zurakowski D., Ezekowitz R.a.B., Van Helden P.D. Mannose-binding protein B allele confers protection against tuberculous meningitis. *Pediatr. Res.* 1999;45(4):459-464. DOI 10.1203/00006450-199904010-00002.

Kilpatrick D. Mannan-binding lectin: clinical significance and applications. *Biochim. Biophys*. *Acta Gen. Subj.* 2002;1572(2-3):401-413. DOI 10.1016/s0304-4165(02)00321-5.

Luo J., Xu F., Lu G.J., Lin H.C., Feng Z.C. Low mannose-binding lectin (MBL) levels and MBL genetic polymorphisms associated with the risk of neonatal sepsis: An updated meta-analysis. *Early Hum. Dev.* 2014;90(10):557-564. DOI 10.1016/j.earlhumdev.2014.07.007.

Madsen H.O., Garred P., Thiel S., Kurtzhals J.A., Lamm L.U., Ryder L.P., Svejgaard A. Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein. *J. Immunol.* 1995;155(6):3013-3020.

Madsen H.O., Satz M.L., Hogh B., Svejgaard A., Garred P. Different molecular events result in low protein levels of mannan-binding lectin in populations from southeast Africa and South America. *J. Immunol.* 1998;161(6):3169-3175.

Monsey L., Best L.G., Zhu J., Decroo S., Anderson M.Z. The association of mannose binding lectin genotype and immune response to *Chlamydia pneumoniae*: The Strong Heart Study. *PLoS One.* 2019; 14(1):e0210640. DOI 10.1371/journal.pone.0210640.

Santos I.K., Costa C.H.N., Krieger H., Feitosa M.F., Zurakowski D., Fardin B., Gomes R.B.B., Weiner D.L., Harn D.A., Ezekowitz R.A.B., Epstein J.E. Mannan-binding lectin enhances susceptibility to visceral leishmaniasis. *Infect. Immun.* 2001;69(8):5212-5215. DOI 10.1128/iai.69.8.5212-5215.2001.

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

Skalnikova H., Freiberger T., Chumchalova J., Grombirikova H., Sediva A. Cost-effective genotyping of human *MBL2* gene mutations using multiplex PCR. *J. Immunol. Methods.* 2004;295(1-2):139-147. DOI 10.1016/j.jim.2004.10.007.

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

Steffensen R., Thiel S., Varming K., Jersild C., Jensenius J.C. Detection of structural gene mutations and promoter polymorphisms in the mannan-binding lectin (MBL) gene by polymerase chain reaction with sequence-specific primers. *J. Immunol. Methods.* 2000; 241(1-2):33-42. DOI 10.1016/s0022-1759(00)00198-8.

Sullivan K.E., Wooten C., Goldman D., Petri M. Mannose-binding protein genetic polymorphisms in black patients with systemic lupus erythematosus. *Arthritis Rheumatol.* 1996;39(12):2046-2051. DOI 10.1002/art.1780391214.

Tereshchenko S.Y., Kasparov E.V., Smol'nikova M.V., Kuvshinova E.V. Mannose-binding lectin deficiency in respiratory diseases. *Rus. Pulmonol.* 2016;26(6):748-752. DOI 10.18093/0869-0189-2016-26-6-748-752.

Tong X., Wan Q., Li Z., Liu S., Huang J., Wu M., Fan H. Association between the mannose-binding lectin (MBL)-2 gene variants and serum MBL with pulmonary tuberculosis: An update meta-analysis and systematic review. *Microb. Pathog.* 2019;132:374-380. DOI 10.1016/j.micpath.2019.04.023.

Troldborg A., Hansen A., Hansen S.W., Jensenius J.C., Stengaard-Pedersen K., Thiel S. Lectin complement pathway proteins in healthy individuals. *Clin. Exp. Immunol.* 2017;188(1):138-147. DOI 10.1111/cei.12909.

Valles X., Sarrias M.R., Casals F., Farnos M., Piner R., Suarez B., Morais L., Mandomando I., Sigauque B., Roca A., Alonso P.L., Torres A., Thielens N.M., Lozano F. Genetic and structural analysis of *MBL2* and *MASP2* polymorphisms in South-Eastern African children. *Tissue Antigens.* 2009;74(4):298-307. DOI 10.1111/j.1399-0039.2009.01328.x.

Van Kempen G., Meijvis S., Endeman H., Vlaminckx B., Meek B., De Jong B., Rijkers G., Bos W.J. Mannose-binding lectin andl-ficolin polymorphisms in patients with community-acquired pneumonia caused by intracellular pathogens. *Immunol.* 2017;151(1):81-88. DOI 10.1111/imm.12705.

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

Zelensky A.N., Gready J.E. The C-type lectin-like domain superfamily. *FEBS J.* 2005;272(24):6179-6217. DOI 10.1111/j.1742-4658.2005.05031.x.

Zhang J.X., Gong W.P., Zhu D.L., An H.R., Yang Y.R., Liang Y., Wang J., Tang J., Zhao W.G., Wu X.Q. Mannose-binding lectin 2 gene polymorphisms and their association with tuberculosis in a Chinese population. *Infect. Dis. Poverty.* 2020;9(1):46. DOI 10.1186/s40249-020-00664-9.

ORCID ID

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

Acknowledgements. This work was supported by the Presidium of the Russian Academy of Sciences, basic research program "Pilot basic research targeted at the development of the Arctic belt of the Russian Federation", 2016–2019; and State Contract register no. NIOKTR 1201351112 "Analysis of prevalence and risk factors and track of major psychosomatic disorders in Siberian children and adolescents".
The authors are grateful to Dr. Maxim B. Freydin, Senior Researcher of the Laboratory of Populational Genetics, Research Institute for Medicinal Genetics, Tomsk National Medical Research Center of the Russian Academy of Sciences, for haplotype analysis and valuable remarks on the study.

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

Received July 8, 2020. Revised August 27, 2020. Accepted August 27, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):876-884
DOI 10.18699/VJ20.686

The GWAS-MAP platform for aggregation of results of genome-wide association studies and the GWAS-MAP|homo database of 70 billion genetic associations of human traits

T.I. Shashkova[1], D.D. Gorev[1], E.D. Pakhomov[1, 2], A.S. Shadrina[1], S.Zh. Sharapov[1], Y.A. Tsepilov[1], L.C. Karssen[2], Y.S. Aulchenko[1]

[1] Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University, Novosibirsk, Russia
[2] PolyKnomics BV, 's-Hertogenbosch, The Netherlands
e-mail: y.aulchenko@nsu.ru

Abstract. Hundreds of genome-wide association studies (GWAS) of human traits are performed each year. The results of GWAS are often published in the form of summary statistics. Information from summary statistics can be used for multiple purposes – from fundamental research in biology and genetics to the search for potential biomarkers and therapeutic targets. While the amount of GWAS summary statistics collected by the scientific community is rapidly increasing, the use of this data is limited by the lack of generally accepted standards. In particular, the researchers who would like to use GWAS summary statistics in their studies have to become aware that the data are scattered across multiple websites, are presented in a variety of formats, and, often, were not quality controlled. Moreover, each available summary statistics analysis tools will ask for data to be presented in their own internal format. To address these issues, we developed GWAS-MAP, a high-throughput platform for aggregating, storing, analyzing, visualizing and providing access to a database of big data that result from region- and genome-wide association studies. The database currently contains information on more than 70 billion associations between genetic variants and human diseases, quantitative traits, and "omics" traits. The GWAS-MAP platform and database can be used for studying the etiology of human diseases, building predictive risk models and finding potential biomarkers and therapeutic interventions. In order to demonstrate a typical application of the platform as an approach for extracting new biological knowledge and establishing mechanistic hypotheses, we analyzed varicose veins, a disease affecting on average every third adult in Russia. The results of analysis confirmed known epidemiologic associations for this disease and led us to propose a hypothesis that increased levels of MICB and CD209 proteins in human plasma may increase susceptibility to varicose veins.
Key words: database; genome-wide association studies; quantitative genetics; varicose veins; GWAS-MAP.

For citation: Shashkova T.I., Gorev D.D., Pakhomov E.D., Shadrina A.S., Sharapov S.Zh., Tsepilov Y.A., Karssen L.C., Aulchenko Y.S. The GWAS-MAP platform for aggregation of results of genome-wide association studies and the GWAS-MAP|homo database of 70 billion genetic associations of human traits. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):876-884. DOI 10.18699/VJ20.686

Платформа GWAS-MAP для агрегации результатов полногеномных исследований ассоциаций и база данных GWAS-MAP|homo 70 миллиардов генетических ассоциаций признаков человека

Т.И. Шашкова[1], Д.Д. Горев[1], Е.Д. Пахомов[1, 2], А.С. Шадрина[1], С.Ж. Шарапов[1], Я.А. Цепилов[1], Л.К. Карссен[2], Ю.С. Аульченко[1]

[1] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[2] PolyKnomics BV, Хертогенбос, Нидерланды
e-mail: y.aulchenko@nsu.ru

Аннотация. Ежегодно проводятся сотни полногеномных исследований ассоциаций (genome-wide association studies, GWAS) человеческих признаков. Результаты GWAS часто публикуют в форме сводных статистик. Информацию из сводных статистик можно использовать для решения различных задач – от фундаментальных исследований в области биологии и генетики до поиска потенциальных биомаркеров и мишеней терапевтического воздействия. Количество собранных научным сообществом сводных статистик GWAS быстро растет, однако использование данных затруднено из-за отсутствия общепринятых стандартов. В частности, исследователи, которые хотели бы применить сводные статистики GWAS в своей работе, сталкиваются с тем, что

данные разбросаны по нескольким веб-сайтам, представлены в различных форматах, нередко без контроля качества. Более того, каждый доступный инструмент анализа сводных статистик запрашивает данные в своем собственном внутреннем формате. Для решения этих проблем мы разработали высокопроизводительную платформу GWAS-MAP для агрегации, хранения, анализа, визуализации и доступа к базе данных результатов полногеномных и региональных исследований ассоциаций. В настоящий момент на платформе содержится информация о более чем 70 миллиардах ассоциаций между вариантами геномной последовательности и болезнями, количественными и «омиксными» признаками человека. Платформа и база данных могут использоваться для изучения этиологии заболеваний человека, разработки предиктивных моделей риска, а также для поиска потенциальных биомаркеров и терапевтических воздействий. Применение платформы как инструмента для извлечения новых биологических знаний и формулировки гипотез о механизмах генетического контроля продемонстрировано на примере варикозной болезни нижних конечностей, заболевания, встречающегося у каждого третьего взрослого жителя России. Результаты проведенного анализа подтвердили известные эпидемиологические ассоциации для данного заболевания и позволили выдвинуть гипотезу о том, что уровень белков MICB и CD209 в плазме крови человека может влиять на риск варикозного расширения вен.
Ключевые слова: база данных; полногеномное исследование ассоциаций; количественная генетика; варикозная болезнь нижних конечностей; GWAS-MAP.

Introduction

Genome-wide association studies (GWAS) are one of the main approaches for identifying associations between genetic variants and traits (Visscher et al., 2017). One of the most important advantages of this approach is that it is agnostic to the molecular mechanisms or biochemical nature of the traits or diseases under study, thus allowing fundamentally new knowledge to be obtained. Based on the functions of the genes mapped by a GWAS, researchers aim to discover new molecular mechanisms underlying the development of traits and pathologies under consideration.

GWAS are performed on large samples of genotyped and phenotyped individuals to identify statistically significant associations between single-nucleotide polymorphisms (SNPs) and traits (Bush, Moore, 2012). SNPs are located relatively homogeneously and with sufficient density, consequently, functional variants occurring at high frequency in the population are detected with a high probability, either because the causative allele is being tested directly or because it is in linkage disequilibrium with genotyped markers. A special case of GWAS is a regional genetic study of associations or a region-wide association study (RWAS), where the analysis is applied to SNPs in a particular region instead of the whole genome. RWAS is used, for example, to find cis-SNPs associated with the expression of a certain gene (GTEx Consortium et al., 2017).

The GWAS approach has become very popular over the past decade. Since 2007 the number of GWAS has increased exponentially and hundreds of original genome-wide studies are published every year. The earliest GWAS addressed the associations between a single trait and several hundreds of thousands SNPs, using samples of several hundreds or thousands individuals (Klein, 2005; International Schizophrenia Consortium et al., 2009).

Currently, both the number of analyzed traits and the genomic coverage of GWAS have increased by many orders of magnitude (Timmers et al., 2019). This has become possible due to the advent of new sequencing and genotyping technologies and the improvement of existing ones, as well as other methods for studying biological objects, leading to an increase in the resolution of sequencing, genotyping and phenotyping. Modern GWAS normally assess associations with millions of SNPs and in some cases the sample size exceeds one million people (Timmers et al., 2019). The number of phenotypes studied can go to hundreds (Demirkan et al., 2012; Shen et al., 2017), thousands (Sun et al., 2018) and even tens of thousands (GTEx Consortium et al., 2017), e. g. for "-omics" traits. The same trait is analyzed in multiple studies, often with progressively increasing sample sizes, as well as in new populations, offering increased power and generalizability.

The direct results of GWAS/RWAS consist of files with summary statistics. These files can include up to ten of millions rows, where each row contains information about the association between a given SNP and the investigated trait. Taking into account the number of GWAS studies and the size of the files with results, GWAS results qualify as Big Data (Wu et al., 2013; Fabregat-Traver et al., 2014). Importantly, not only does this body of data grow, but so do the rates of data acquisition.

GWAS results can be used to address a large number of problems ranging from fundamental biology and genetics to the search for biomarkers and targets for therapeutic interventions. Currently, a range of methods has been developed that implement the solution of these problems based on summary statistics data.

In particular, methods have been developed to define sets of SNPs that are most likely to contain the true functional variant at loci suggested by GWAS (Kichaev et al., 2014; Benner et al., 2016; Schaid et al., 2018). For example, this problem is addressed by the PAINOR (Kichaev et al., 2014) software and the conditional and joint analysis as implemented in the GCTA tool (Yang et al., 2011).

Also, identification of causal genes influencing a trait of interest is possible through the use of summary statistics (Giambartolomei et al., 2014; Zhu et al., 2016; Momozawa et al., 2018). By regulating the expression of those genes or by manipulating their products through the use of, for example, pharmacological interventions, the trait of interest can be addressed in a targeted manner. Several instruments implement these methods, for example, the SMR (Summary-level Mendelian Randomization) tool (Zhu et al., 2016). The same methods can often be used to study pleiotropic effects (Klarić et al., 2020; Shadrina et al., 2020). The results of studies of pleiotropy can be used for drug repositioning, for predicting

possible side effects of gene editing, and for prediction of possible side effects of pharmacological manipulation of the products of these genes.

Over the past decade, the number of studies using Mendelian randomization methods has increased substantially, providing important new information about disease etiology (Elgaeva et al., 2019). Mendelian randomization methods combined with the use of summary statistics and multiple instrumental variables (Hemani et al., 2016; O'Connor, Price, 2018) can help to reconstruct the theoretical hierarchy of cause and effect relationships between traits and, in practice, have the potential to be used for the identification of traits that can be targeted by therapeutic interventions.

Methods for studying genetic correlations (Bulik-Sullivan et al., 2015; Speed, Balding, 2019) can be particularly useful in addressing fundamental questions related to the genetic architecture of complex traits. One of these methods is implemented in a popular LDsr (Linkage Disequilibrium score regression) python package (Bulik-Sullivan et al., 2015).

Finally, summary statistics from GWAS can be used in methods to develop models for the prediction of quantitative traits and disease risks for a given individual or group (Mak et al., 2017; Choi, O'Reilly, 2019; Lloyd-Jones et al., 2019). The simplest of these models use effects of the most significant independent SNPs (Evans et al., 2009). If a GWAS involves a large number of cases and controls, powerful predictors can be developed for some traits even with simple models, breast cancer being a well-known example (Mavaddat et al., 2019). Methods allowing the researcher to manage information about millions of SNPs and whole-genome LD structure while developing a prediction model have recently become popular (Vilhjálmsson et al., 2015). Such models were used for predicting the risk of ischemic heart disease (Khera et al., 2018), type 2 diabetes (Khera et al., 2018), and obesity (Khera et al., 2019).

Although the amount of GWAS results obtained by the scientific community is constantly growing, as are the number of methods for their analysis, they have currently found only limited use. The problems researchers face when working with these data are multiple. First, summary statistics files from GWAS are large (more than tens of terabytes), and so their storage and processing require dedicated infrastructure. Secondly, data are produced by different laboratories using different protocols, and consequently, quality control and a harmonization procedure for storing such data in a common format are required. Thirdly, the existing tools for analyses of summary statistics data from GWAS are implemented using different languages, hosted at different repositories and websites and require custom input data formats. Finally, large-scale adoption of these methods and data require user interfaces for researchers without specialized bioinformatics skills.

The existing solutions are incomplete or partial. On the one hand, resources such as GWAS Central (https://www.gwascentral.org) or GWAS Catalog (https://www.ebi.ac.uk/gwas/) can do as much (that is, aggregate, store and provide access to GWAS results), but originally they were intended for handling "small data" (that is, the most statistically significant associations), and so their architecture does not scale well enough to handle big data and the requirements of new methods for processing GWAS results. On the other hand, most software applications (for example, SMR (Zhu et al., 2016), GCTA (Yang et al., 2011), LDsr (Bulik-Sullivan et al., 2015)) are intended for analyzing data rather than for aggregating, storing or providing access to them. Finally, the portals MR-Base (http://www.mrbase.org/) and LD Hub (http://ldsc.broadinstitute.org/) can aggregate and store GWAS results and allow the user to conduct specify types of analysis; however, these portals do not offer anything for other methods of analysis, and software solutions for data aggregation and storage are not available.

To help address these issues, we developed the GWAS-MAP platform for aggregating, storing, analyzing, visualizing and providing access to big data obtained from GWAS. The name GWAS-MAP means both a map between phenotypes and genotypes, but is also an abbreviation of Multiple Analyses Platform. Using GWAS-MAP we collected GWAS-MAP|homo database of GWAS and RWAS results for human traits. Currently, the database contains more than 70 billion associations between SNPs and human traits. GWAS-MAP provides an opportunity to carry out research that will contribute to the search for new biomarkers that has a bearing on the development of high-efficacy drugs and also reveal side effects in existing drugs. We have performed a genetic analysis of varicose veins to demonstrate how the platform works.

The GWAS-MAP platform

GWAS-MAP software architecture

The GWAS-MAP platform consists of two data processing modules (one for integration and one for analysis of GWAS/RWAS results) and a database (DB) module (see the Figure).

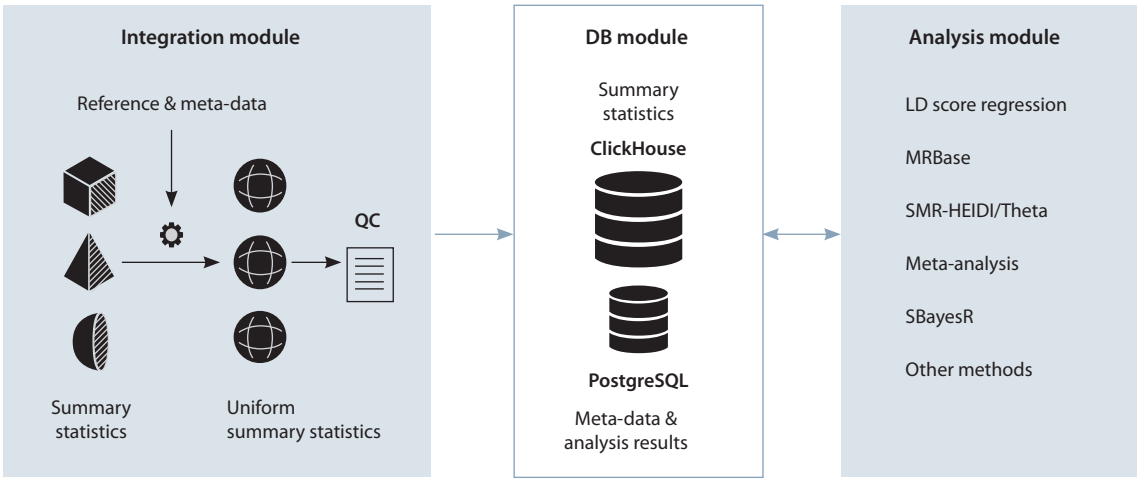
Data integration starts with the conversion of summary statistics files collected from various sources into a universal data format. After conversion, we perform quality control (QC) and if the summary statistics pass, they are uploaded to the databases.

The DB module is the part responsible for setting up the databases and tables structure required for the GWAS-MAP platform. The DB module consists of two components, each controlled by a separate open source database management system (DBMS). One of the components is used to store the GWAS summary statistics; for this component the ClickHouse DBMS version 19.16.2. revision 54427 (https://clickhouse.tech/) is used. A record in this system contains the parameters of association between certain SNP and a trait. The other component contains (1) meta-data that gives particular information about the summary statistics collected from articles, study web-sites or other sources, and (2) the results of analyses; for this component the PostgreSQL DBMS version 10.6 (https://www.postgresql.org/) is used.

With the analysis module a user can run various analyses on the GWAS/RWAS summary statistics using the integrated analytical tools written in Python, which are accessible through command-line utilities.

Integration and quality control of GWAS/RWAS results

The platform offers users the option to upload GWAS summary statistics files of their own original research. Because these data were generated using different protocols, the resulting summary statistics files may appear in different formats.



GWAS-MAP software architecture.

Grey blocks: data processing modules. White block: database module with database management systems. Arrows between modules: saving/retrieving data to/from the DB. QC, quality control; DB, database; LD, linkage disequilibrium. Description of the software modules is provided in the text.

To address this, GWAS-MAP provides an integration module converting summary statistics files to a common format and performing QC on the data.

To ensure data consistency within the DB, information about a SNP's identifier, position in the genome, and alleles and allele frequencies are compared with reference data. The reference is a list of SNPs with their main characteristics: the identifier (rsID), chromosome, position, alleles and allele frequencies. At present, the reference is based on the 503 genomes of European-ancestry individuals from the "1000 Genomes" project phase 3 version 5 (The 1000 Genomes Project Consortium et al., 2015).

In general, summary statistics contain all fields required for unification in a universal format. If some fields are absent, then the missing information is added from the reference (for example, allele frequency) or calculated from the information in the input file. For example, it is possible to recover the standard error of the effect size based on the effect size and *p*-value.

Before uploading GWAS/RWAS data to the DB, it is absolutely necessary to have them passed through QC. QC is indispensable not only for meta-analyses of GWAS/RWAS results, but also for verifying separate studies, because seemingly insignificant data errors may lead to heavily biased results later on.

We have developed a QC module which spots outlying SNPs (that is, those with characteristics other than expected) and assesses the overall quality of the input data. More specifically, QC includes (1) a comparison of the frequencies of alleles from the input data with those from the reference set, a comparison of the *p*-values provided in the study and those calculated from the *Z*-statistics (if present), (2) an analysis of the distribution of estimates of the allele effect sizes, (3) calculation of the trait variance and (4) genomic control factor (λ_{GC}). SNPs whose characteristics depart by more than a threshold value from those expected are labeled as outliers and can be filtered out by the user. If the summary statistics from GWAS have more than 5 % outliers, or the effect size distribution is not symmetric, this data will be not recommended for upload, although the final decision is up to a user. We should notice that all current data in DB have passed the above described criteria.

Analysis methods using GWAS/RWAS summary statistics implemented in GWAS-MAP

GWAS-MAP incorporates several widely used methods for the analysis of GWAS/RWAS summary statistics with special emphasis on the identification of genes, molecules, traits and functional SNPs that appear as potential targets of therapeutic interventions. In particular, data processing can be carried out using the following methods.

1. Linkage disequilibrium score regression is a method to assess the heritability of a trait and to calculate genetic correlations between two traits (Bulik-Sullivan et al., 2015). This method was implemented in Python 2 by Bulik-Sullivan and co-authors (2015). We have re-written it in Python 3 because it is the main programming language used for GWAS-MAP and because Python 2 has been deprecated since January 1, 2020. This also allowed us to optimize it for working with our DBs.
2. Mendelian randomization methods – a set of tests that allow to infer causal relationships between two traits (Hemani et al., 2016). Hemani and colleagues provided an open source R package, which includes such methods. To this, we added a module for reading summary statistics from the GWAS-MAP DB in the required format.
3. Summary-level Mendelian randomization (SMR) and heterogeneity in dependent instruments (HEIDI) are the tests to ascertain whether two different traits are associated with the same locus (SMR) and whether this association can be explained by the null hypothesis of pleiotropy or by an alternative hypothesis that each trait is associated with different SNPs in linkage disequilibrium (LD) (HEIDI) (Zhu et al., 2016). We implemented the SMR-HEIDI tests ourselves for the GWAS-MAP platform. The rationale behind this was mainly that the SMR tool developed by Zhu

Database content

Domain	Collection	Number of sets of summary statistics	Associations (billions)
Complex traits	UKB_NealeLab (Neale Lab, 2018), UKB_GeneAtlas (Canela-Xandri et al., 2018), CVD (Schunkert et al., 2011; Nikpay et al., 2015; Howson et al., 2017), and others	2475	25.5
Metabolomics (mQTL)	Metabolomics (Kettunen et al., 2016), GLGC (Willer et al., 2013)	127	1.3
Proteomics (pQTL)	SomaLogic_2017 (Suhre et al., 2017), SomaLogic_2018 (Sun et al., 2018), OLINK (Folkersen et al., 2017)	4489	33.5
Glycomics (glyQTL)	Plasma_Glycome (Sharapov et al., 2019), Glycomics_ IgG (Klarić et al., 2020)	190	1.1
Transcriptomics (eQTL)	GTEx_v7 (GTEx Consortium et al., 2017), blood_eQTL (Westra et al., 2013), CEDAR (Momozawa et al., 2018)	1137406	7.9
		Total:	70

Note. List of collections in the DB, the domains to which they have been assigned and the corresponding numbers of GWAS/RWAS and SNP summary statistics. Domains: complex traits, mQTL (metabolite levels), pQTL (protein levels), glycomics (glycan levels), and eQTL (gene expression data). UKB, UK Biobank; CVD, cardiovascular diseases; GLGC, Global Lipids Genetics Consortium; IgG, immunoglobulins G.

and colleagues (2016) specializes in testing pleiotropy between the level of gene expression (RWAS) and a complex trait (GWAS), but not between two sets of GWAS results summary statistics.

4. We also implemented the θ metric defined by Momozawa et al., which assesses the similarity between association profiles using only summary statistics and is an alternative to the HEIDI test. This method is preferable when the LD information of the population used in a GWAS is lacking or unreliable (Momozawa et al., 2018). The θ metric as implemented in GWAS-MAP is based on the equations provided in the article (Momozawa et al., 2018).
5. Finally, the GWAS-MAP platform implements several standard methods for meta-analysis which can be applied to a pool of GWAS results of the same trait in order to obtain enhance power (Winkler et al., 2014). GWAS-MAP has a module for checking the quality of the GWAS results to be used in meta-analyses and a module for meta-analysis. We have implemented two methods for meta-analysis: inverse-variance weighting and *Z*-score (Evangelou, Ioannidis, 2013).

GWAS-MAP|homo database content

To allow the researcher to filter GWAS/RWAS according to certain criteria, the platform offers key information including the publication data, reference set used for imputations, the name/type of the DNA microarray (e. g. Metabochip, Affymetrix, Illumina SNP arrays) or whole-genome sequencing used in each study. Currently, the DB contains more than 70 billion associations between SNPs and traits, collected from 7281 GWAS and more than a million RWAS (see the Table). To give a reader an idea of the context, such popular databases as "GWAS central" (Beck et al., 2020) provides information on 71 million of associations, while Phenoscanner (Staley et al., 2016) – 65 billion.

The GWAS and RWAS in the DBs are assigned to the following domains: complex traits (including diseases), metabolites (mQTL), proteins (pQTL), glycans and gene expression data. Additionally, the GWAS and RWAS results coming from the same study are pooled in a collection. The presence of GWAS traits from different domains enables the researcher to conduct a comprehensive study of the trait of interest, to identify ways of how the trait of interest is influenced by the expression levels of genes, proteins and metabolites, and to look for associations with other diseases or quantitative traits.

GWAS-MAP application: a genetic analysis of varicose veins

Varicose veins (VV) is a widely prevalent disease affecting on average every third adult in Russia (Zolotukhin et al., 2017). The genetic basis of this pathology has long been poorly studied. Shadrina and co-authors have performed the first large-scale study of its genetic architecture using a range of modern methods in bioinformatics as implemented in GWAS-MAP (Shadrina et al., 2019).

The study used UK Biobank (http://www.ukbiobank.ac.uk/) data on 408,455 individuals of European descent. GWAS summary statistics of VV were retrieved from the open access databases Gene ATLAS (Canela-Xandri et al., 2018) and the Neale Lab website (Neale Lab, 2018). Shadrina and co-authors identified 12 genetic loci associated with VV which account for 13.4 % of the SNP-based heritability. A gene or a group of genes most probably involved in VV pathogenesis was prioritized for each locus. The SMR-HEIDI implementation in GWAS-MAP was used as one of the prioritization methods. With SMR-HEIDI, we searched for the genes for which the expression levels are associated with SNPs affecting VV risk (cases of the so-called colocalization of associated loci). The analysis relied on data from the eQTL (expression quantitative trait loci) domain, namely data of 44 tissues in the GTEx_v7

(GTEx Consortium et al., 2017) and blood_eQTL (Westra et al., 2013) collections. Colocalization was demonstrated for the following loci: rs3101725 (associated with the expression level of the long non-coding RNA *LINC01184* in 9 tissues), rs2241173 (associated with the expression level of the non-coding RNA *AC005152.3* in the lower extremity skin) and rs2861819 (associated with expression levels of the *PPP3R1* gene in blood). Because the functions of *LINC01184* and *AC005152.3* are not yet known, we may only speculate about the role of these RNAs in VV. As far as *PPP3R1* is concerned, its association with VV appears to be more sound. Its product is involved in the inflammatory response in the vascular wall, stimulating the production of the chemokine MCP-1 (Satonaka et al., 2004), which is consistent with the modern view of the pathogenesis of chronic venous disease (Lim, Davies, 2009; del Rio Solá et al., 2009). Additionally, Smetanina and co-workers demonstrated enhanced *PPP3R1* expression in VV specimens compared to unaffected veins (Smetanina et al., 2018).

In addition to gene prioritization, SMR-HEIDI was used to search for traits associated with VV-related functional variants. The analysis involved 2219 traits, including various diseases, levels of metabolites and proteins in blood, and revealed 32 traits associated with 6 loci. The traits can conventionally be divided into three main groups: one associated with body weight and the total metabolic rate; a second with blood test results, and a last one with all others.

The GWAS-MAP platform was also used for the analysis of genetic correlations between VV and 861 traits, the summary statistics of which were obtained by analyzing more than 10 thousand individuals. The analysis showed the presence of common genetic variance between VV cases and 62 traits. Some of these traits were already known from previous epidemiological studies: overweight, standing and heavy physical work, deep venous thrombosis, gonarthrosis, and pain in the legs when walking. Other traits that, at the genetic level, correlate with VV, such as intellect, memory, educational attainment, or whole-body pain, have not previously been reported as associated with VV.

Finally, Shadrina and co-workers used Mendelian randomization for the analysis of causal relationships between various traits and VV. Analysis results showed that the following traits directly influence the risk for VV: height (irrespective of weight), body weight; waist and hip circumferences, and the blood levels of two proteins, MICB and CD209 (also known as DC-SIGN). Curiously, the risk for VV increased with an increase of both body fat and fat-free mass. Data on height as a risk factor for VV are consistent with the Edinburgh Vein Study results (Lee et al., 2003). MICB and CD209 participate in the innate and the adaptive immune response. Because presented work is the first to propose that these proteins have roles in VV pathogenesis, we think it reasonable to repeat the analysis with an independent dataset. If the Mendelian randomization results are confirmed, these proteins can be regarded as promising candidates for further *in vivo* and *in vitro* studies aimed at finding therapeutic targets.

GWAS-MAP benefits and future development

The GWAS-MAP platform offers a broad range of opportunities for comprehensive analysis of GWAS results. We expect that GWAS-MAP will be helpful both for bioinformatics studies and as a reference source for medical researchers. For example, given a trait of interest, it is possible to compute what other traits it is genetically associated with, i. e., is controlled by overlapping sets of genetic variants. More generally, not only correlations between traits can be calculated, but also all pairwise correlations between the traits in the DB. The results can be used to cluster the traits and/or to build a network connecting traits that have a shared genetic basis (see, for example, Fig. 4 from Shadrina et al., 2019). Furthermore, the Mendelian randomization methods implemented in the platform will help to elucidate which of these associations are causal. Thus, it is possible to build a directed graph for interactions between traits. By considering a particular vertex, for example, "disease", it is possible to infer what metabolites, glycans and/or proteins can be used as its biomarkers.

If a researcher's interest lies with a locus or loci associated with a certain GWAS of interest, it is also interesting to consider colocalization. With SMR-HEIDI and the θ metric, it is possible to understand with the expression of what genes the GWAS loci are associated. Additionally, by analysis of RWAS results for the genes of interest and GWAS results in the domains for metabolites, proteins and/or glycans, it is possible to infer what biological processes are associated with changes in the expression of these genes. A large-scale analysis of colocalization will help to build networks of associations between traits in the DB and genes. These networks will be helpful in developing medications. Not only will they show what genes can be targeted, but also what implications and side effects of manipulations with the gene may entail. However, it should be kept in mind that these analyses are done *in silico* and therefore, and experimental validation is absolutely required.

A large number of methods have already been implemented in the platform – however, there are certainly more to come. Our short-term plans include the addition of new analysis methods: Depict (Pers et al., 2015), CoJo (Deng, Pan, 2018), and SbayesR (Lloyd-Jones et al., 2019). We are planning to develop a web-interface to allow external users to access our DBs and perform analyses. Such a web-interface will guide the user through the search for information about the association between a SNP and traits and will be convenient for e. g. medical researchers. We continue adding new data to the database and we are working on making GWAS-MAP useful for human populations other than Europeans.

Conclusion

We have developed the GWAS-MAP platform for aggregating, storing, analyzing, visualizing and providing access to summary statistics from GWAS and RWAS. Using the platform we collected GWAS-MAP|homo DB which contains over 70 billion associations between SNPs and traits. The GWAS-MAP user interface offers a universal workspace for operating on public and private data, and allows for rapid implementation of new analysis methods in the platform. The user communicates with the platform through command-line utilities, allowing him to upload data to the platform and run analyses.

The analysis of the genetic basis of varicose veins demonstrates the power of the platform for generating new biological hypotheses such as, for example, ours postulating a causal

relationship between the levels of the proteins MICB and CD209 in blood and the risk for this disease.

GWAS-MAP is a powerful platform for the analysis of summary statistics from GWAS and RWAS, it is actively used in research work and can be useful to a broad range of scientists. The platform evolves continuously through constant acquisition of more functionalities, and the DBs are updated with the actual data from GWAS and RWAS.

References

Beck T., Shorter T., Brookes A.J. GWAS Central: a comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies. *Nucleic Acids Res*. 2020;8(48):D933-D940. DOI 10.1093/nar/gkz895.

Benner C., Spencer C.C.A., Havulinna A.S., Salomaa V., Ripatti S., Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics.* 2016; 32(10):1493-1501. DOI 10.1093/bioinformatics/btw018.

Bulik-Sullivan B.K., Loh P.-R., Finucane H.K., Ripke S., Yang J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N., Daly M.J., Price A.L., Neale B.M. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 2015;47(3):291-295. DOI 10.1038/ng.3211.

Bush W.S., Moore J.H. Genome-wide association studies. *PLoS Comput. Biol.* 2012;8(12):e1002822. DOI 10.1016/B978-0-12-809633-8.20232-X.

Canela-Xandri O., Rawlik K., Tenesa A. An atlas of genetic associations in UK Biobank. *Nat. Genet*. 2018;50(11):1593-1599. DOI 10.1038/s41588-018-0248-z.

Choi S.W., O'Reilly P.F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience.* 2019;8(7). DOI 10.1093/gigascience/giz082.

del Rio Solá L., Aceves M., Dueñas A.I., González-Fajardo J.A., Vaquero C., Crespo M.S., García-Rodríguez C. Varicose veins show enhanced chemokine expression. *Eur. J. Vasc. Endovasc. Surg.* 2009; 38(5):635-641. DOI 10.1016/j.ejvs.2009.07.021.

Demirkan A., van Duijn C.M., Ugocsai P., Isaacs A., Pramstaller P.P., Liebisch G., Wilson J.F., Johansson Å., Rudan I., Aulchenko Y.S., Kirichenko A.V., … Meitinger T., Hicks A.A., Hayward C., DIAGRAM Consortium, CARDIoGRAM Consortium, CHARGE Consortium & EUROSPAN Consortium. Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS Genet.* 2012;8(2):e1002490. DOI 10.1371/journal.pgen.1002490.

Deng Y., Pan W. Improved use of small reference panels for conditional and joint analysis with GWAS summary statistics. *Genetics.* 2018;209(2):401-408. DOI 10.1534/genetics.118.300813.

Elgaeva E.E., Tsepilov Y., Freidin M.B., Williams F.M.K., Aulchenko Y., Suri P. ISSLS Prize in Clinical Science 2020. Examining causal effects of body mass index on back pain: a Mendelian randomization study. *Eur. Spine J*. 2019;686-391. DOI 10.1007/s00586-019-06224-6.

Evangelou E., Ioannidis J.P.A. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* 2013;14(6): 379-389. DOI 10.1038/nrg3472.

Evans D.M., Visscher P.M., Wray N.R. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum. Mol. Genet.* 2009; 18(18):3525-3531. DOI 10.1093/hmg/ddp295.

Fabregat-Traver D., Sharapov S.Z., Hayward C., Rudan I., Campbell H., Aulchenko Y., Bientinesi P. High-performance mixed models based genome-wide association analysis with omicABEL software. *F1000Research.* 2014;3:200. DOI 10.12688/f1000research.4867.1.

Folkersen L., Fauman E., Sabater-Lleal M., Strawbridge R.J., Frånberg M., Sennblad B., Baldassarre D., Veglia F., Humphries S.E., Rauramaa R., de Faire U., Smit A.J., Giral P., Kurl S., Mannarino E., Enroth S., Johansson Å., Enroth S.B., Gustafsson S., Lind L., Lindgren C., Morris A.P., Giedraitis V., Silveira A., Franco-Cereceda A., Tremoli E., Gyllensten U., Ingelsson E., Brunak S., Eriksson P., Ziemek D., Hamsten A., Mälarstig A. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet.* 2017;13(4):e1006706. DOI 10.1371/journal.pgen.1006706.

Giambartolomei C., Vukcevic D., Schadt E.E., Franke L., Hingorani A.D., Wallace C., Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383. DOI 10.1371/journal.pgen.1004383.

GTEx Consortium et al. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204-213. DOI 10.1038/nature24277.

Hemani G., Zheng J., Wade K.H., Laurin C., Elsworth B., Burgess S., Bowden J., Langdon R., Tan V., Yarmolinsky J., Shihab H.A., Timpson N., Evans D.M., Relton C., Martin R.M., Smith G.D., Gaunt T.R., Haycock P.C. MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. *BioRxiv.* 2016;18092. DOI 10.1101/078972.

Howson J.M.M., Barnes D.R., Ho W.K., Young R., Paul D.S., Freitag D.F., Sun B.B., Lin W.Y., Surendran P., Di Angelantonio E., Chowdhury R., … Wang T.D., Rasheed A., Frossard P., Alam D.S., Majumder A.A.S. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat. Genet*. 2017; 49(7):1113-1119. DOI 10.1038/ng.3874.

International Schizophrenia Consortium, Purcell S.M., Wray N.R., Stone J.L., Visscher P.M., O'Donovan M.C., Sullivan P.F., Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009;460(7256):748-752. DOI 10.1038/nature08185.

Kettunen J., Demirkan A., Würtz P., Draisma H.H.M., Haller T., Rawal R., Vaarhorst A., Kangas A.J., Lyytikäinen L.-P., Pirinen M., Pool R., … Raitakari O., Salomaa V., Slagboom P.E., Waldenberger M., Ripatti S., Ala-Korpela M. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* 2016;7:11122. DOI 10.1038/ncomms11122.

Khera A.V., Chaffin M., Aragam K.G., Haas M.E., Roselli C., Choi S.H., Natarajan P., Lander E.S., Lubitz S.A., Ellinor P.T., Kathiresan S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* 2018;50(9):1219-1224. DOI 10.1038/s41588-018-0183-z.

Khera A.V., Chaffin M., Wade K.H., Zahid S., Brancale J., Xia R., Distefano M., Senol-Cosar O., Haas M.E., Bick A., Aragam K.G., Lander E.S., Smith G.D., Mason-Suares H., Fornage M., Lebo M., Timpson N.J., Kaplan L.M., Kathiresan S. Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell.* 2019; 177(3):587-596. DOI 10.1016/j.cell.2019.03.028.

Kichaev G., Yang W.-Y., Lindstrom S., Hormozdiari F., Eskin E., Price A.L., Kraft P., Pasaniuc B. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* 2014;10(10):e1004722. DOI 10.1371/journal.pgen.1004722.

Klarić L., Tsepilov Y.A., Stanton C.M., Mangino M., Sikka T.T., Esko T., Pakhomov E., Salo P., Deelen J., McGurnaghan S.J., Keser T., … Zoldoš V., Vitart V., Spector T., Aulchenko Y.S., Lauc G., Hayward C. Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. *Sci. Adv.* 2020;6(8):eaax0301. DOI 10.1126/sciadv.aax0301.

Klein R.J. Complement factor H polymorphism in age-related macular degeneration. *Science.* 2005;308(5720):385-389. DOI 10.1126/science.1109557.

Lee A.J., Evans C.J., Allan P.L., Ruckley C.V., Fowkes F.G.R. Lifestyle factors and the risk of varicose veins: Edinburgh Vein Study. *J. Clin. Epidemiol.* 2003;56(2):171-179. DOI 10.1016/s0895-4356(02)00518-8.

Lim C.S., Davies A.H. Pathogenesis of primary varicose veins. *Br. J. Surg.* 2009;96(11):1231-1242. DOI 10.1002/bjs.6798.

Lloyd-Jones L.R., Zeng J., Sidorenko J., Yengo L., Moser G., Kemper K.E., Wang H., Zheng Z., Magi R., Esko T., Metspalu A., Wray N.R., Goddard M.E., Yang J., Visscher P.M. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* 2019;10(1):5086. DOI 10.1038/s41467-019-12653-0.

Mak T.S.H., Porsch R.M., Choi S.W., Zhou X., Sham P.C. Polygenic scores via penalized regression on summary statistics. *Genet. Epidemiol.* 2017;41(6):469-480. DOI 10.1002/gepi.22050.

Mavaddat N., Michailidou K., Dennis J., Lush M., Fachal L., Lee A., Tyrer J.P., Chen T.H., Wang Q., Bolla M.K., Yang X., … Antoniou A.C., Chatterjee N., Kraft P., García-Closas M., Simard J., Easton D.F. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am. J. Hum. Genet*. 2019;104(1):21-34. DOI 10.1016/j.ajhg.2018.11.002.

Momozawa Y., Dmitrieva J., Théâtre E., Deffontaine V., Rahmouni S., Charloteaux B., Crins F., Docampo E., Elansary M., Gori A.S., Mariman R., … Tremelling M., Wei Z., Winkelmann J., Zhang C.K., Zhao H., Zhang H. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.* 2018;9(1):2427. DOI 10.1038/s41467-018-04365-8.

Neale Lab. 2018. GWAS database available at http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-bioban.

Nikpay M., Goel A., Won H.-H., Hall L.M., Willenborg C., Kanoni S., Saleheen D., Kyriakou T., Nelson C.P., Hopewell J.C., Webb T.R., … McPherson R., Deloukas P., Schunkert H., Samani N.J., Farrall M., CARDIoGRAMplusC4D Consortium. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet*. 2015;47(10):1121-1130. DOI 10.1038/ng.3396.

O'Connor L.J., Price A.L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet.* 2018;50(12):1728-1734. DOI 10.1038/s41588-018-0255-0.

Pers T.H., Karjalainen J.M., Chan Y., Westra H.-J., Wood A.R., Yang J., Lui J.C., Vedantam S., Gustafsson S., Esko T., Frayling T., Speliotes E.K., GIANT Consortium, Boehnke M., Raychaudhuri S., Fehrmann R.S.N., Hirschhorn J.N., Franke L. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* 2015;6:5890. DOI 10.1038/ncomms6890.

Satonaka H., Suzuki E., Nishimatsu H., Oba S., Takeda R., Goto A., Omata M., Fujita T., Nagai R., Hirata Y. Calcineurin promotes the expression of monocyte chemoattractant protein-1 in vascular myocytes and mediates vascular inflammation. *Circ. Res.* 2004;94(5):693-700. DOI 10.1161/01.RES.0000118250.67032.5E.

Schaid D.J., Chen W., Larson N.B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* 2018;19(8):491-504. DOI 10.1038/s41576-018-0016-z.

Schunkert H., König I.R., Kathiresan S., Reilly M.P., Assimes T.L., Holm H., Preuss M., Stewart A.F.R., Barbalic M., Gieger C., Absher D., … Roberts R., Thorsteinsdottir U., O'Donnell C.J., McPherson R., Erdmann J., Samani N.J. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* 2011;43(4):333-338. DOI 10.1038/ng.784.

Shadrina A.S., Sharapov S.Z., Shashkova T.I., Tsepilov Y.A. Varicose veins of lower extremities: insights from the first large-scale genetic study. *PLoS Genet.* 2019;15(4):e1008110. DOI 10.1371/journal.pgen.1008110.

Shadrina A.S., Shashkova T.I., Torgasheva A.A., Sharapov S.Z., Klarić L., Pakhomov E.D., Alexeev D.G., Wilson J.F., Tsepilov Y.A., Joshi P.K., Aulchenko Y.S. Prioritization of causal genes for coronary artery disease based on cumulative evidence from experimental and *in silico* studies. *Sci. Rep.* 2020;10(1):1-15. DOI 10.1038/s41598-020-67001-w.

Sharapov S.Z., Tsepilov Y.A., Aulchenko Y.S., Shadrina A.S., Klaric L., Vilaj M., Vuckovic F., Stambuk J., Trbojevic-Akmacic I., Kristic J., Simunovic J., Momcilovic A., Pucic-Bakovic M., Lauc G., Mangino M., Spector T., Williams F.M.K., Thareja G., Suhre K., Simurina M., Pavic T., Dagostino C., Dmitrieva J., Georges M., Campbell H., Dunlop M.G., Farrington S.M., Doherty M., Gieger C., Allegri M., Louis E. Defining the genetic control of human blood plasma N-glycome using genome-wide association study. *Hum. Mol. Genet.* 2019;28(12):2062-2077. DOI 10.1093/hmg/ddz054.

Shen X., Klarić L., Sharapov S., Mangino M., Ning Z., Wu D., Trbojević-Akmačić I., Pučić-Baković M., Rudan I., Polašek O., Hayward C., Spector T.D., Wilson J.F., Lauc G., Aulchenko Y.S. Multivariate discovery and replication of five novel loci associated with immunoglobulin G N-glycosylation. *Nat. Commun.* 2017;8(1):447. DOI 10.1038/s41467-017-00453-3.

Smetanina M.A., Kel A.E., Sevost'ianova K.S., Maiborodin I.V., Shevela A.I., Zolotukhin I.A., Stegmaier P., Filipenko M.L. DNA methylation and gene expression profiling reveal MFAP5 as a regulatory driver of extracellular matrix remodeling in varicose vein disease. *Epigenomics.* 2018;10(8):1103-1119. DOI 10.2217/epi-2018-0001.

Speed D., Balding D.J. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet.* 2019;51(2):277-284. DOI 10.1038/s41588-018-0279-5.

Staley J.R., Blackshaw J., Kamat M.A., Ellis S., Surendran P., Sun B.B., Paul D.S., Freitag D., Burgess S., Danesh J., Young R., Butterworth A.S. PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics.* 2016;20(15):3207-3209. DOI 10.1093/bioinformatics/btw373.

Suhre K., Arnold M., Bhagwat A.M., Cotton R.J., Engelke R., Raffler J., Sarwath H., Thareja G., Wahl A., DeLisle R.K., Gold L., Pezer M., Lauc G., El-Din Selim M.A., Mook-Kanamori D.O., Al-Dous E.K., Mohamoud Y.A., Malek J., Strauch K., Grallert H., Peters A., Kastenmüller G., Gieger C., Graumann J. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* 2017;8:14357. DOI 10.1038/ncomms14357.

Sun B.B., Maranville J.C., Peters J.E., Stacey D., Staley J.R., Blackshaw J., Burgess S., Jiang T., Paige E., Surendran P., Oliver-Williams C., Kamat M.A., Prins B.P., Wilcox S.K., Zimmerman E.S., Chi A., Bansal N., Spain S.L., Wood A.M., Morrell N.W., Bradley J.R., Janjic N., Roberts D.J., Ouwehand W.H., Todd J.A., Soranzo N., Suhre K., Paul D.S., Fox C.S., Plenge R.M., Danesh J., Runz H., Butterworth A.S. Genomic atlas of the human plasma proteome. *Nature.* 2018;558(7708):73-79. DOI 10.1038/s41586-018-0175-2.

The 1000 Genomes Project Consortium, Auton A., Brooks L.D., Durbin R.M., Garrison E.P., Kang H.M., Korbel J.O., Marchini J.L., McCarthy S., McVean G.A., Abecasis G.R. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74. DOI 10.1038/nature15393.

Timmers P.R., Mounier N., Lall K., Fischer K., Ning Z., Feng X., Bretherick A.D., Clark D.W., eQTLGen Consortium, Agbessi M., Ahsan H., Alves I., Andiappan A., Awadalla P., Battle A., Bonder M.J., Boomsma D., Christiansen M., Claringbould A., … Shen X., Esko T., Kutalik Z., Wilson J.F., Joshi P.K. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *eLife.* 2019;8:e39856. DOI 10.7554/eLife.39856.

Vilhjálmsson B.J., Yang J., Finucane H.K., Gusev A., Lindström S., Ripke S., Genovese G., Loh P.-R., Bhatia G., Do R., Hayeck T., Won H.-H., Schizophrenia Working Group of the Psychiatric Genomics Consortium, DRIVE study, Kathiresan S., Pato M., Pato C., Tamimi R., Stahl E., Zaitlen N., Pasaniuc B., Belbin G., Kenny E.E., Schierup M.H., De Jager P., Patsopoulos N.A., McCarroll S., Daly M., Purcell S., Chasman D., Neale B., Goddard M., Visscher P.M., Kraft P., Patterson N., Price A.L. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* 2015;97(4):576-592. DOI 10.1016/j.ajhg.2015.09.001.

Visscher P.M., Wray N.R., Zhang Q., Sklar P., McCarthy M.I., Brown M.A., Yang J. 10 years of GWAS discovery: biology, func-

tion, and translation. *Am. J. Hum. Genet.* 2017;101(1):5-22. DOI 10.1016/j.ajhg.2017.06.005.

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

Willer C.J., Schmidt E.M., Sengupta S., Peloso G.M., Gustafsson S., Kanoni S., Ganna A., Chen J., Buchkovich M.L., Mora S., Beckmann J.S., … Ripatti S., Cupples L.A., Sandhu M.S., Rich S.S., Boehnke M., Deloukas P., Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet*. 2013;45(11):1274-1283. DOI 10.1038/ng.2797.

Winkler T.W., Day F.R., Croteau-Chonka D.C., Wood A.R., Locke A.E., Mägi R., Ferreira T., Fall T., Graff M., Justice A.E., Luan J.A., Gustafsson S., Randall J.C., Vedantam S., Workalemahu T., Kilpeläinen T.O., Scherag A., Esko T., Kutalik Z., Heid I.M., Alavere H., Fischere K., Metspalu A., Mihailov E., Milani L., Morris A.P., Nelis M., Perola M., Tammesoo M.-L., Teder-Laving M., Loos R.J.F., GIANT Consortium. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* 2014;9(5):1192-1212. DOI 10.1038/nprot.2014.071.

Wu X., Zhu X., Wu G.Q., Ding W. Data mining with big data. *IEEE Trans. Knowl. Data Eng*. 2013;26(1):97-107. DOI 10.1109/TKDE.2013.109

Yang J., Lee S.H., Goddard M.E., Visscher P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 2011;88(1):76-82. DOI 10.1016/j.ajhg.2010.11.011.

Zhu Z., Zhang F., Hu H., Bakshi A., Robinson M.R., Powell J.E., Montgomery G.W., Goddard M.E., Wray N.R., Visscher P.M., Yang J. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 2016;48(5):481-487. DOI 10.1038/ng.3538.

Zolotukhin I.A., Seliverstov E.I., Shevtsov Y.N., Avakiants I.P., Nikishkov A.S., Tatarintsev A.M., Kirienko A.I. Prevalence and risk factors for chronic venous disease in the general Russian population. *Eur. J. Vasc. Endovasc. Surg.* 2017;54(6):752-758. DOI 10.1016/j.ejvs.2017.08.033.

Acknowledgements. We are grateful to Lucija Klaric and Peter Joshi for discussion and help with testing the platform. We thank Tatiana Iosifovna Aksenovich for discussion of the manuscript. We thank Viktoria Voroshilova, Alexander Severinov and Ksenia Khudyakova for help with platform development and testing.

Funding. The development of a database of genetic associations of human traits was supported by the Russian Ministry of Science and Education under Excellence Programme 5-100. The development of the program modules of GWAS-MAP was supported by PolyKnomics BV.

Conflict of interest. YSA and LCK are co-owners of Maatschap PolyOmica and PolyKnomics BV, private organizations, providing services, research and development in the field of computational, statistical, and quantitative (gen)omics.

Received June 23, 2020. Revised September 30, 2020. Accepted October 1, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):885-896
DOI 10.18699/VJ20.687

The role of microRNAs in learning and long-term memory

L.N. Grinkevich

Pavlov Institute of Physiology of the Russian Academy of Sciences, St. Petersburg, Russia
e-mail: Larisa_Gr_spb@mail.ru

Abstract. The mechanisms of long-term memory formation and ways to improve it (in the case of its impairment) remain an extremely difficult problem yet to be solved. Over the recent years, much attention has been paid to microRNAs in this regard. MicroRNAs are unique endogenous non-coding RNAs about 22 nucleotides in length; each can regulate translation of hundreds of messenger RNA targets, thereby controlling entire gene networks. MicroRNAs are widely represented in the central nervous system. A large number of studies are currently being conducted to investigate the role of microRNAs in the brain functioning. A number of microRNAs have been shown to be involved in the process of synaptic plasticity, as well as in the long-term memory formation. Disruption of microRNA biogenesis leads to significant cognitive dysfunctions. Moreover, impaired microRNA biogenesis is one of the causes of the pathogenesis of mental disorders, neurodegenerative illnesses and senile dementia, which are often accompanied by deterioration in the learning ability and by memory impairment. Optimistic predictions are made that microRNAs can be used as targets for therapeutic treatment and for diagnosing the above pathologies. The importance of applications related to microRNAs significantly raises interest in studying their functions in the brain. Thus, this review is focused on the role of microRNAs in cognitive processes. It describes microRNA biogenesis and the role of miRNAs in the regulation of gene expression, as well as the latest achievements in studying the functional role of microRNAs in learning and in long-term memory formation, depending on the activation or inhibition of their expression. The review presents summarized data on the effect of impaired microRNA biogenesis on long-term memory formation, including those associated with sleep deprivation. In addition, analysis is provided of the current literature related to the prospects of improving cognitive processes by influencing microRNA biogenesis via the use of CRISPR/Cas9 technologies and active mental and physical exercises.
Key words: epigenetics; miRNA; learning; long-term memory; cognitive impairment; sleep deprivation; environmental enrichment.

For citation: Grinkevich L.N. The role of microRNAs in learning and long-term memory. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):885-896. DOI 10.18699/VJ20.687

Роль микроРНК в обучении и долговременной памяти

Л.Н. Гринкевич

Институт физиологии им. И.П. Павлова Российской академии наук, Санкт-Петербург, Россия
e-mail: Larisa_Gr_spb@mail.ru

Аннотация. Механизмы формирования долговременной памяти и способы ее улучшения (в случае нарушения) остаются сложнейшей нерешенной проблемой. В последние годы большое внимание в этой связи уделяется микроРНК. МикроРНК являются уникальными эндогенными некодирующими РНК длиной около 22 нуклеотидов, каждая из которых может регулировать трансляцию сотен матричных РНК, тем самым управляя целыми сетями генов. МикроРНК широко представлены в центральной нервной системе. В настоящее время значительное количество исследований посвящено изучению роли микроРНК в функционировании мозга. Показано, что целый ряд микроРНК вовлечен в процесс синаптической пластичности, а также в формирование долговременной памяти. При этом нарушение биогенеза микроРНК приводит к значительным когнитивным дисфункциям. Более того, нарушение биогенеза микроРНК является одной из причин патогенеза заболеваний, связанных с психическими расстройствами, нейродегенеративными патологиями и старческой деменцией, которые часто сопровождаются ухудшением способности к обучению и нарушением памяти. Высказываются оптимистичные прогнозы, что микроРНК могут быть использованы в качестве мишеней для терапевтического лечения и диагностики данных патологий. Важное прикладное значение микроРНК увеличивает интерес к изучению их функций в работе мозга. Представленный обзор посвящен роли микроРНК в когнитивных процессах. Описаны биогенез микроРНК и роль микроРНК в регуляции экспрессии генов. Рассмотрены последние достижения в изучении функциональной роли микроРНК в обучении и формировании долговременной памяти, в зависимости от активации или ингибирования их экспрессии, и о влиянии нарушения биогенеза микроРНК на формирование долговременной

памяти. Небольшой раздел посвящен влиянию депривации сна на когнитивные процессы, зависимые от микроРНК. Кроме того, приведен анализ текущей литературы, связанной с перспективами улучшения когнитивных функций посредством влияния на биогенез микроРНК путем применения CRISPR/Cas9 технологий и активных умственных и физических нагрузок.
Ключевые слова: эпигенетика; микроРНК; долговременная память; когнитивные нарушения; депривация сна, обогащенная среда.

Introduction

The mechanisms of forming long-term memory (LTM) and of its improvement in case of impairment resulting from trauma, neurological and neurodegenerative diseases, and age-related dysfunctions are among the most challenging issues to be solved by the science. Back in the middle of the last century, researchers realized that LTM formation requires active involvement of the genome. Later, it was shown that newly synthesized proteins are necessary for modification of synaptic contacts and rearrangements of the neural networks involved in consolidating new experiences (Sweatt, 2016). The main difficulties in studying the molecular basis of LTM are associated with both complexity of the structure of the central nervous system and with variety of the regulatory processes acting at the genome level. The latter include regulation of gene expression by DNA-binding transcription factors, as well as epigenetic modifications that regulate the structure of chromatin (Berger, 2007). These epigenetic processes are widely involved in brain functioning, including neuronal differentiation and adaptive behavior, inter alia LTM formation (Fischer, 2014; Kim, Kaang, 2017).

Somewhat later, studies were started on involvement of microRNA in the epigenetic regulation of gene expression. MicroRNAs are unique non-coding molecules, each of which able to regulate translation of hundreds to thousands of messenger RNAs (mRNAs) targets. MicroRNAs are most widely represented in the central nervous system, and many of them are expressed at a high level (Chen, Qin, 2015). A large number of studies have been published on microRNAs participation in neuronal differentiation (Baek et al., 2014; Chen, Qin, 2015), in the LTM formation (Rajasethupathy et al., 2009; Gao et al., 2010; Hu Z., Li, 2017), as well as in the pathogenesis of diseases associated with mental disorders, neurodegenerative pathologies and senile dementia (Beveridge et al., 2010; Danka Mohammed et al., 2017; Wingo et al., 2020). Optimistic predictions are expressed that a number of microRNAs can be used as targets for therapeutic treatment and diagnosis of diseases accompanied by cognitive impairment (Liu et al., 2017). This important applied value of microRNAs raises interest in the study of their functions. This review will examine in detail the role of microRNAs in cognitive processes.

MicroRNAs – biogenesis and mechanism of regulating gene expression

For the first time, microRNAs, as functionally significant molecules capable of regulating gene expression, were described in the nematode *C. elegans* in 1993 (Lee R. et al., 1993). In 2000, highly conserved microRNA let-7, which is necessary for organism development, was discovered in the same animal (Reinhart et al., 2000). In parallel, in 1998, A. Fire and C. Mello published an article in which it was shown that by means of small double-stranded RNAs (siRNA) it is possible to silence genes – this mechanism was called RNA interference. Moreover, already in 2006, in view of the importance of this discovery, Andrew Fire and Craig Mello were awarded the Nobel Prize in Physiology or Medicine. It has been shown that gene silencing by means of RNA interference is also carried out by microRNA (He, Hannon, 2004). Since the 2000s, an avalanche of studies on the functional role of microRNA and its biogenesis in many animal species started.

MicroRNAs are a family of small, highly conserved endogenous non-coding RNAs about 22 nucleotides long (He, Hannon, 2004; Bitetti et al., 2018). MicroRNA biogenesis is a complex and multistep process, including transcription from DNA of a rather long primary transcript (pri-miRNA) with characteristic stem-loop structures, its processing to form pre-miRNA, translocation of pre-miRNA into the cytoplasm and its further processing to form microRNA (miRNA). Next, the microRNA interacts with the RISC complex (RNA-induced silencing complex) in which the microRNA binds to the target mRNA and induces degradation and/or mRNA translational repression (Bartel, 2009; Aksoy-Aksel et al., 2014). Evolutionarily conserved proteins control all of the above stages: DROSHA, DGCR8, EXP5, RAN, DICER, TARBP2, AGO, and PIWI. Proteins DROSHA and DGCR8 are endonucleases and control the processing of the primary microRNA transcript. EXP5 and RAN are involved in translocation of pre-miRNAs from the nucleus to the cytoplasm. DICER endonuclease regulates cleavage of pre-miRNAs to form mature miRNAs and association of mature miRNAs with the RISC. The current data on microRNA biogenesis are described in detail in the review by Smith and Kenny (Smith, Kenny, 2018).

In the cytoplasm of neuronal cell bodies, miRNAs are often associated with processing structures that are responsible for storage and degradation of mRNA, and can also be found in stress granules that are formed in response to stress (Leung, 2015; Smith, Kenny, 2018). Expression of miRNAs in neurons is induced by electrical activity and downstream regulatory cascades at several levels, including the synthesis and processing of the primary transcript, processing of pre-miRNAs, and assembly of the RISC (Aksoy-Aksel et al., 2014). However, these processes are still poorly understood. MicroRNAs are highly stable molecules (up to 10 times more stable than mRNAs) (Gantier et al., 2011).

The complexity of studying the role of miRNAs in the brain is determined by the variety of neuronal and glial cells that perform different functions and express different patterns of miRNAs (McNeill, Van Vactor, 2012; Malmevik et al., 2016). Moreover, each miRNA can have hundreds of different mRNAs as targets, and expression of a particular mRNA can be regulated by several miRNAs (Lewis et al., 2003; John et al., 2004). Therefore, dysregulation of a single miRNA can have a large polygenic effect. The number of identified miRNAs already amounts to several thousands, and according to various estimates, miRNAs are capable of regulating the expression of 30 to 70 % of all genes encoding proteins at the posttranscriptional level (Selbach et al., 2008). It is important that about 70 % of miRNAs are expressed in the brain, and quite differentially in different regions (cited after Chen, Qin, 2015). In addition, microRNAs can be secreted into the extracellular space, including the circulatory system, to ensure intercellular and interorgan communication (Lesseur et al., 2014; Smith, Kenny, 2018). The pattern of extracellular miRNAs changes in a number of pathologies, and these data are beginning to be used for diagnostic purposes (Lesseur et al., 2014; Smith, Kenny, 2018). Thus, miRNAs are the key regulators of many gene networks and, accordingly, can coordinate the most important processes in the organism.

MicroRNAs in learning and long-term memory

Currently, several learning models are used to study the molecular mechanisms of LTM, which can be divided into associative and non-associative ones. The former includes development of various conditioned reflexes, and the latter includes non-associative analogs of learning, such as sensitization (facilitation), depression (habituation), post-tetanic potentiation and post-tetanic depression.

Sensitization is "enhancement of a pre-existing response of an animal to a stimulus as a result of application of another, nociceptive (painful) stimulus". Sensitization is necessary for an animal to respond to a stimulus that was previously insignificant for it (Kandel, 1982). Depression (habituation) is weakening of the response to a previously significant stimulus as a result of its periodic reiteration. Through habituation, an animal learns to ignore stimuli that have lost their novelty or meaning. At the cellular level, sensitization is associated with an increase in the efficiency of synaptic communication between neurons, and depression – with its weakening (Kandel, 1982).

Post-tetanic potentiation is improvement in synapse conduction after a series of frequent (tetanizing) stimulations of incoming fibers. Post-tetanic depression is deterioration in synaptic conduction after a series of weak rhythmic stimuli.

Long-term changes in the efficiency of synaptic transmission in these models require involvement of the genome and are caused by long-term plastic changes in the synapses through remodeling of presynaptic and/or postsynaptic structures, the spiny morphogenesis, and/or growth or elimination of synapses. Thus, synaptic plasticity is the mechanism by which the brain encodes and stores information. It is believed that these processes underlie LTM formation in both animals and humans.

The role of microRNAs in synaptic plasticity and local biosynthesis in neurites

The role of microRNAs in synaptic plasticity has been most fully investigated using models of long-term post-tetanic depression (LTD) and post-tetanic potentiation (LTP). Local protein synthesis plays an important role in these processes. It has been shown that the synthesis of receptors, proteins involved in the transport of synaptic vesicles and proteins needed for modeling the growth of spines can occur in neurites, since neurites contain the necessary set for translation of mRNA, including ribosomes, and, moreover, pre-microRNA, microRNA and Dicer enzyme, which allows microRNA to regulate local biosynthesis (Lugli et al., 2005; Bicker et al., 2013; Smalheiser, 2014; Hu Z., Li, 2017). Thus, during LTD formation in the dendrites of hippocampal neurons, local synthesis of the glutamate receptor GluA1 occurs, its expression is regulated by miR-501-3p, and this process is required for remodeling of dendritic spines, the density of which determines the efficiency of the synapse (Hu Z. et al., 2015).

Long-term remodeling of the dendritic tree also involves miR-191, miR-135, and miR-137 (Hu Z. et al., 2014; Siegert et al., 2015). On the other hand, miR-26a and miR-384-5p participate in the formation of LTP, the expression of which decreases during tetanization in an RSK3-dependent manner. Structural and signaling proteins of synapses (Gu Q. et al., 2015) are the targets of these microRNAs. At the same time, LTP is accompanied by expansion of spines and formation of new ones, while the opposite picture is observed for LTD (Hu Z., Li, 2017). It has been suggested that modulation of protein synthesis in synapses may be based on local Ca^{2+}-dependent activation of the Dicer enzyme (Lugli et al., 2005).

The effect of microRNA biogenesis disorders on LTM formation

In the initial studies of involvement of miRNAs in learning and memory, animals with genetic impairments of enzymes involved in miRNA biogenesis, in particular, with dysfunction of the Dicer, were used (Konopka et al., 2010; Fiorenza, Barco, 2016; Fiorenza et al., 2016). Induced by Tamoxifen injection, deletion of Dicer in the forebrain of mice (mutation $Dicer1^{CaMKCreERT2}$) has been shown to cause the loss of a number of brain-enriched microRNAs, including miR-124, miR-132, miR-137, miR-138, miR-29a/c, and these mice show improved memory (Konopka et al., 2010). In animals with Dicer suppression, the excitability of pyramidal neurons in the CA1 region of the hippocampus also increased, as well as induction of "early genes" required for LTM formation (Fiorenza et al., 2016). The above data are supported by the studies of Hansen et al. (Hansen et al., 2010) and Siegert et al. (Siegert et al., 2015), which showed that overexpression of miR-132 in forebrain neurons in adult mice (transgenic

mice), and miR-137 in the dentate gyrus (using the lentiviral technology) leads to LTM impairment.

Contradictory to the above are the studies carried out on aging animals in which the content of many microRNAs decreases with age; however, cognitive impairment occurs (Inukai et al., 2012; Chmielarz et al., 2017). Dicer dysfunction in the cerebellum leads to progressive loss of microRNA and death of Purkinje cells, and in the forebrain it causes abnormal hyper phosphorylation of the tau protein and neurodegeneration similar to that in Alzheimer's disease, which accordingly impairs cognitive processes (Hébert et al., 2010; Dimmeler, Nicotera, 2013). In addition, impairment of microRNA biosynthesis in dopaminergic neurons due to suppression or depletion of Dicer (tissue-specific inducible suppression) causes dopaminergic cell dysfunction, while pharmacological stimulation is neuroprotective (Chmielarz et al., 2017). In addition, the pharmacological inhibition of Dicer activity with poly-lysine (Poly-L-lysine hydrobromide) disrupts formation of a conditioned reflex with single-trial induced LTM in the mollusk *Lymnaea stagnalis* (Korneev et al., 2018) and impairs formation of the conditioned defense reflex in the mollusk *Helix* (Grinkevich, 2019). Thus, the last two studies show that short-term Dicer dysfunction can lead not to improvement, but to impairment of LTM.

As molecular genetic studies continued, it became clear that miRNAs are capable of not only inhibiting LTM, but also improving its formation. Thus, in the lateral amygdala 7 microRNAs upregulated and 32 downregulated by auditory fear training (Griggs et al., 2013). MicroRNAs miR-9 and miR-34 do not suppress, but support the capacity of spatial learning (the Morris water maze) and reference memory, respectively (Malmevik et al., 2016). The expression of a number of microRNAs in the hippocampus is activated during contextual fear formation (Vetere et al., 2014; Jovasevic et al., 2015). Thus, it became clear that the effect of microRNA on cognitive processes can be multidirectional, which was shown in numerous further studies.

MicroRNAs that negatively regulate LTM formation

The best-studied microRNAs whose expression decreases during learning include miR-124, miR-134, and miR-206.

miR-124 is one of the first studied microRNAs associated with LTM formation. It is a highly conserved microRNA with a high level of expression in the central nervous system. In 2009, a comprehensive work was published, which for the first time demonstrated involvement of miR-124 in LTM formation and studied its function (Rajasethupathy et al., 2009). As a learning model, the authors used long-term facilitation of synaptic connection between sensory and motor neurons of the mollusk *Aplysia*. As noted above, facilitation is an essential component in formation of a number of conditioned reflexes, including defensive ones, and is successfully used in the studies of the mechanisms of LTM formation (Kandel, 2012). It was shown that during development of facilitation, the level of miR-124 decreases, and, accordingly, translation of the target of miR-124, the transcription factor CREB-1, is activated (Rajasethupathy et al., 2009). As a result, CREB-1-dependent induction of the genes involved in synaptic modifications takes place, leading to a long-term increase in efficiency of synaptic transmission. At the same time, regulation of miR-124 expression is effected by the modulatory mediator serotonin, which mediates the action of the sensitizing pain stimulus, through PKA-MAPK/ERK-dependent signaling cascades (Rajasethupathy et al., 2009).

Further studies showed that miR-124 also plays an important role in LTM formation in vertebrates, in which, similarly to *Aplysia*, miR-124 is inhibited during learning (Yang et al., 2012; Malmevik et al., 2016). For example, the amount of miR-124 decreases in the hippocampus during spatial learning and social interactions in mice (Yang et al., 2012). In this case, the target of miR-124 is the transcription factor Zif268, which takes an active part in cognitive processes; accordingly, a decrease in the miR-124 amount induces translation of Zif268 (Yang et al., 2012). Increased expression of miR-124 (vector rAAV1/2-miR-124), or knockdown of Zif268 (LNA-Zif268 antisense) have a negative effect on LTM, and knockdown of miR-124 (LNA-miR-124 antisense) restores expression Zif268 and reverses of LTM formation.

The expression of miR-124 is regulated through cAMP and its intracellular receptors EPAC1 and EPAC2. Moreover, in EPAC$^{-/-}$ mice, impairment of spatial learning and memory, as well as social interactions, suppression of synaptic transmission, and impairment of long-term post-tetanic potentiation in the hippocampus are observed (Yang et al., 2012). It has been shown that inhibition of miR-124 in hippocampal neurons leads to improvement in LTM, potentially through an increase in the level of expression of genes associated with synaptic plasticity and neuronal transmission (Malmevik et al., 2016). At the same time, genes associated with translation and neurodegenerative diseases are suppressed.

A decrease in the miR-124 expression level associated with improvement in LTM formation in mice was noted in the work of Konopka et al. (Konopka et al., 2010). miR-124 is also involved in memory consolidation during sleep (Karabulut et al., 2019). Post-learning sleep deprivation during specific time windows induces expression of miR-124 in the hippocampus, inhibits synthesis of the neurotrophic factor BDNF, which is the target of miR-124, and, accordingly, disrupts LTM consolidation.

miR-134 is highly expressed in the brain and is detected not only in the bodies of neurons, but also in dendrites (Bicker et al., 2013). As in the case of miR-124, overexpression of miR-134 (the lentiviral technology) in the CA1 region of the hippocampus leads to significant deterioration in LTM formation in the contextual fear-conditioning paradigm and to abrogated long-term potentiation in this structure (Gao et al., 2010). miR-134, like miR-124, affects synaptic plasticity through post-transcriptional regulation of CREB-1 and BDNF in a CREB-dependent way. In turn, miR-134

expression is regulated by SIRT1 deacetylase. In mutant mice lacking the catalytic activity of SIRT1 in the brain, an increase in levels miR-134 is observed, followed by repression of target genes, and, accordingly, impairment of LTM (Gao et al., 2010). The increase in miR-134 levels is sufficient to mimic the behavioral and electrophysiological phenotypes of SIRT1-deficient mice. Conversely, inhibition of miR-134 reverses memory in SIRT1 knockdown mice and restores long-term potentiation in the CA1 region of the hippocampus (Gao et al., 2010).

Memory impairment under stress is also associated with suppression of the SIRT1/miR-134 pathway and the down-regulation expression of BDNF and synaptic proteins in the hippocampus (Shen et al., 2019). Thus, miR-134 and miR-124 can have a synergistic effect on the expression of genes involved in plastic rearrangements. In addition, disruption of miR-134 and miR-124-dependent regulation is an important mechanism underlying cognitive dysfunction in Alzheimer's disease (Wang X. et al., 2018; Baby et al., 2020).

miR-206. Increased levels of miR-206 are observed in the brain of Tg2576 mice (a model of Alzheimer's disease) and in the temporal cortex of the human brain in Alzheimer's disease (Lee S. et al., 2012). Decreased miR-206 levels lead to improved memory through induction of the neurotrophic factor BDNF. Improving memory through decrease in the content of miR-206-3p in the hippocampus and cortex is also facilitated by administration of donepezil, a drug with an antidementional effect (Wang C. et al., 2017).

Summarizing the above data, down-regulation expression of miR-124, miR-134, and miR-206 is necessary for successful formation of LTM, since these microRNAs normally block the expression of genes the products of which are necessary for plastic rearrangements.

MicroRNAs positively regulating learning and LTM formation

miR-9-3p positively influences hippocampus-dependent memory. Inhibition of miR-9-3p in the hippocampus leads to impairment of long-term post-tetanic potentiation (LTP) and disruption of LTM through increased expression of Dmd (dystrophin) and SAP97 (synapse-associated protein 97) genes, which are negatively correlated with LTP (Sim et al., 2016). At the same time, miR-9-5p, which is formed from a common precursor, is not involved in these processes. The miR-9 family is also involved in the regulation of synaptogenesis during early brain development, and a link was found between these developmental events and cognitive functions later in the adult life (Lin et al., 2017).

miR-92. The level of miR-92 increases in the hippocampus during the contextual fear memory formation in mice, which reduces expression of several miR-92 targets, including proteins KCC2, CPEB3 and MEF2D, which negatively regulates memory-induced structural plasticity (Vetere et al., 2014). Selective inhibition of miR-92 in CA1 neurons of the hippocampus (lentiviral technology) leads to up-regulation of KCC2, CPEB3 and MEF2D, prevents the learning-induced increase in the spine density and impairs this type of memory.

miR-195. Overexpression of miR-195 in the rat hippocampus using lenti-pre-miR-195 protects against development of dementia, and its inhibition (knockdown by antisense microRNA – lenti-pre-AMO-miR-195) leads to impairment of spatial memory (Morris water maze) (Ai et al., 2013). Potential targets for miR-195 are APP and BACE1 proteins associated with β-amyloid aggregation.

MicroRNA cluster miR-183/96/182. Enhanced expression of the microRNA cluster miR-183/96/182 in the hippocampus promotes LTM formation (an object recognition task), and miR-183/96/182 expression is regulated by protein phosphatase PP1 (Woldemichael et al., 2016). An increase in the levels of miR-183/96/182 leads to suppression of histone deacetylase HDAC9 activity and promotes LTM formation. It is known that HDAC9 negatively affects LTM through deacetylation of histones and chromatin remodeling (Grinkevich, 2012; Fischer, 2014). Downregulation of the miR-183/96/182 cluster leads to memory impairment in the old age, and memory can be improved by overexpression of this cluster (Jawaid et al., 2019).

Thus, miRNAs (miR-9-3p, miR-92, miR-195 and the miR-183/96/182 miRNA cluster), which contribute to LTM formation, repress mRNAs encoding proteins that inhibit LTP (Dmd, Sap97), structural plasticity (KCC2, CPEB3, MEF2D) and gene silencing (histone deacetylase HDAC9), as well as proteins causing β-amyloid aggregation (APP and BACE1). Below is a simplified diagram of microRNA-dependent regulation of long-term memory formation (Fig. 1).

MicroRNAs that can influence learning and long-term memory in both positive and negative ways

On the other hand, to date, a large number of miRNAs have been described that can influence formation of LTM in both positive and negative ways.

miR-132. CREB1-dependent activation of miR-132 expression in the hippocampus is observed during formation of fear-induced memory, matches the kinetics of inducing immediate early genes and regulates the spine size in the presynapses (Nudelman et al., 2010). While in transgenic mice (tTA::miR132), with increased expression of miR-132 in forebrain neurons, violation of LTM formation is observed (Hansen et al., 2010). It should be noted that in this case, another type of learning was studied – recognition of new objects. Nevertheless, the authors note that, in tTA::miR132 transgenic mice, there is a decrease in the expression of MeCP2, a protein involved in development of the Rett syndrome and other mental disorders, as well as a noticeable increase in dendritic spine density in the hippocampus, which ultimately should lead to improving LTM. The authors associate all these inconsistencies with too high expression of miR-132.

miR-34. The data on the effect of the miR-34 family, consisting of three members, miR-34a, miR-34b, and miR-34c,

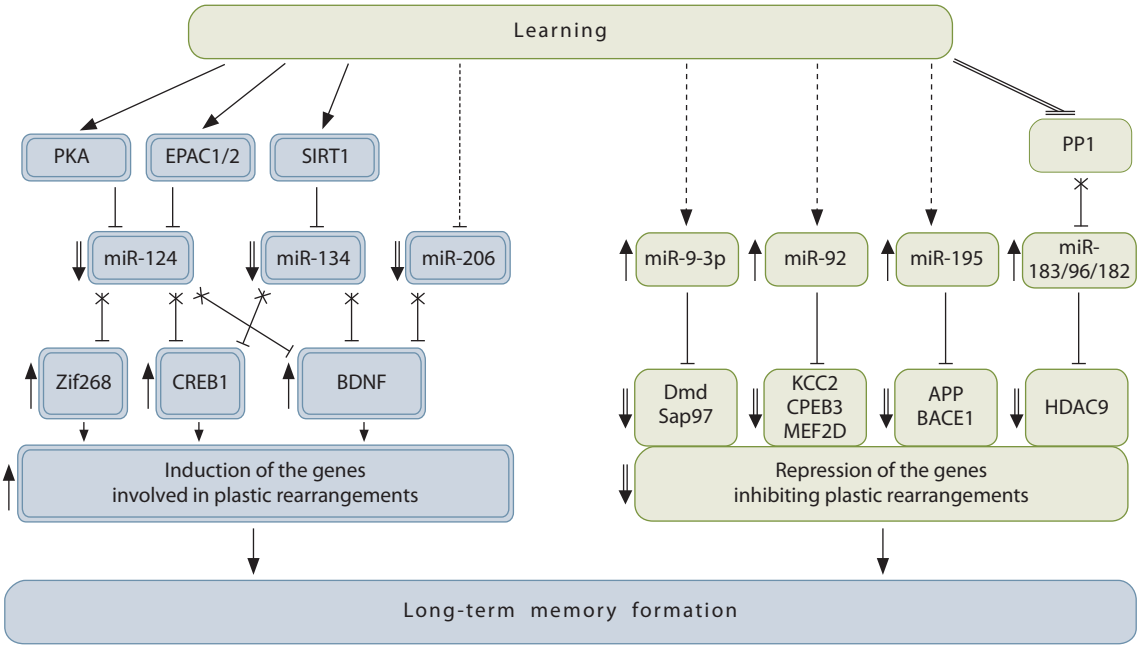


Fig. 1. MicroRNA-dependent regulation of gene expression in the long-term memory formation.
During training, both inhibition of expression of a number of miRNAs (miR-124, miR-134, miR-206 – marked with a double line) and activation (miR-9-3p, miR-92, miR-195 and cluster miR-183/96/182 – marked with a single line) can occur. Downregulation of miRNAs is needed to reactivate genes, the products of which are necessary for plastic rearrangements and are regulated by transcription factors Zif268, CREB1 and by the growth factor BNDF (all marked with a double line). MicroRNAs which are activated during training repress mRNAs encoding proteins that inhibit structural and synaptic plasticity (Dmd, Sap97, KCC2, CPEB3 and MEF2D), proteins promoting β-amyloid aggregation (APP and BACE1) and also proteins repressing synthesis of histone deacetylase HDAC9 (all marked with a single line).
PKA – protein kinase A; EPAC1/EPAC2 – intracellular cAMP receptors; SIRT1-deacetylase and PP1-protein phosphatase are involved in the regulation of microRNAs shown in the diagram. The pathways of regulation of microRNAs that are not described in the literature analyzed in this review are indicated by dashed lines.

on the learning of microRNAs are even more diverse. It has been shown that inhibition of all the members of miRNAs of this family in hippocampal neurons with using AAV-delivered miRNA sponges reduces the ability for reference memory and causes transcriptome changes associated with transduction of neuroactive ligand-receptors and cell communication (Malmevik et al., 2016). Inhibition of miR-34a alone has a similar effect on LTM in rats. In this case, we are talking about amygdala-dependent memory (auditory fear conditioning) (Dias et al., 2014). The learning-induced increase in miR-34a amount in the basolateral amygdala suppresses the Notch pathway. Given the leading role of the Notch pathway proteins in embryonic development and synapse maturation, the authors believe that Notch signaling normally maintains the steady state of synaptic stability by suppressing synaptic plasticity. Fear-mediated, transient increases in miR-34a in the amygdala reduce Notch signaling, thereby creating an environment that temporarily allows synaptic modification and hence long-term memory consolidation.

Unlike miR-34a, miR-34c is a negative factor in memory consolidation. The level of miR-34c increases in the hippocampus during aging, which contributes to impairment of learning and memory, can be restored by inhibiting this microRNA (Zovoilis et al., 2011). In this study, the effect of miR-34c was attributed, at least in part, to a decrease in the target of miR-34c deacetylase SIRT1. The content of miR-34c is also increased in the hippocampus in patients with Alzheimer's dementia and in hippocampal neurons in APPPS1-21 transgenic mice exhibiting β-amyloid pathology and cognitive deficits at an early age (Zovoilis et al., 2011). In turn, SIRT1 modulates synaptic plasticity and memory formation through a miR-134-mediated mechanism (Gao et al., 2010). Thus, miR-34a is positively associated with amygdala-dependent LTM, while miR-34c is negatively associated with hippocampus-dependent LTM.

miR-137. This microRNA is being intensively studied both in connection with the mechanisms of LTM and with various cognitive pathologies, and its functions are very diverse. It has been shown that activation of miR-137 is required for long-term memory formation in the pond snail *Lymnaea* (Korneev et al., 2018). At the same time, miR-137 inhibits the transcription factor CREB2, a negative regulator of the expression of genes necessary for LTM formation. In mice knocked out of the miR-137 gene, spatial learning and memory are impaired, which is potentially associated with increased expression of the Ezh2 gene (Yan et al., 2019). Ezh2 encodes histone-lysine N-methyltransferase involved in the methylation of histones at sites that inhibit gene expression. Thus, it has been shown that miRNAs are capable of modulating another regulatory pathway, namely, epigenetic chromatin remodeling. Data obtained in heterozygous mice with partial loss of miR-137 function support the activating effect of miR-137 on LTM (Cheng et

al., 2018). On the other hand, overexpression of miR-137 in the dentate gyrus of the hippocampus disrupts presynaptic plasticity and impairs fear-induced context memory, while the expression of presynaptic target genes associated with the release of synaptic vesicles (complexin-1, Nsf, and synaptotagmin-1) decreases (Siegert et al., 2015). miR-137 is also implicated in the Pb-induced hippocampus-dependent spatial memory impairment (Gu X. et al., 2019). It was shown that chronic oral administration of lead acetate (PbAc) with drinking water causes a change in the genomic landscape of histone H3 methylation at the H3K27me3 site in the hippocampus. It should be noted that the change in methylation is associated with activated interaction of miR-137 and EZH2 methyltransferase, which make up a mutually inhibitory loop. Overexpression of EZH2 in PbAc-treated rats reverses H3K27me3 methylation and partially restores spatial memory.

miR-153 also has a multidirectional effect on LTM. Thus, the expression of miR-153 is specifically induced in the hippocampus during fear-dependent memory acquisition (Mathew et al., 2016). At the same time, miR-153 inhibits the expression of key components of the vesicular transport system, reduces the level of the glutamate receptor A1 trafficking and neurotransmitter release. On the other hand, knockdown of miR-153 in the hippocampus of adult mice leads to improvement in the fear memory. The authors explain the resulting contradiction by the fact that miR-153, along with, possibly, other fear-induced miRNAs, acts as a component of a feedback loop that blocks neuronal hyperactivity by inhibiting the vesicular transport pathway (Mathew et al., 2016). Dysregulation of miR-153 has been associated with decreased learning and memory ability in autistic mice (You et al., 2019). It has been shown that the target of miR-153 is the LEPR (a leptin receptor) and the JAK-STAT signaling pathway regulated by it. Overexpression of miR-153 suppresses LEPR and the JAK-STAT signaling pathway, which leads to an increase in BDNF expression, an increase in the proliferative capacity of hippocampal neurons, and promotes LTM formation. That is, the high expression of miR-153, and not its knockdown, as stated in the work of Mathew et al. (Mathew et al., 2016), improves LTM formation.

miR-182. An increase in the expression of this miRNA in the hippocampus within the miR-183/96/182 miRNA cluster promotes the hippocampus-dependent LTM formation (Woldemichael et al., 2016). On the contrary, in the amygdala, during formation of amygdala-dependent LTM, a decrease in the amount of miR-182 is noted, and its artificial overexpression leads to a disruption of LTM (Griggs et al., 2013). The miR-182 targets in the amygdala are the key actin-regulating proteins, cortactin and Rac1. Interestingly, another member of the miR-183/96/182 cluster, miR-96, is not expressed in the amygdala. The mechanisms of independent functioning of some microRNAs belonging to clusters are currently not clear, but presumably, this phenomenon is associated with the type of cells in which their differential expression occurs (Banks et al., 2020).

Thus, miR-132, miR-34, miR-137, miR-153, and miR-182 may influence LTM formation in both positive and negative ways, depending on the learning paradigm and the brain structures involved in learning. In addition, the effect of microRNAs on memory may often depend on their concentration. Thus, a moderate increase in miR-212/132 facilitates, and an excessive increase in its expression negatively affects, learning and memory (Benito et al., 2018). The important role of miRNAs in LTM formation is also evidenced by the recent studies related to sleep deprivation.

MicroRNAs, long-term memory and sleep

It is widely known that even a short period of sleep deprivation may impair memory formation. Sleep disturbance, caused by emotional overload, the rugged rhythm of life and chronic life stress, causes a decrease in performance and cognitive functions in a significant number of the world population. It is believed that one of the mechanisms of the effect of sleep deprivation on cognitive processes may be impaired epigenetic regulation of gene expression, including genes associated with microRNA dysfunction (Gaine et al., 2018).

Thus, the recent studies have shown that sleep deprivation significantly changes the profiles of DNA methylation and, accordingly, the synthesis of RNA, including microRNA (Nilsson et al., 2016). On the other hand, miRNAs are involved in the regulation of circadian rhythms that regulate the sleep and wakefulness cycles (Gaine et al., 2018). Disruption of microRNA biogenesis may lead to changes in the circadian rhythms and potentially affect cognitive abilities. In patients with depression and late insomnia, genetic variants of miR-182 miRNA were found that induce inhibition of expression of circadian clock proteins CLOCK and DSIP (Saus et al., 2010). In addition, impaired expression of miR-182, along with miR-132 and miR-124, is observed during the paradoxical sleep phase deprivation and leads to disruption of hippocampus-dependent LTM (Karabulut et al., 2019). At the same time, the synthesis of the growth factor BDNF, which is involved in memory consolidation during sleep, changes markedly (Karabulut et al., 2019). The relation of these microRNAs with numerous cognitive processes has been described in the chapters above. In addition, miR-132 is a key pathway for coupling the circadian rhythm and the rhythm of cognitive abilities (Aten et al., 2018).

Sleep deprivation also disrupts the content of miRNAs let-7b, miR-125a, and miR-138 (Gaine et al., 2018). It is believed that induction of the epigenetic processes caused by sleep deprivation is carried out by signaling cascades that regulate synaptic plasticity (Havekes, Abel, 2017). Sleep disturbance is common in people with fear-related anxiety disorders. It has been shown that some microRNAs, such as miR-132 and miR-144-3p, play an important role both in generation of fear and in suppression of memories of it and are associated with consolidation and reconsolidation of LTM, respectively (Murphy, Singewald, 2018). At the same time, impaired miR-132 expression is observed during sleep deprivation and is accompanied by cognitive dysfunctions (Karabulut et al., 2019).

Perspectives for improving cognitive processes by influencing microRNA biogenesis

In recent years, more and more data have been accumulated that microRNAs play an important role in cognitive disorders in neurological, neurodegenerative, and age-related dysfunctions (Ramakrishna, Muddashetty, 2019; Wingo et al., 2020; Wu, Kuo, 2020). Coverage of these issues requires a separate review. However, it is important to note that there is an increasing number of predictions regarding the possibility of therapeutic treatment of a number of cognitive impairments by influencing microRNA biogenesis (Cao, Zhen, 2018; Paul et al., 2020; Wingo et al., 2020). Optimism in this area is determined by the emergence of new genome editing technologies using the CRISPR/Cas9 system adapted to microRNA (Aquino-Jarquin, 2017). In addition, CRISPR-Cas9 systems have been developed, which allows editing of the genome in a specific cell population without affecting other organs and tissues (Hirosawa et al., 2017; Hoffmann et al., 2019). These technologies are especially important in studying the mechanisms of the central nervous system functioning and the prospects for therapeutic intervention in the pathogenesis of brain diseases.

On the other hand, it has been known for a long time that cognitive processes can be improved through intensification of cognitive processes, fine motor movement work or physical exercise. In recent years, it has been shown that mental and physical activities improve the epigenetic processes involved in formation of LTM and protect neurons from death. For example, running exercise helps to improve memory in mice with traumatic brain injury (Hu T. et al., 2015). At the same time, there is a decrease in the content of miR-21 and an increase in the number of branch points of the hippocampal neurons. Physical exercise also improves the cognitive function in aged mice (Jessop, Toledo-Rodriguez, 2018). The process involves miR-137, which is associated with good memory and neurogenesis in adults (the rate of neurogenesis decreases with age). In addition, the possibility of improving aging-related memory decline by enriching the environment (a combination of cognitive training and physical exercise) has been shown (Jawaid et al., 2019). In this case, an increase in the biogenesis of the miR-183/96/182 cluster is stimulated, which is closely associated with hippocampus-dependent memory. Environmental enrichment also attenuates mild cognitive impairment by activating the SIRT1/miR-134 signaling pathway in the hippocampus, followed by ultrastructure changes of synapses and dendritic remodeling (Shen et al., 2019). Moreover, it has been shown that an environmental enrichment is capable of enhancing synaptic plasticity and cognition even in the next generation, with sperm RNA, and especially miRs 212/132, mediating the effect (Benito et al., 2018). Widely available ways to improve cognitive abilities are shown in Fig. 2.

Mental work, physical exercise, manual creativity, light stress, good sleep and a good mood are able to protect neurons from death and improve cognitive processes via epigenetic mechanisms on chromatin remodeling and miRNA expression.

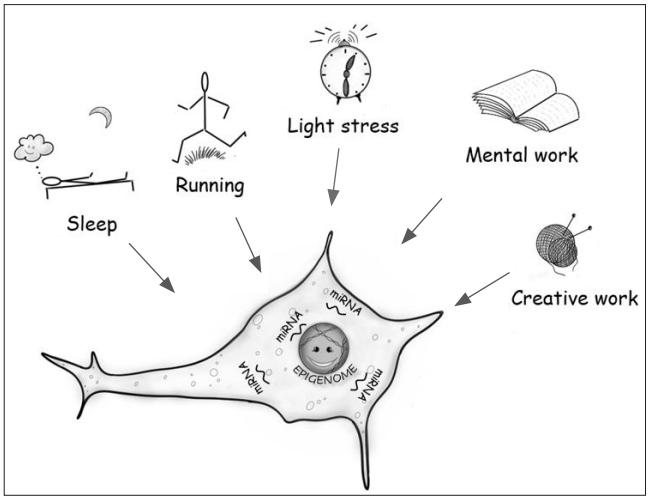


Fig. 2. Ways to improve cognitive abilities.

Conclusion

Thus, miRNAs are widely involved in the regulation of gene expression required for the long-term memory formation. Further study of ways to regulate microRNA activity in individual cell populations, as well as detailed study of their targets using bioinformatics analysis methods, will help to better understand the molecular genetic basis of long-term memory and potentially to develop methods of treatment in case of cognitive dysfunctions in neurodegenerative pathologies and senile dementia.

References

Ai J., Sun L.H., Che H., Zhang R., Zhang T.Z., Wu W.C., Su X.L., Chen X., Yang G., Li K., Wang N., Ban T., Bao Y.N., Guo F., Niu H.F., Zhu Y.L., Zhu X.Y., Zhao S.G., Yang B.F. MicroRNA-195 protects against dementia induced by chronic brain hypoperfusion via its anti-amyloidogenic effect in rats. *J. Neurosci.* 2013;33(9):3989-4001. DOI 10.1523/JNEUROSCI.1997-12.2013. https://www.ncbi.nlm.nih. gov/pmc/articles/PMC6619292.

Aksoy-Aksel A., Zampa F., Schratt G. MicroRNAs and synaptic plasticity – a mutual relationship. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2014;369(1652):20130515. DOI 10.1098/rstb.2013.0515. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4142036.

Aquino-Jarquin G. Emerging role of CRISPR/Cas9 technology for microRNAs editing in cancer research. *Cancer Res.* 2017;77(24):6812-6817. DOI 10.1158/0008-5472.CAN-17-2142. https://cancerres.aacrjournals.org/content/77/24/6812.long.

Aten S., Hansen K.F., Snider K., Wheaton K., Kalidindi A., Garcia A., Alzate-Correa D., Hoyt K.R., Obrietan K. miR-132 couples the circadian clock to daily rhythms of neuronal plasticity and cognition. *Learn. Mem.* 2018;25(5):214-229. DOI 10.1101/lm.047191.117. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5903403.

Baby N., Alagappan N., Dheen S.T., Sajikumar S. MicroRNA-134-5p inhibition rescues long-term plasticity and synaptic tagging/capture in an Aβ(1-42)-induced model of Alzheimer's disease. *Aging Cell.* 2020;19(1):e13046. DOI 10.1111/acel.13046. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6974725.

Baek S., Hwan C., Kim J. Ebf3-miR218 regulation is involved in the development of dopaminergic neurons. *Brain Res.* 2014;1587:

23-32. DOI 10.1016/j.brainres.2014.08.059. https://pubmed.ncbi.nlm.nih.gov/25192643.

Banks S.A., Pierce M.L., Soukup G.A. Sensational microRNAs: neurosensory roles of the microRNA-183 family. *Mol. Neurobiol.* 2020;57(1):358-371. DOI 10.1007/s12035-019-01717-3. https://link.springer.com/article/10.1007%2Fs12035-019-01717-3.

Bartel D.P. MicroRNAs: target recognition and regulatory functions. *Cell.* 2009;136:215-233. DOI 10.1016/j.cell.2009.01.002. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3794896.

Benito E., Kerimoglu C., Ramachandran B., Pena-Centeno T., Jain G., Stilling R.M., Islam M.R., Capece V., Zhou Q., Edbauer D., Dean C., Fischer A. RNA-dependent intergenerational inheritance of enhanced synaptic plasticity after environmental enrichment. *Cell Rep.* 2018;23(2):546-554. DOI 10.1016/j.celrep.2018.03.059. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5912949.

Berger S.L. The complex language of chromatin regulation during transcription. *Nature.* 2007;447(7143):407-412. DOI 10.1038/nature 05915. https://pubmed.ncbi.nlm.nih.gov/17522673.

Beveridge N.J., Gardiner E., Carroll A.P., Tooney P.A., Cairns M.J. Schizophrenia is associated with an increase in cortical microRNA biogenesis. *Mol. Psychiatry.* 2010;15(12):1176-1189. DOI 10.1038/mp.2009.84. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC299 0188.

Bicker S., Khudayberdiev S., Weiss K., Zocher K., Baumeister S., Schratt G. The DEAH-box helicase DHX36 mediates dendritic localization of the neuronal precursor-microRNA-134. *Genes Dev.* 2013;27(9):991-996. DOI 10.1101/gad.211243.112. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3656329.

Bitetti A., Mallory A.C., Golini E., Carrieri C., Carreño Gutiérrez H., Perlas E., Pérez-Rico Y.A., Tocchini-Valentini G.P., Enright A.J., Norton W.H.J., Mandillo S., O'Carroll D., Shkumatava A. MicroRNA degradation by a conserved target RNA regulates animal behavior. *Nat. Struct. Mol. Biol.* 2018;25(3):244-251. DOI 10.1038/s41594-018-0032-x. https://pubmed.ncbi.nlm.nih.gov/29483647.

Cao T., Zhen X.C. Dysregulation of miRNA and its potential therapeutic application in schizophrenia. *CNS Neurosci. Ther.* 2018;24(7):586-597. DOI 10.1111/cns.12840.

Chen W., Qin C. General hallmarks of microRNAs in brain evolution and development. *RNA Biol.* 2015;12(7):701-708. DOI 10.1080/15476286.2015.1048954. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4615839.

Cheng Y., Wang Z.M., Tan W., Wang X., Li Y., Bai B., Li Y., Zhang S.F., Yan H.L., Chen Z.L., Liu C.M., Mi T.W., Xia S., Zhou Z., Liu A., Tang G.B., Liu C., Dai Z.J., Wang Y.Y., Wang H., Wang X., Kang Y., Lin L., Chen Z., Xie N., Sun Q., Xie W., Peng J., Chen D., Teng Z.Q., Jin P. Partial loss of psychiatric risk gene miR137 in mice causes repetitive behavior and impairs sociability and learning via increased Pde10a. *Nat. Neurosci.* 2018;21(12):1689-1703. DOI 10.1038/s41593-018-0261-7. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6261680.

Chmielarz P., Konovalova J., Najam S.S., Alter H., Piepponen T.P., Erfle H., Sonntag K.C., Schütz G., Vinnikov I.A., Domanskyi A. Dicer and microRNAs protect adult dopamine neurons. *Cell Death Dis.* 2017;8(5):e2813. DOI 10.1038/cddis.2017.214. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5520729.

Danka Mohammed C.P., Park J.S., Nam H.G., Kim K. MicroRNAs in brain aging. *Mech. Ageing Dev.* 2017;168:3-9. DOI 10.1016/j.mad.2017.01.007. https://pubmed.ncbi.nlm.nih.gov/28119001.

Dias B.G., Goodman J.V., Ahluwalia R., Easton A.E., Andero R., Ressler K.J. Amygdala-dependent fear memory consolidation via miR-34a and notch signaling. *Neuron.* 2014;83(4):906-918. DOI 10.1016/j.neuron.2014.07.019. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4172484.

Dimmeler S., Nicotera P. MicroRNAs in age-related diseases. *EMBO Mol. Med.* 2013;5(2):180-190. DOI 10.1002/emmm.201201986. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3569636.

Fiorenza A., Barco A. Role of Dicer and the miRNA system in neuronal plasticity and brain function. *Neurobiol. Learn. Mem.* 2016;135:3-12. DOI 10.1016/j.nlm.2016.05.001. https://pubmed.ncbi.nlm.nih.gov/27163737.

Fiorenza A., Lopez-Atalaya J.P., Rovira V., Scandaglia M., Geijo-Barrientos E., Barco A. Blocking miRNA biogenesis in adult forebrain neurons enhances seizure susceptibility, fear memory, and food intake by increasing neuronal responsiveness. *Cereb. Cortex.* 2016;26:1619-1633. DOI 10.1093/cercor/bhu332. https://pubmed.ncbi.nlm.nih.gov/25595182.

Fischer A. Epigenetic memory: the Lamarckian brain. *EMBO J.* 2014;33(9):945-967. DOI 10.1002/embj.201387637. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4193930.

Gaine M.E., Chatterjee S., Abel T. Sleep deprivation and the epigenome. *Front. Neural Circuits*. 2018;12:14. DOI 10.3389/fncir.2018.00014. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5835037.

Gantier M.P., McCoy C.E., Rusinova I., Saulep D., Wang D., Xu D., Irving A.T., Behlke M.A., Hertzog P.J., Mackay F., Williams B.R. Analysis of microRNA turnover in mammalian cells following *Dicer1* ablation. *Nucleic Acids Res.* 2011;39(13):5692-5703. DOI 10.1093/nar/gkr148. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3141258.

Gao J., Wang W.Y., Mao Y.W., Gräff J., Guan J.S., Pan L., Mak G., Kim D., Su S.C., Tsai L.H. A novel pathway regulates memory and plasticity via SIRT1 and miR-134. *Nature.* 2010;466(7310):1105-1109. DOI 10.1038/nature09271. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2928875.

Griggs E.M., Young E.J., Rumbaugh G., Miller C.A. MicroRNA-182 regulates amygdala-dependent memory formation. Version 2. *J. Neurosci.* 2013;33(4):1734-1740. DOI 10.1523/JNEUROSCI.2873-12.2013. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3711533.

Grinkevich L.N. Epigenetics and long-term memory formation. *Rossiyskiy Fiziologicheskiy Zhurnal im. I.M. Sechenova = I.M. Sechenov Physiological Journal.* 2012;98(5):553-574. https://pubmed.ncbi.nlm.nih.gov/22838191/ (in Russian)

Grinkevich L.N. Influence of PLL treatment on the long-term memory formation in *Helix* mollusk. *Meditsynskiy Akademicheskiy Zhurnal = Medical Academic Journal.* 2019;19(4):87-92. DOI 10.17816/MAJ19080. https://journals.eco-vector.com/MAJ/article/view/19080. (in Russian)

Gu Q.H., Yu D., Hu Z., Liu X., Yang Y., Luo Y., Zhu J., Li Z. miR-26a and miR-384-5p are required for LTP maintenance and spine enlargement. *Nat. Commun.* 2015;6:6789. DOI 10.1038/ncomms7789. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4403380.

Gu X., Xu Y., Xue W.Z., Wu Y., Ye Z., Xiao G., Wang H.L. Interplay of miR-137 and EZH2 contributes to the genome-wide redistribution of H3K27me3 underlying the Pb-induced memory impairment. *Cell Death Dis.* 2019;10(9):671. DOI 10.1038/s41419-019-1912. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6739382.

Hansen K.F., Sakamoto K., Wayman G.A., Impey S., Obrietan K. Transgenic miR132 alters neuronal spine density and impairs novel object recognition memory. *PLoS One.* 2010;5(11):e15497. DOI 10.1371/journal.pone.0015497. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2993964.

Havekes R., Abel T. The tired hippocampus: the molecular impact of sleep deprivation on hippocampal function. *Curr. Opin. Neu-*

robiol. 2017;44:13-19. DOI 10.1016/j.conb.2017.02.005. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5511071.

He L., Hannon G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 2004;5(7):522-531. DOI 10.1038/nrg1379. https://pubmed.ncbi.nlm.nih.gov/15211354.

Hébert S.S., Papadopoulou A.S., Smith P., Galas M.C., Planel E., Silahtaroglu A.N., Sergeant N., Buée L., De Strooper B. Genetic ablation of Dicer in adult forebrain neurons results in abnormal tau hyperphosphorylation and neurodegeneration. *Hum. Mol. Genet.* 2010;19(20):3959-3969. DOI 10.1093/hmg/ddq311. https://pubmed.ncbi.nlm.nih.gov/20660113.

Hirosawa M., Fujita Y., Parr C.J.C., Hayashi K., Kashida S., Hotta A., Woltjen K., Saito H. Cell-type-specific genome editing with a microRNA-responsive CRISPR-Cas9 switch. *Nucleic Acids Res.* 2017;45(13):e118. DOI 10.1093/nar/gkx309. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5570128.

Hoffmann M.D., Aschenbrenner S., Grosse S., Rapti K., Domenger C., Fakhiri J., Mastel M., Börner K., Eils R., Grimm D., Niopek D. Cell-specific CRISPR-Cas9 activation by microRNA-dependent expression of anti-CRISPR protein. *Nucleic Acids Res.* 2019;47(13):e75. DOI 10.1093/nar/gkz271. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6648350.

Hu T., Zhou F.J., Chang Y.F., Li Y.S., Liu G.C., Hong Y., Chen H.L., Xiyang Y.B., Bao T.H. miR21 is associated with the cognitive improvement following voluntary running wheel exercise in TBI mice. *J. Mol. Neurosci.* 2015;57(1):114-122. DOI 10.1007/s12031-015-0584-8. https://pubmed.ncbi.nlm.nih.gov/26018937.

Hu Z., Li Z. miRNAs in synapse development and synaptic plasticity. *Curr. Opin. Neurobiol.* 2017;45:24-31. DOI 10.1016/j.conb.2017.02.014. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5554733.

Hu Z., Yu D., Gu Q.H., Yang Y., Tu K., Zhu J., Li Z. miR-191 and miR-135 are required for long-lasting spine remodelling associated with synaptic long-term depression. *Nat. Commun.* 2014;5:3263. DOI 10.1038/ncomms4263. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3951436.

Hu Z., Zhao J., Hu T., Luo Y., Zhu J., Li Z. miR-501-3p mediates the activity-dependent regulation of the expression of AMPA receptor subunit GluA1. *J. Cell Biol.* 2015;208(7):949-959. DOI 10.1083/jcb.201404092. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC43 84731.

Inukai S., de Lencastre A., Turner M., Slack F. Novel microRNAs differentially expressed during aging in the mouse brain. *PLoS One.* 2012;7:e40028. DOI 10.1371/journal.pone.0040028. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402511.

Jawaid A., Woldemichael B.T., Kremer E.A., Laferriere F., Gaur N., Afroz T., Polymenidou M., Mansuy I.M. Memory decline and its reversal in aging and neurodegeneration involve miR-183/96/182 biogenesis. *Mol. Neurobiol.* 2019;56(5):3451-3462. DOI 10.1007/s12035-018-1314-3. https://pubmed.ncbi.nlm.nih.gov/30128653.

Jessop P., Toledo-Rodriguez M. Hippocampal *TET1* and *TET2* expression and DNA hydroxymethylation are affected by physical exercise in aged mice. *Front. Cell Dev. Biol.* 2018;6:45. DOI 10.3389/fcell.2018.00045. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC 5922180.

John B., Enright A.J., Aravin A., Tuschl T., Sander C., Marks D.S. Human microRNA targets. *PLoS Biol.* 2004;2(11):e363. DOI 10.1371/journal.pbio.0020363. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC521178.

Jovasevic V., Corcoran K.A., Leaderbrand K., Yamawaki N., Guedea A.L., Chen H.J., Shepherd G.M., Radulovic J. GABAergic mechanisms regulated by miR-33 encode state-dependent fear. *Nat. Neurosci.* 2015;18(9):1265-1271. DOI 10.1038/nn.4084. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4880671.

Kandel E. Small neuron systems. In: The Brain. Scientific American, 1979.

Kandel E. The molecular biology of memory: cAMP, PKA, CRE, CREB-1, CREB-2, and CPEB. *Mol. Brain.* 2012;5(14):1-12. DOI 10.1186/1756-6606-5-1426. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3514210.

Karabulut S., Korkmaz Bayramov K., Bayramov R., Ozdemir F., Topaloglu T., Ergen E., Yazgan K., Taskiran A.S., Golgeli A. Effects of post-learning REM sleep deprivation on hippocampal plasticity-related genes and microRNA in mice. *Behav. Brain Res.* 2019;361:7-13. DOI 10.1016/j.bbr.2018.12.045. https://pubmed.ncbi.nlm.nih.gov/30594545.

Kim S., Kaang B.K. Epigenetic regulation and chromatin remodeling in learning and memory. *Exp. Mol. Med.* 2017;49(1):e281. DOI 10.1038/emm.2016.140. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5291841.

Konopka W., Kiryk A., Novak M., Herwerth M., Parkitna J.R., Wawrzyniak M., Kowarsch A., Michaluk P., Dzwonek J., Arnsperger T., Wilczynski G., Merkenschlager M., Theis F.J., Köhr G., Kaczmarek L., Schütz G. MicroRNA loss enhances learning and memory in mice. *J. Neurosci.* 2010;30(44):14835-14842. DOI 10.1523/JNEUROSCI.3030-10.2010. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6633640.

Korneev S.A., Vavoulis D.V., Naskar S., Dyakonova V.E., Kemenes I., Kemenes G. A CREB2-targeting microRNA is required for long-term memory after single-trial learning. *Sci. Rep.* 2018;8(1):3950. DOI 10.1038/s41598-018-22278-w. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5834643.

Lee R.C., Feinbaum R.L., Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell.* 1993;75(5):843-854. DOI 10.1016/0092-8674(93)90529-y. https://pubmed.ncbi.nlm.nih.gov/8252621.

Lee S.T., Chu K., Jung K.H., Kim J.H., Huh J.Y., Yoon H., Park D.K., Lim J.Y., Kim J.M., Jeon D., Ryu H., Lee S.K., Kim M., Roh J.K. miR-206 regulates brain-derived neurotrophic factor in Alzheimer disease model. *Ann. Neurol.* 2012;72:269-277. DOI 10.1002/ana.23588. https://pubmed.ncbi.nlm.nih.gov/22926857.

Lesseur C., Paquette A.G., Marsit C.J. Epigenetic regulation of infant neurobehavioral outcomes. *Med. Epigenet.* 2014;2(2):71-79. DOI 10.1159/000361026. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4116357.

Leung A.K.L. The whereabouts of microRNA actions: cytoplasm and beyond. *Trends Cell Biol.* 2015;25(10):601-610. DOI 10.1016/j.tcb.2015.07.005. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4610250.

Lewis B.P., Shih I.-H., Jones-Rhoades M.W., Bartel D.P., Burge C.B. Prediction of mammalian microRNA targets. *Cell.* 2003;115(7):787-798. DOI 10.1016/s0092-8674(03)01018-3. https://pubmed.ncbi.nlm.nih.gov/14697198.

Lin Q., Ponnusamy R., Widagdo J., Choi J.A., Ge W., Probst C., Buckley T., Lou M., Bredy T.W., Fanselow M.S., Ye K., Sun Y.E. MicroRNA-mediated disruption of dendritogenesis during a critical period of development influences cognitive capacity later in life. *Proc. Natl. Acad. Sci. USA.* 2017;114(34):9188-9193. DOI 10.1073/pnas.1706069114. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC 5576812.

Liu E.Y., Cali C.P., Lee E.B. RNA metabolism in neurodegenerative disease. *Dis. Model. Mech.* 2017;10(5):509-518. DOI 10.1242/dmm.028613. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5451173.

Lugli G., Larson J., Martone M.E., Jones Y., Smalheiser N.R. Dicer and eIF2c are enriched at postsynaptic densities in adult mouse brain and are modified by neuronal activity in a calpain-dependent manner. *J. Neurochem.* 2005;94(4):896-905. DOI 10.1111/

j.1471-4159.2005.03224.x. https://pubmed.ncbi.nlm.nih.gov/16092937.

Malmevik J., Petri R., Knauff P., Brattas P.L., Akerblom M., Jakobsson J. Distinct cognitive effects and underlying transcriptome changes upon inhibition of individual miRNAs in hippocampal neurons. *Sci. Rep.* 2016;6:19879. DOI 10.1038/srep19879. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4728481.

Mathew R.S., Tatarakis A., Rudenko A., Johnson-Venkatesh E.M., Yang Y.J., Murphy E.A., Todd T.P., Schepers S.T., Siuti N., Martorell A.J., Falls W.A., Hammack S.E., Walsh C.A., Tsai L.H., Umemori H., Bouton M.E., Moazed D.A. microRNA negative feedback loop downregulates vesicle transport and inhibits fear memory. *eLife.* 2016;5:e22467. DOI 10.7554/eLife.22467. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5293492.

McNeill E., Van Vactor D. MicroRNAs shape the neuronal landscape. *Neuron.* 2012;75(3):363-379. DOI 10.1016/j.neuron.2012.07.005. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3441179.

Murphy C.P., Singewald N. Potential of microRNAs as novel targets in the alleviation of pathological fear. *Genes Brain Behav.* 2018;17(3):e12427. DOI 10.1111/gbb.12427. https://onlinelibrary.wiley.com/doi/full/10.1111/gbb.12427.

Nilsson E.K., Boström A.E., Mwinyi J., Schiöth H.B. Epigenomics of total acute sleep deprivation in relation to genome-wide DNA methylation profiles and RNA expression. *OMICS.* 2016;20(6):334-342. DOI 10.1089/omi.2016.0041. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4926204.

Nudelman A.S., DiRocco D.P., Lambert T.J., Garelick M.G., Le J., Nathanson N.M., Storm D.R. Neuronal activity rapidly induces transcription of the CREB-regulated microRNA-132, *in vivo*. *Hippocampus.* 2010;20(4):492-498. DOI 10.1002/hipo.20646. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2847008.

Paul S., Reyes P.R., Garza B.S., Sharma A. MicroRNAs and child neuropsychiatric disorders: a brief review. *Neurochem. Res.* 2020;45(2):232-240. DOI 10.1007/s11064-019-02917-y. https://pubmed.ncbi.nlm.nih.gov/31773374.

Rajasethupathy P., Fiumara F., Sheridan R., Betel D., Puthanveettil S.V., Russo J.J., Sander C., Tuschl T., Kandel E. Characterization of small RNAs in *Aplysia* reveals a role for miR-124 in constraining synaptic plasticity through CREB. *Neuron.* 2009;63(6):803-817. DOI 10.1016/j.neuron.2009.05.029. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2875683.

Ramakrishna S., Muddashetty R.S. Emerging role of microRNAs in dementia. *J. Mol. Biol.* 2019;431(9):1743-1762. DOI 10.1016/j.jmb.2019.01.046. https://pubmed.ncbi.nlm.nih.gov/30738891.

Reinhart B.J., Slack F.J., Basson M., Pasquinelli A.E., Bettinger J.C., Rougvie A.E., Horvitz H.R., Ruvkun G. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature.* 2000;403(6772):901-906. DOI 10.1038/35002607. https://pubmed.ncbi.nlm.nih.gov/10706289.

Saus E., Soria V., Escaramis G., Vivarelli F., Crespo J.M., Kagerbauer B., Menchón J.M., Urretavizcaya M., Gratacòs M., Estivill X. Genetic variants and abnormal processing of pre-miR-182, a circadian clock modulator, in major depression patients with late insomnia. *Hum. Mol. Genet.* 2010;19(20):4017-4025. DOI 10.1093/hmg/ddq316. https://pubmed.ncbi.nlm.nih.gov/20656788.

Selbach M., Schwanhäusser B., Thierfelder N., Fang Z., Khanin R., Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. *Nature.* 2008;455(7209):58-63. DOI 10.1038/nature07228. https://pubmed.ncbi.nlm.nih.gov/18668040.

Shen J., Li Y., Qu C., Xu L., Sun H., Zhang J. The enriched environment ameliorates chronic unpredictable mild stress-induced depressive-like behaviors and cognitive impairment by activating the SIRT1/miR-134 signaling pathway in hippocampus. *J. Affect Disord.* 2019;248:81-90. DOI 10.1016/j.jad.2019.01.031. https://pubmed.ncbi.nlm.nih.gov/30716615.

Siegert S., Seo J., Kwon E.J., Rudenko A., Cho S., Wang W., Flood Z., Martorell A.J., Ericsson M., Mungenast A.E., Tsai L.H. The schizophrenia risk gene product miR-137 alters presynaptic plasticity. *Nat. Neurosci.* 2015;18(7):1008-1016. DOI 10.1038/nn.4023. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4506960.

Sim S.E., Lim C.S., Kim J.I., Seo D., Chun H., Yu N.K., Lee J., Kang S.J., Ko H.G., Choi J.H., Kim T., Jang E.H., Han J., Bak M.S., Park J.E., Jang D.J., Baek D., Lee Y.S., Kaang B.K. The brain-enriched microRNA miR-9-3p regulates synaptic plasticity and memory. *J. Neurosci.* 2016;36(33):8641-8652. DOI 10.1523/JNEUROSCI.0630-16.2016. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6601897.

Smalheiser N.R. The RNA-centred view of the synapse: non-coding RNAs and synaptic plasticity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2014;369(1652):20130504. DOI 10.1098/rstb.2013.0504. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4142025.

Smith A.C.W., Kenny P.J. MicroRNAs regulate synaptic plasticity underlying drug addiction. *Genes Brain Behav.* 2018;17(3):e12424. DOI 10.1111/gbb.12424. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5837931.

Sweatt J.D. Neural plasticity and behavior – sixty years of conceptual advances. *J. Neurochem.* 2016;139(Suppl.2):179-199. DOI 10.1111/jnc.13580. https://pubmed.ncbi.nlm.nih.gov/26875778.

Vetere G., Barbato C., Pezzola S., Frisone P., Aceti M., Ciotti M., Cogoni C., Ammassari-Teule M., Ruberti F. Selective inhibition of miR-92 in hippocampal neurons alters contextual fear memory. *Hippocampus.* 2014;24(12):1458-1465. DOI 10.1002/hipo.22326. https://pubmed.ncbi.nlm.nih.gov/24990518.

Wang C.N., Wang Y.J., Wang H., Song L., Chen Y., Wang J.L., Ye Y., Jiang B. The anti-dementia effects of Donepezil involve miR-206-3p in the hippocampus and cortex. *Biol. Pharm. Bull.* 2017;40(4):465-472. DOI 10.1248/bpb.b16-00898. https://pubmed.ncbi.nlm.nih.gov/28123152.

Wang X., Liu D., Huang H.Z., Wang Z.H., Hou T.Y., Yang X., Pang P., Wei N., Zhou Y.F., Dupras M.J., Calon F., Wang Y.T., Man H.Y., Chen J.G., Wang J.Z., Hébert S.S., Lu Y., Zhu L.Q. A novel microRNA-124/PTPN1 signal pathway mediates synaptic and memory deficits in Alzheimer's disease. *Biol. Psychiatry.* 2018;83(5):395-405. DOI 10.1016/j.biopsych.2017.07.023. https://pubmed.ncbi.nlm.nih.gov/28965984.

Wingo T.S., Yang J., Fan W., Min Canon S., Gerasimov E.S., Lori A., Logsdon B., Yao B., Seyfried N.T., Lah J.J., Levey A.I., Boyle P.A., Schneider J.A., De Jager P.L., Bennett D.A., Wingo A.P. Brain microRNAs associated with late-life depressive symptoms are also associated with cognitive trajectory and dementia. *NPJ Genom. Med.* 2020;5:6. DOI 10.1038/s41525-019-0113-8. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7004995.

Woldemichael B.T., Jawaid A., Kremer E.A., Gaur N., Krol J., Marchais A., Mansuy I.M. The microRNA cluster miR-183/96/182 contributes to long-term memory in a protein phosphatase 1-dependent manner. *Nat. Commun.* 2016;7:12594. DOI 10.1038/ncomms12594. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5007330.

Wu Y.Y., Kuo H.C. Functional roles and networks of non-coding RNAs in the pathogenesis of neurodegenerative diseases. *J. Biomed. Sci.* 2020;27(1):49. DOI 10.1186/s12929-020-00636-z. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7140545.

Yan H.L., Sun X.W., Wang Z.M., Liu P.P., Mi T.W., Liu C., Wang Y.Y., He X.C., Du H.Z., Liu C.M., Teng Z.Q. MiR-137 deficiency causes anxiety-like behaviors in mice. *Front. Mol. Neuro-*

sci. 2019;12:260. DOI 10.3389/fnmol.2019.00260. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6831983.

Yang Y., Shu X., Liu D., Shang Y., Wu Y., Pei L., Xu X., Tian Q., Zhang J., Qian K., Wang Y.X., Petralia R.S., Tu W., Zhu L.Q., Wang J.Z., Lu Y. EPAC null mutation impairs learning and social interactions via aberrant regulation of miR-124 and Zif268 translation. *Neuron.* 2012;73(4):774-788. DOI 10.1016/j.neuron.2012.02.003. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3307595.

You Y.H., Qin Z.Q., Zhang H.L., Yuan Z.H., Yu X. MicroRNA-153 promotes brain-derived neurotrophic factor and hippocampal neuron proliferation to alleviate autism symptoms through inhibition of JAK-STAT pathway by LEPR. *Biosci. Rep.* 2019;39(6):BSR20181904. DOI 10.1042/BSR20181904. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6591574.

Zovoilis A., Agbemenyah H.Y., Agis-Balboa R.C., Stilling R.M., Edbauer D., Rao P., Farinelli L., Delalle I., Schmitt A., Falkai P., Bahari-Javan S., Burkhardt S., Sananbenesi F., Fischer A. MicroRNA-34c is a novel target to treat dementias. *EMBO J.* 2011;30:4299-4308. DOI 10.1038/emboj.2011.327. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3199394.

ORCID ID
L.N. Grinkevich orcid.org/0000-0003-3744-5946

Acknowledgements. This work was supported by the Program of Fundamental Research of State Academies for 2013–2020 (GP-14, section 63).
Conflict of interest. The author declares no conflict of interest.
Received July 2, 2020. Revised October 11, 2020. Accepted October 15, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):897-904
DOI 10.18699/VJ20.688

Evaluation of various RNA-seq approaches for identification of gene outrons in the flatworm *Opisthorchis felineus*

N.I. Ershov[1] ✉, D.E. Maslov[2], N.P. Bondar[1, 2]

[1] Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
✉ e-mail: ershov@bionet.nsc.ru

Abstract. The parasitic flatworm *Opisthorchis felineus* is one of the causative agents of opisthorchiasis in humans. Recently, we assembled the *O. felineus* genome, but the correct genome annotation by means of standard methods was hampered by the presence of spliced leader trans-splicing (SLTS). As a result of SLTS, the original 5′-end (outron) of the transcripts is replaced by a short spliced leader sequence donated from a specialized SL RNA. SLTS is involved in the RNA processing of more than half of *O. felineus* genes, making it hard to determine the structure of outrons and *bona fide* transcription start sites of the corresponding genes and operons, being based solely on mRNA-seq data. In the current study, we tested various experimental approaches for identifying the sequences of outrons in *O. felineus* using massive parallel sequencing. Two of them were developed by us for targeted sequencing of already processed branched outrons. One was based on sequence-specific reverse transcription from the SL intron toward the 5′-end of the Y-branched outron. The other used outron hybridization with an immobilized single-stranded DNA probe complementary to the SL intron. Additionally, two approaches to the sequencing of rRNA-depleted total RNA were used, allowing the identification of a wider range of transcripts compared to mRNA-seq. One is based on the enzymatic elimination of overrepresented cDNAs, the other utilizes exonucleolytic degradation of uncapped RNA by Terminator enzyme. By using the outron-targeting methods, we were not able to obtain the enrichment of RNA preparations by processed outrons, which is most likely indicative of a rapid turnover of these trans-splicing intermediate products. Of the two rRNA depletion methods, a method based on the enzymatic normalization of cDNA (Zymo-Seq RiboFree) showed high efficiency. Compared to mRNA-seq, it provides an approximately twofold increase in the fraction of reads originating from outrons and introns. The results suggest that unprocessed nascent transcripts are the main source of outron sequences in the RNA pool of *O. felineus*.
Key words: opisthorchiasis; spliced leader trans-splicing; outron; start of transcription; transcriptome; ribosomal RNA.

For citation: Ershov N.I., Maslov D.E., Bondar N.P. Evaluation of various RNA-seq approaches for identification of gene outrons in the flatworm *Opisthorchis felineus*. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):897-904. DOI 10.18699/VJ20.688

Апробация различных вариантов RNA-seq для идентификации аутронов генов у плоского червя *Opisthorchis felineus*

Н.И. Ершов[1] ✉, Д.Е. Маслов[2], Н.П. Бондарь[1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
✉ e-mail: ershov@bionet.nsc.ru

Аннотация. *Opisthorchis felineus* – представитель паразитических плоских червей, один из возбудителей описторхоза человека. Недавно нами была проведена сборка генома *O. felineus*, однако корректная аннотация генов в этом геноме стандартными методами оказалась затруднена наличием сплайс-лидер зависимого транс-сплайсинга (SLTS). В результате SLTS исходный 5′-конец (аутрон) транскриптов заменяется короткой сплайс-лидерной последовательностью, донором которой выступает специализированная молекула SL РНК. SLTS вовлечен в процессинг РНК более половины всех генов *O. felineus*, из-за чего становится невозможным установить последовательности аутронов и реальные старты транскрипции соответствующих генов и оперонов, опираясь только на данные mRNA-seq. В настоящей работе мы провели апробацию различных экспериментальных подходов для идентификации последовательностей аутронов у *O. felineus* с помощью массового параллельного секвенирования. Два подхода были спланированы нами для прицельного секвенирования процессированных разветвленных аутронов. Первый заключался в сиквенс-специфичной обратной транскрипции с SL-интрона в направлении 5′-конца аутрона. Во втором использовалась гибридизация аутронов с иммобилизованным одноцепочечным ДНК-зондом, комплементарным SL-интрону. Также были использованы два подхода к секвенированию тотальной РНК, обедненной по рРНК, позволяющих иденти-

фицировать более широкий спектр транскриптов, чем mRNA-seq. Один из них основан на ферментативной элиминации перепредставленных кДНК, другой – на ферментативной деградации некэпированных РНК экзонуклеазой Terminator. С помощью селективных методов нам не удалось получить обогащения препаратов РНК по процессированным аутронам, что, наиболее вероятно, связано с коротким временем жизни этих промежуточных продуктов транс-сплайсинга. Из двух методов обеднения по рРНК высокую эффективность показал метод, основанный на ферментативной нормализации кДНК (Zymo-Seq RiboFree). Он позволил примерно вдвое увеличить долю прочтений, соответствующих аутронам и интронам, по сравнению с mRNA-seq. Полученные результаты предполагают, что основным ресурсом последовательностей аутронов в пуле РНК *O. felineus* служат новосинтезированные непроцессированные транскрипты.
Ключевые слова: описторхоз; сплайс-лидер зависимый транс-сплайсинг; аутрон; старт транскрипции; транскриптом; рибосомальная фракция РНК.

Introduction

Opisthorchis felineus is a representative of parasitic flatworms (Trematoda: Opisthorchiidae), which has a complex life cycle with two intermediate hosts and the mammalian definitive host, including humans (Beer, 2005). Opisthorchiasis caused by parasitism of this fluke in the bile ducts of the human liver has a chronic course and leads to a number of serious concomitant disorders of the hepatobiliary system, including cholangitis, cholecystitis, pancreatitis, and is also a risk factor for the development of cholangiocarcinoma (Sripa et al., 2007; Pakharukova, Mordvinov, 2016; Pakharukova et al., 2019). According to WHO study, more than 1 million people estimated to be infected with *O. felineus*, with the largest focus of opisthorchiasis located in the Ob-Irtysh basin, Russia, with the incidence in some regions of the Tomsk and Tyumen regions up to 60 % (FAO/WHO, 2014; Fedorova et al., 2018).

One of the necessary steps to study the biology of this flatworm, and the corresponding development of molecular genetic approaches to the diagnostics and pharmacotherapy of opisthorchiasis, was to obtain a reference genome assembly and its annotation (Ershov et al., 2019). However, the existing gene annotation based on polyA-mRNA-seq data, which is commonly used for this task, has significant drawbacks due to one of the features of RNA processing in flatworms – spliced leader trans-splicing (SLTS). During SLTS, the exons of two independent transcripts are fused: a specialized short capped spliced leader RNA (SL RNA) carrying a 5′-splice site, and one or the other pre-RNA with a corresponding 3′-site (see Fig. 1, *b*). At the same time, the original 5′-region of the pre-RNA, together with the SL RNA intron attached to its branchpoint through the 2′–5′-phosphodiester bond, is excised (Murphy et al., 1986; Sutton, Boothroyd, 1988). It is assumed that this Y-branched product (Y-outron) undergoes rapid debranching and degradation (Sutton, Boothroyd, 1988; Lasda, Blumental, 2011). Thus, in the polyA-mRNA-seq data, the sequences of excised outrons are practically absent, and with them, localization of the actual transcriptional start sites and promoters of the corresponding genes, as well as information on the operonic organization of gene groups, becomes inaccessible. It should be noted that due to SLTS, a large group of experimental methods for identifying transcription starts based on selective sequencing of the capped 5′-ends of RNAs is also ineffective.

According to our estimates, products of more than half of all genes in the *O. felineus* genome undergo trans-splicing (Ershov et al., 2019). Thus, the correct annotation of such genes requires alternative high-throughput sequencing methods for large-scale identification of outrons. Most often, to describe gene outrons, researchers combine information on transcription initiation sites obtained by such methods as GRO-seq, GRO-cap, ChIP-seq to RNA polymerase II, with data on SLTS sites detected by mRNA-seq (Chen et al., 2013; Kruesi et al., 2013). In this way, the region of the genome from the start of transcription to the 3′-site of the SLTS is accepted as an outron. It should be noted that most of the mentioned methods require a large amount of starting biological material and, in the case of a small size of individuals, is suitable mainly for organisms cultured in the laboratory. There are few known examples when it was possible to detect outrons directly by RNA sequencing methods. Thus, in one of the studies on *C. elegans*, growth of individuals at low temperature led to the accumulation of unprocessed transcripts, which made it possible to detect the 5′-ends of outrons by SAGE of the nuclear fraction of RNA (Saito et al., 2013).

On non-model organisms, including trematodes, RNA-seq "SL Trapping" methods have been successfully used for direct massive detection of trans-spliced mRNAs, based on selection for the universal 5′-sequence of processed transcripts, the SL exon (Nilsson et al., 2010; Boroni et al., 2018). Since all Y-outrons similarly contain the universal 3′-sequence of the SL intron, it seems attractive to use this property for selective identification of outrons and symmetric confirmation of SLTS sites. In addition, direct identification of Y-outrons would serve as direct evidence of the generally accepted mechanism of SLTS on a large sample of genes.

In the current study, to identify *O. felineus* outrons, we evaluated two candidate approaches based on targeted enrichment for Y-outrons, in one case, by sequence-specific reverse transcription primed from the SL intron towards the 5′-end of the outron, in the other – by hybridization of Y-outrons with an immobilized single-stranded DNA probe complementary to the SL-intron. In addition, two approaches to sequencing of rRNA-depleted total RNA that do not use hybridization probes were used as an alternative: one that use rRNA cleavage by Terminator exonuclease, and the commercial Zymo-Seq RiboFree kit based on enzymatic normalization of cDNA. The proportion of sequences containing outrons in these libraries should certainly be higher than in standard polyA-mRNA-seq libraries.

Materials and methods

Biomaterial. Adult *O. felineus* worms were isolated from the bile ducts of golden hamsters *Mesocricetus auratus* after 3–4 months after infection with metacercariae obtained from the tissues of naturally infected fish (*Leuciscus idus*) from the

Ob river. The worms were washed in saline buffer and used immediately or frozen and stored at –80 °C.

RNA isolation. RNA was isolated from fresh or frozen *O. felineus* samples using PureZOL (BioRad, USA) according to the manufacturer's protocol. Precipitation of RNA was carried out by adding an equal volume of isopropanol and 5 μg of linear polyacrylamide (LPA) as a co-precipitant; the resulting solution was left overnight at –20 °C. The precipitate was dissolved in bidistilled water, and additional purification of samples was carried out using an Aurum total RNA mini kit (BioRad, USA), including a stage of treatment with DNase I. The quality and amount of isolated RNA was assessed using a NanoDrop 2000 spectrophotometer.

Terminator exonucleolytic cleavage. Total RNA was treated with Terminator enzyme (5′-Phosphate-Dependent Exonuclease, Epicentre, USA) according to the manufacturer's protocol. The reaction mixture contained 500 ng of total RNA, 1 μl of RiboLock (ThermoFisher Scientific, USA), 2 μl of 10× buffer A, 1 U of Terminator exonuclease. The reaction volume was brought up to 20 μl with bidistilled water. The sample was incubated for 60 min at 30 °C, the reaction was stopped by the addition of 1 μl of 100 mM EDTA. The products of the reaction were purified using Agencourt RNAClean XP beads (Beckman Coulter, USA). The decrease in the amount of ribosomal RNA was checked on an Agilent 2100 Bioanalyzer using an RNA 6000 Pico chip.

Isolation of RNA fraction enriched in Y-outrons (SLi-BC). A biotinylated ssDNA probe ([biotin-TEG]-5′-GGC TAGCCAAATAATTCATCCGACCATAGGCCGGAGTC GATTCTT-3′) was immobilized on magnetic beads (Dynabeads M-280 Streptavidin, Invitrogen, USA) in accordance with the manufacturer's protocol, monitoring the approximate correspondence of the amount of bound probe to the declared binding capacity (~200 pmol/mg). 1 μg of RNA in a buffer for hybridization (1 M NaCl, 5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 0.1 % Tween-20) was added to the magnetic beads covered with a DNA probe and incubated with constant stirring according to the following protocol: 75 °C for 2 minutes, then the temperature was lowered by 1 °C/min to 55 °C and fixed for 15 min at 55 °C. The particles were washed three times with warm (50 °C) buffer. 10 μl of water was added to the particles and heated to 94 °C for 5 min. The RNA-containing solution was collected in a clean tube and the reverse transcription reaction was carried out immediately. For the reverse transcription reaction, 200 U RevertAid Reverse Transcriptase (ThermoFisher Scientific, USA) and random hexamers were used. The reaction was carried out under the following conditions: 5 min at 25 °C, 60 min at 42 °C, and 5 min at 70 °C. The resulting cDNA was used to prepare the library.

Reverse transcription reaction with sequence-specific primers (SLi-RT). The thermostable Maxima H Minus Reverse Transcriptase (ThermoFisher Scientific, USA) was used for the reverse transcription reaction. The reaction mixture containing 300 ng RNA, 2 pmol of a specific primer to the SL-RNA intron (SLi_r1, 5′-AGGCCGGAGTCGATTCTT-3′), 1 μl 10 mM dNTP, 4 μl 5 M betaine, was incubated for 2 min at 75 °C, then the temperature was lowered at a rate of 2 °C/min to 55 °C; 20 U RiboLock RNase Inhibitor and 100 U Maxima H Minus Reverse Transcriptase were added and the mixture incubated for 30 min at 55 °C and for 5 min at 85 °C to inactivate the reaction. The resulting cDNA was used to prepare the library.

Evaluation of Y-outron enrichment of SLi-RT and SLi-BC cDNA samples using real-time PCR (RT-PCR). To assess the enrichment for Y-outrons during the elaboration of the SLi-BC and SLi-RT protocols, RT-PCR with primers to the outron (MMCE_ou_f: 5′-CCTGGCGACACACATCTG AA-3′, MMCE_ou_r: 5′-ACATGGACATGGCTGAAGCA-3′) and exons (MMCE_ex_f: 5′-TGCAACCTCTCTTGTGTT CCT-3′, MMCE_ex_r: 5′-CCACCTGGACACCGAATG TAT-3′) of the *mmce* gene was carried out (Supplementary Material 1)[1]. Enrichment was calculated by the $\Delta\Delta C_t$ method as the change in the difference of outron from exon between control and selected cDNA. The control was cDNA obtained by reverse transcription of total RNA primed from random hexamers. In the case of SLi-RT, the selected cDNA was obtained by reverse transcribing the total RNA from the SLi_r1 primer, and in the case of SLi-BC, by converting the enriched RNA fraction using random hexamers.

PCR was carried out in a 20 μl reaction mixture containing 0.25 mM dNTP, 2 μL 10X PCR buffer B+EVAGreen, 2.5 mM $MgCl_2$, 10 pmol of each primer, 2 μl cDNA sample, 0.3 U/μl SynTaq DNA polymerase with antibodies inhibiting polymerase activity (Syntol, Russia). The reaction was carried out with preheating to 95 °C for 5 minutes followed by 39 amplification cycles including denaturation at 95 °C for 15 seconds, primer annealing and elongation at 60 °C for 20 seconds. Melting curves were acquired within the temperature range 65 to 95 °C. All reactions were performed in two technical replicates.

Library preparation. To prepare RNA-seq libraries after Terminator treatment, NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, USA) was used according to the standard protocol. To prepare libraries from cDNA samples obtained by the SLi-BC and SLi-RT protocols, NEBNext Ultra II Directional RNA Library Prep Kit for Illumina was used, starting from the stage of the second strand cDNA synthesis. Fragmentation of the cDNA was performed using the dsDNA Fragmentase enzyme simultaneously with the repair of the DNA ends. Fragments of the required length were selected using Agencourt AMPure XP beads, after which several cycles of library amplification were performed.

rRNA-depleted RNA-seq libraries were prepared using the Zymo-Seq RiboFree Total RNA Library Prep Kit (Zymo Research, USA) according to the standard protocol. 500 ng of total RNA was used in the reaction; depletion after renaturation was carried out for 60 min.

The size and quantity of the resulting libraries was determined using an Agilent 2100 Bioanalyzer (Suppl. Material 2). The resulting libraries were sequenced in paired-end mode (2 × 250 bp) on the Illumina MiSeq platform (the service was provided by Vector-Best, Russia). Raw sequencing data has been deposited in the open-access repository Zenodo (Ershov, 2020). Previously published mRNA-seq data used in the study are available in the NCBI repository (PRJNA257351).

Computational processing of sequencing data. Sequencing data in the FASTQ format was processed using the Cutadapt software (Martin, 2011) to remove adapter sequences

[1] Supplementary Materials 1–3 are available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/SupplErshov_engl.pdf

and mapped to the *O. felineus* reference genome (GenBank ID GCA_004794785.1) using the STAR aligner (Dobin et al., 2013). The annotation of genomic elements was constructed using the corresponding gene annotation and previously obtained genome-wide data on positions of trans-splicing sites (Ershov et al., 2019) using in-house Perl scripts. The assembly of the genomic rRNA repeat was carried out manually using the data of genomic DNA sequencing (Ershov et al., 2019). The boundaries of rRNA genes in the genome were determined using the Rfam database (https://rfam.xfam.org) and the RNAmmer tool (Lagesen et al., 2007). Read counts in the defined genomic intervals were computed using the featureCounts tool from the Subread package (Liao et al., 2019). Statistical processing and graphical data visualization were carried out in the R environment.

Results and discussion

Description of the approaches

At the moment, we have not been able to find any reference in the literature to experimental methods for the direct massive identification of Y-outrons in the transcriptome. Upon examination of the previously obtained *O. felineus* mRNA-seq data (Ershov et al., 2019), we found that the 5′-regions upstream of the trans-splicing site (potential outrons) of many genes have the same weak read coverage as their introns (Fig. 1, *a*), despite the depletion of non-coding and unprocessed transcripts as a result of polyA selection. In total, about 0.15 % of mRNA-seq reads were mapped to outronic regions (5 kb upstream of the site) of trans-spliced transcripts (on average with 0.8-fold coverage). The source of such outronic reads can be: (1) unprocessed nascent pre-RNA, (2) Y-outrons – intermediate Y-branched byproducts of trans-splicing, as well as (3) products of their further debranching, which lack the covalently bound SL-intronic arm (see Fig. 1, *b*). We assumed that the amount of Y-outrons (2) in the total RNA pool could potentially be sufficient for its isolation by targeted enrichment methods.

The possibility of targeting Y-outrons is due to the presence of a universal sequence – an SL-RNA intronic arm, covalently bound at a branchpoint to outron of pre-RNA (see Fig. 1, *b*). Using this property, we have developed two different approaches to direct selection of Y-outrons. The first, designated SLi-RT (SL-intron Reverse Transcription), is a targeted reverse transcription from the 2′-arm (SL-intron) towards the varying 5′-end of the outron (see Fig. 1, *c*). It is known that M-MLV reverse transcriptase lacking RNase H activity, when primed from the 2′-arm, is able to quite frequently bypass the branchpoint with the introduction of single-mismatch errors and successfully reverse the 5′-segment (Bitton et al., 2014; Döring, Hurek, 2017). Among the expected drawbacks of the method are the reduced efficiency of such reverse transcription, the accumulation of products of mispriming, and the underrepresentation of the 5′-regions of outrons due to a frequent accidental termination of the first strand cDNA synthesis. In accordance with the latter, the choice of enzyme and high temperature reaction conditions, including hot start, were optimized for maximum processivity and synthesis specificity.

The second approach, SLi-BC (SL-intron Biotin Capture), is based on hybridization of the 2′-arm of outrons with a biotinylated single-stranded DNA probe homologous to the SL intron, followed by immobilization and purification of the corresponding fraction on magnetic particles coated with streptavidin (see Fig. 1, *d*). This approach should be more sensitive to the proportion of the target within the total RNA sample, but it has a better chance of identifying full-length outrons.

Both approaches are aimed at the intact Y-structure of the processed outron and therefore would allow differentiating them from fragments of nascent transcripts that did not undergo trans-splicing. In addition, we also applied two methods for sequencing of the rRNA-depleted total RNA, that allow identification of noncoding transcripts, including introns and outrons. At the same time, since *O. felineus* is a non-model species, available depletion kits based on hybridization with rRNA probes are not applicable. We therefore tested a protocol for treatment of total RNA with the Terminator exonuclease, which specifically hydrolyzes nucleic acids with a phosphorylated 5′-end (including rRNA), but does not affect capped transcripts. Alternatively, we applied the commercial Zymo-Seq RiboFree protocol, based on DSN-normalization of RNA:DNA hybrids immediately after the synthesis of the first cDNA strand, as a result of which all overrepresented transcripts, including globins and rRNA, are hydrolyzed. Unlike the first two, these methods do not make it possible to distinguish the processed outrons from their precursors – unprocessed and nascent transcripts, and therefore cannot serve as direct evidence of the trans-splicing event.

The sequences of the previously identified intact SL-RNA (Ershov et al., 2019) were used to design the corresponding primers and oligonucleotide probes. Optimization of the enrichment conditions for Y-outrons in the SLi-BC and SLi-RT methods was controlled by RT-PCR with primers to the outron and exons of the *mmce* gene, the trans-spliced product of which is not polycistronic and is among the most highly expressed transcripts (see Suppl. Material 1). The enrichment was calculated as the difference of ΔCt (outron, exon) between the control and the enriched cDNA samples. In the case of cDNA samples used for the subsequent preparation of the SLi-BC and SLi-RT libraries, there was 4- and 100-fold enrichment in the outron region relative to the *mmce* exons, respectively (see Suppl. Material 1, B).

RNA-seq libraries were prepared from the RNA or cDNA samples obtained by various methods and then sequenced with low coverage, which gives a reasonable indication of the overall performance for enrichment or depletion of target sequences.

Residual rRNA content in libraries

To compare the applied approaches against the standard mRNA-seq method, the obtained sequencing data from five libraries were mapped to a reference genome supplemented with an rRNA repeat sequence. Since the latter repeat had a sufficiently high coverage in all libraries, it turned out to be useful not only for assessing the efficiency of rRNA depletion, but also for assessing the strand-specificity of the resulting libraries (Fig. 2). Of the libraries, only TerminatorExo had a relatively low degree of strand-specificity (68 %). Interestingly, each of the libraries was characterized by its own specific rRNA coverage profile. It is worth mentioning that,

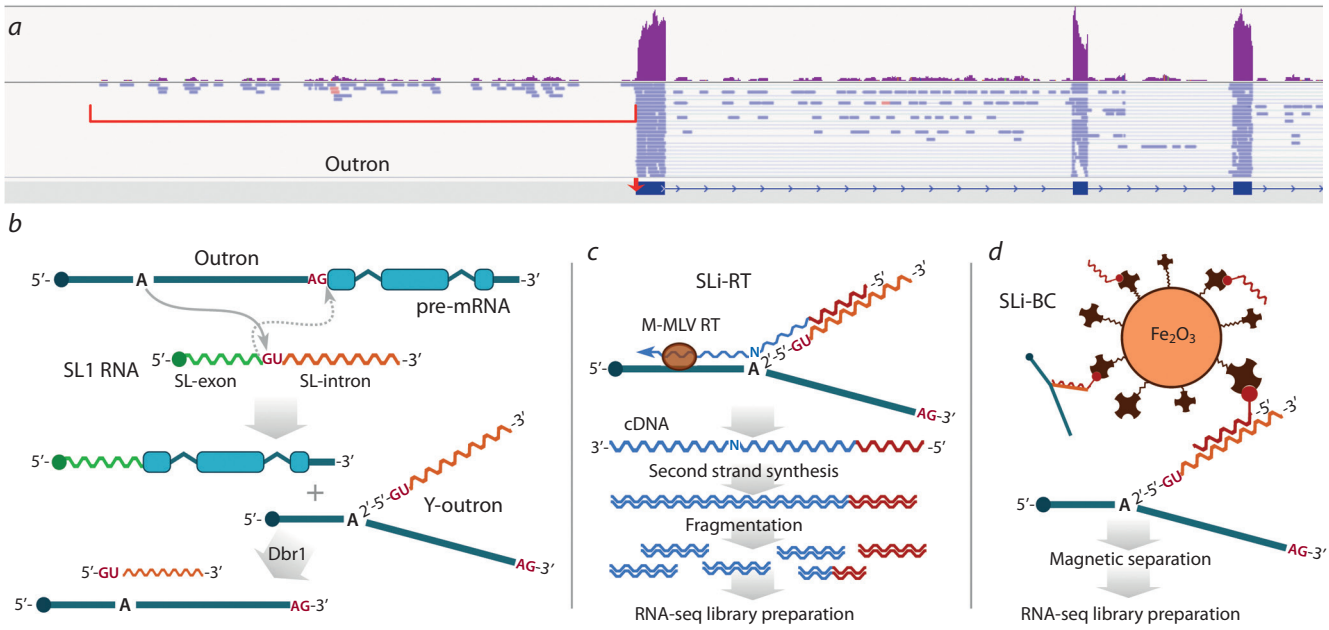


Fig. 1. The presence of outronic sequences in the polyA-mRNA-seq data of *O. felineus* and the graphical overview of the proposed RNA-seq methods for targeted identification of Y-outrons.

a – an example of read coverage of an outronic region of a trans-spliced gene (*Vps39l*) in mRNA-seq data of *O. felineus* adult worm. The arrow indicates the confirmed trans-splicing site; *b* – a schematic overview of trans-splicing of pre-RNA with the formation of an Y-branched outron containing an SL-intron arm, and its subsequent degradation with the participation of Dbr1; *c* – scheme of reverse transcription primed from the SL-intron within the Y-outron, used in the SLi-RT approach; M-MLV RT, M-MLV [H-] reverse transcriptase; *d* – scheme of hybridization of a single-stranded DNA probe modified with 5′-biotin-TEG, with an SL-intron within the Y-outron, and subsequent purification on magnetic beads coated with streptavidin.

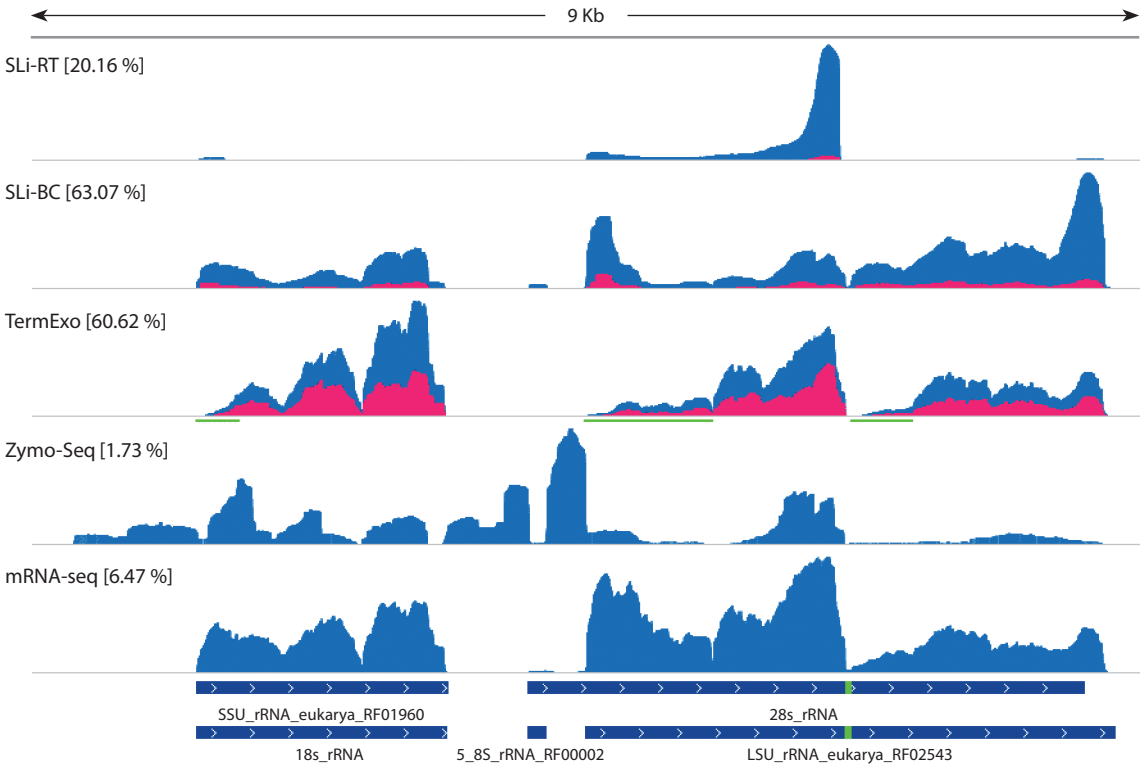


Fig. 2. Coverage profiles of genomic rRNA repeat containing 18S, 5.8S, 28Sα, and 28Sβ rDNA.

Coverage profiles with antisense and sense cDNA reads are highlighted in blue and red. In the designations of the libraries, the percentage of reads mapped to the rRNA repeat is indicated in square brackets. The dark green lines mark the 5′-regions of rRNA that are most susceptible to hydrolysis by the Terminator exonuclease. The bottom track shows the positions of the rRNA genes predicted by the Rfam and RNAmmer 1.2 services, indicating the hidden break locus of the 28S rRNA (green).

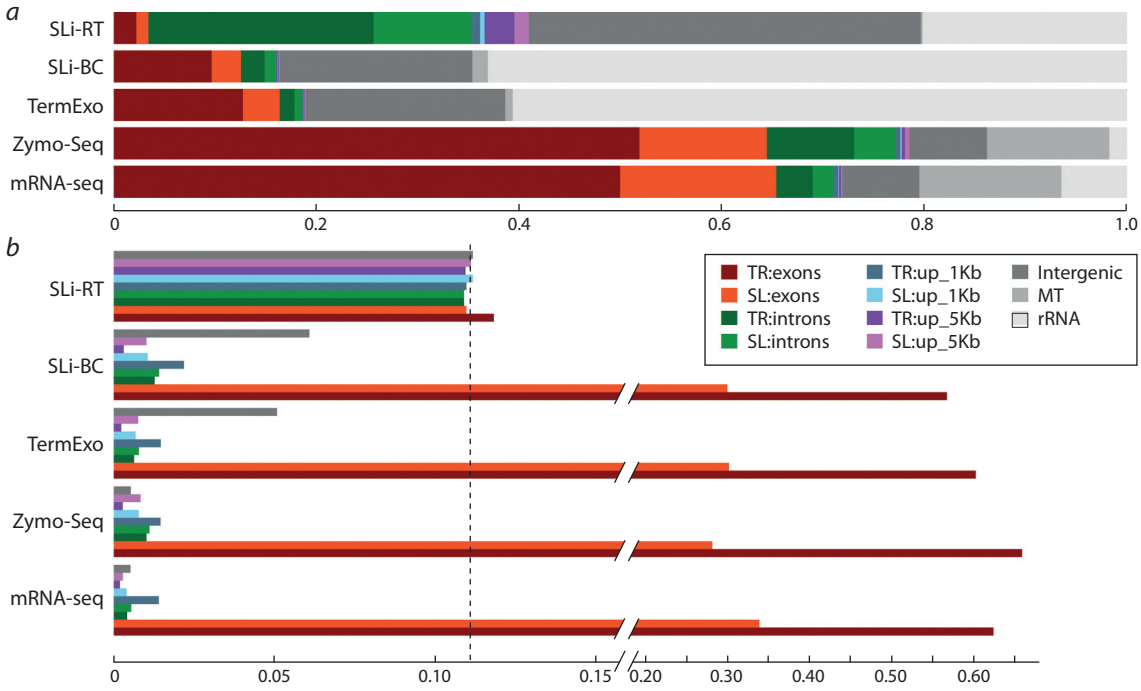


Fig. 3. Representation of various genomic elements in the RNA-seq data produced by different approaches.

a – fractional distribution of mapped reads across 11 categories of genomic elements; *b*– chart showing normalized read counts in each of the 9 genomic categories (excluding mtRNA and rRNA). Values are normalized to total category length and sequencing depth (similar to TPM metric). The dashed line indicates the expected level of values under the condition of uniform genome coverage.

for a large number of protostomes, 28S rRNA is additionally processed at the "hidden break" site with the formation of approximately equivalent in size 28Sα and 28Sβ fragments (Ishikawa, 1977). This feature is also observed for *O. felineus*: a coverage drop in the 28S rRNA (see Fig. 2) and a single rRNA peak in the electropherogram (see Suppl. Material 2).

The observed distribution of the few rRNA reads of the Zymo-Seq library in comparison with mRNA-seq (1.7 and 6.5 % of reads, respectively) reflects extremely effective subtractive depletion in the most overrepresented regions of this repeat.

The TerminatorExo library profile indicates a fairly effective initial enzymatic degradation of the 5′ regions of 18S, 28Sα, and 28Sβ rRNA (see Fig. 2, marked with green lines), which, however, rapidly fades away at certain sites. This is probably due to the sensitivity of the enzyme to the complex tertiary structure of the substrate, or to the frequent hydrolysis of rRNA at these sites with the formation of 5′-OH ends, which protect against further exonucleolytic cleavage by the enzyme. Consequently, the resulting TerminatorExo library contained about 60 % of rRNA reads.

Targeted enrichment methods SLi-BC and SLi-RT also showed high levels of background rRNA (63.1 and 20.2 %). Namely, the SLi-RT library contained mainly the product of off-target reverse transcription from the 3′-region of 28Sα rRNA. Since the 3′-region of 28Sα rRNA can potentially form a stable hairpin (according to the secondary structure predicted by the RNAfold service, Suppl. Material 3), it is likely that such a structure determines the priming of reverse transcription without any participation of the SLi_r1 primer.

Representation of various genomic features in sequenced libraries

For the mapped data, we analyzed the distribution of reads across various mutually exclusive groups of genomic elements in the following priority order: rRNA repeats, mtDNA, exons, introns, promoter or outron regions (1 and 5 Kb upstream of the 5′-end of the predicted gene) and intergenic loci. The calculation was carried out separately for genes undergoing highly efficient trans-splicing (SL) and for all other genes (TR). As can be seen from the results presented in Fig. 3, *a*, in the TerminatorExo, SLi-BC and SLi-RT libraries, the overwhelming proportion of reads is represented by the uninformative rRNA fraction and the category of intergenic loci. In contrast, in the Zymo-Seq and mRNA-seq libraries, more than 60 % of the reads are mapped to exons of the annotated genes.

If we exclude the rRNA and mtRNA groups from consideration (see Fig. 3, *b*), the distribution of read counts across categories of genomic elements, normalized to the total length of the latter, turns out to be quite similar for the mRNA-seq, SLi-BC, TerminatorExo, and Zymo-Seq libraries. The SLi-RT library generally reproduces a random (uniform) distribution of reads across the genome. Obviously, SLi-RT contains mainly noise, which was probably originated from genomic DNA, despite the identical DNase treatment of all RNA samples. It is worth noting that this library contained 227 (0.12 %) read pairs containing the SLi_r1 primer at the beginning of the fragment (next to the adapter sequence). These target sequences were expected to be overrepresented in the SLi-RT library. Although they most likely do not belong to genuine Y-outrons, their presence confirms the efficiency of the primer in the reverse transcription reaction.

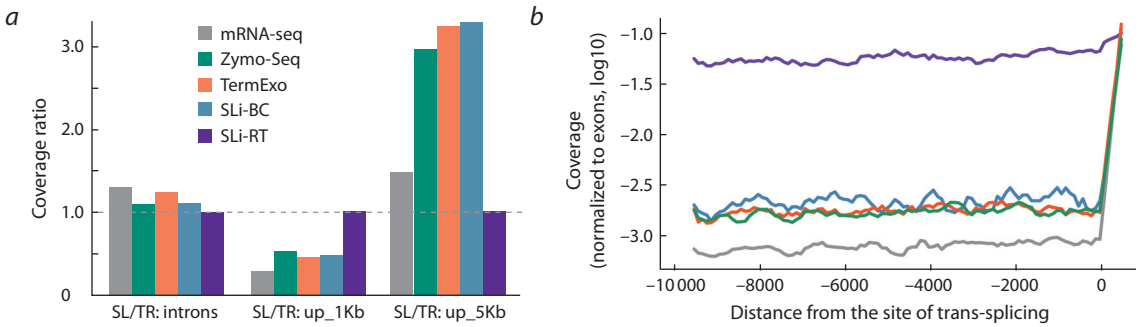


Fig. 4. Enrichment of the libraries by outronic sequences.

a – barplot of the enrichment in introns and outrons of trans-spliced genes (SL) compared to similar regions of non-trans-spliced genes (TR). The coverage is normalized to the total length of the categories; *b* – profiles of the total normalized coverage of potential outrons in the genome in each of five libraries. The *X*-axis shows the distance (bp) from the identified trans-splicing site. The color coding corresponds to the legend in section *a*.

It was assumed that in the case of effective enrichment of libraries by outrons we should observe a significant increase in the coverage of 5′-adjacent regions of trans-spliced genes (SL:up_1Kb and SL:up_5Kb) relative to the promoter regions of ordinary genes (TR:up_1Kb and TR:up_5Kb). This enrichment was indeed observed for the SLi-BC, TerminatorExo and Zymo-Seq libraries, but only in the case of the SL:up_5Kb category (Fig. 4, *a*). In contrast, these libraries demonstrated depletion in potential outrons in SL:up_1Kb. The depletion is most likely due to the fact that for many "TR" genes their 5′-UTRs are not annotated correctly and therefore a large number of reads originating from mature mRNAs are falsely assigned to the TR:up_1Kb category. More revealing is the comparison with mRNA-seq, in which all three mentioned methods were more enriched in both categories of outronic regions. Since the coefficient of enrichment is approximately the same for the three libraries, it describes just the actual fraction of outrons in the initial total RNA. Consequently, the SLi-BC method did not result in preferential selection of target Y-outron sequences.

Thus, of all the tested methods, only Zymo-Seq protocol made it possible to significantly get rid of the uninformative rRNA fraction, while preserving the original fraction of noncoding transcripts, including outrons. At the same time, the coverage of outrons compared to exons in the Zymo-Seq library remains very low (see Fig. 4, *b*).

Taken together, these results suggest that sequences corresponding to outrons including unprocessed or nascent transcripts and intact or partially degraded Y-outrons, are represented by a very small fraction in the *O. felineus* RNA pool. Apparently, intact Y-outrons make up the smallest part of them, and therefore the SLi-BC and SLi-RT libraries targeting the intact Y-outron structure contained mainly unspecific noise. Although the mechanism and rate of degradation of Y-branched trans-splicing products remain unknown, it is assumed that they degrade rapidly (Lasda, Blumenthal, 2011). Thus, in an *in vitro* experiment on *C. elegans*, all trans-splicing intermediates were observed, except for the Y-branched outrons (Hannon et al., 1990).

The main source of noise in the SLi-BC method was the nonspecific sorption of RNA, while in the SLi-RT method, it was off-target reverse transcription of rRNA without the participation of a primer, as well as a minimal admixture of genomic DNA. The detection of the latter only in this library once again indicates that the target RNA template was almost absent in the reverse transcription reaction, namely, outrons with an intact Y-branched structure.

Conclusion

Thus, of the RNA-seq approaches considered, which theoretically allow identification of the outrons of trans-spliced transcripts, the Zymo-Seq RiboFree approach, which uses enzymatic cDNA normalization, turned out to be the most promising. The ineffectiveness of the targeted SLi-BC and SLi-RT methods is likely due to undetectable amounts of Y-branched outrons in the total RNA pool.

References

Beer S.A. Biology of the Agent of Opisthorchiasis. Moscow, 2005. (in Russian)

Bitton D.A., Rallis C., Jeffares D.C., Smith G.C., Chen Y.Y.C., Codlin S., Marguerat S., Bähler J. LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq. *Genome Res.* 2014;24(7):1169-1179. DOI 10.1101/gr.166819.113.

Boroni M., Sammeth M., Gava S.G., Jorge N.A.N., Macedo A.M., Machado C.R., Mourão M.M., Franco G.R. Landscape of the spliced leader trans-splicing mechanism in *Schistosoma mansoni*. *Sci. Rep.* 2018;8(1):3877. DOI 10.1038/s41598-018-22093-3.

Chen R.A.-J., Down T.A., Stempor P., Chen Q.B., Egelhofer T.A., Hillier L.W., Jeffers T.E., Ahringer J. The landscape of RNA polymerase II transcription initiation in *C. elegans* reveals promoter and enhancer architectures. *Genome Res.* 2013;23(8):1339-1347. DOI 10.1101/gr.153668.112.

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21. DOI 10.1093/bioinformatics/bts635.

Döring J., Hurek T. Arm-specific cleavage and mutation during reverse transcription of 2′,5′-branched RNA by Moloney murine leukemia virus reverse transcriptase. *Nucleic Acids Res.* 2017;45(7):3967-3984. DOI 10.1093/nar/gkx073.

Ershov N.I. Evaluation of various RNA-seq approaches for identification of outrons in the flatworm *Opisthorchis felineus* (Version 1.0.0). *Zenodo.* 2020. DOI 10.5281/zenodo.3901531.

Ershov N.I., Mordvinov V.A., Prokhortchouk E.B., Pakharukova M.Y., Gunbin K.V., Ustyantsev K., Genaev M.A., Blinov A.G., Mazur A., Boulygina E., Tsygankova S., Khrameeva E., Chekanov N., Fan G.,

Xiao A., Zhang H., Xu X., Yang H., Solovyev V., Lee S.M., Liu X., Afonnikov D.A., Skryabin K.G. New insights from *Opisthorchis felineus* genome: update on genomics of the epidemiologically important liver flukes. *BMC Genom.* 2019;20(1):399. DOI 10.1186/s12864-019-5752-8.

FAO/WHO [Food and Agriculture Organization of the United Nations/World Health Organization]. Multicriteria-Based Ranking for Risk Management of Food-Borne Parasites. Microbiological Risk Assessment Series. No. 23. Rome, 2014.

Fedorova O.S., Fedotova M.M., Sokolova T.S., Golovach E.A., Kovshirina Y.V., Ageeva T.S., Kovshirina A.E., Kobyakova O.S., Ogorodova L.M., Odermatt P. *Opisthorchis felineus* infection prevalence in Western Siberia: a review of Russian literature. *Acta Trop.* 2018;178:196-204. DOI 10.1016/j.actatropica.2017.11.018.

Hannon G.J., Maroney P.A., Denker J.A., Nilsen T.W. Trans splicing of nematode pre-messenger RNA *in vitro*. *Cell*. 1990;61(7):1247-1255. DOI 10.1016/0092-8674(90)90689-c.

Ishikawa H. Evolution of ribosomal RNA. *Comp. Biochem. Physiol.* 1977;58(1):1-7. DOI 10.1016/0305-0491(77)90116-X.

Kruesi W.S., Core L.J., Waters C.T., Lis J.T., Meyer B.J. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife.* 2013;2:e00808. DOI 10.7554/eLife.00808.

Lagesen K., Hallin P., Rødland E.A., Staerfeldt H.H., Rognes T., Ussery D.W. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007;35(9):3100-3108. DOI 10.1093/nar/gkm160.

Lasda E.L., Blumenthal T. Trans-splicing. *Wiley Interdiscip. Rev. RNA*. 2011;2(3):417-434. DOI 10.1002/wrna.71.

Liao Y., Smyth G.K., Shi W. The R package *Rsubread* is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 2019;47(8):e47. DOI 10.1093/nar/gkz114.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):10-12. DOI 10.14806/ej.17.1.200.

Murphy W.J., Watkins K.P., Agabian N. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans-splicing. *Cell.* 1986;47:517-525. DOI 10.1016/0092-8674(86)90616-1.

Nilsson D., Gunasekera K., Mani J., Osteras M., Farinelli L., Baerlocher L., Roditi I., Ochsenreiter T. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathog.* 2010;6(8):e1001037. DOI 10.1371/journal.ppat.1001037.

Pakharukova M.Y., Mordvinov V.A. The liver fluke *Opisthorchis felineus*: biology, epidemiology, and carcinogenic potential. *Trans. R. Soc. Trop. Med. Hyg.* 2016;110:28-36. DOI 10.1093/trstmh/trv085.

Pakharukova M.Y., Zaparina O.G., Kapushchak Y.K., Baginskaya N.V., Mordvinov V.A. *Opisthorchis felineus* infection provokes time-dependent accumulation of oxidative hepatobiliary lesions in the injured hamster liver. *PLoS One.* 2019;14(5):e0216757. DOI 10.1371/journal.pone.0216757.

Saito T.L., Hashimoto S., Gu S.G., Morton J.J., Stadler M., Blumenthal T., Fire A., Morishita S. The transcription start site landscape of *C. elegans*. *Genome Res.* 2013;23(8):1348-1361. DOI 10.1101/gr.151571.112.

Sripa B., Kaewkes S., Sithithaworn P., Mairiang E., Laha T., Smout M., Pairojkul C., Bhudhisawasdi V., Tesana S., Thinkamrop B., Bethony J.M., Loukas A., Brindley P.J. Liver fluke induces cholangiocarcinoma. *PLoS Med.* 2007;4(7):e201. DOI 10.1371/journal.pmed.0040201.

Sutton R.E., Boothroyd J.C. Trypanosome trans-splicing utilizes 2′-5′ branches and a corresponding debranching activity. *EMBO J.* 1988;7(5):1431-1437.

ORCID ID

N.I. Ershov orcid.org/0000-0003-3423-3497
N.P. Bondar orcid.org/0000-0002-5602-5149

Acknowledgements. This work was supported by the Russian Science Foundation, project 18-74-00101.

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

Received March 21, 2020. Revised November 24, 2020. Accepted November 24, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):905-912
DOI 10.18699/VJ20.689

Transplastomic tobacco plants producing the hydrophilic domain of the sheep pox virus coat protein L1R

D.K. Beisenov ✉, G.E. Stanbekova, B.K. Iskakov

M.A. Aitkhozhin Institute of Molecular Biology and Biochemistry, Almaty, Kazakhstan
✉ e-mail: beisenov.d@gmail.com

Abstract. Sheep pox has a wide geographical range of distribution and poses a threat to sheep breeding worldwide, as the disease is highly contagious and is accompanied by large economic losses. Vaccines based on live attenuated virus strains are currently being used for prevention of this disease. Such vaccines are effective, but potentially dangerous because of the possible virus reversion to a pathogenic state. The development of safe recombinant subunit vaccines against sheep pox is very relevant. The high ploidy level of the plant chloroplasts makes it possible to obtain large quantities of foreign proteins. The purpose of this study was to create transplastomic *Nicotiana tabacum* plants producing one of the candidate vaccine proteins of sheep pox virus L1R. A vector containing a deletion variant of the *SPPV_56* gene, which encodes the N-terminal hydrophilic part of the viral coat protein L1R, was constructed to transform tobacco plastids. It provides integration of the transgene into the *trnG*/*trnfM* region of the chloroplast tobacco genome by homologous recombination. Spectinomycin-resistant tobacco lines were obtained by biolistic gun-mediated genetic transformation. PCR analysis in the presence of gene-specific primers confirmed integration of the transgene into the plant genome. Subsequent Northern and Western blot analysis showed the gene expression at the transcriptional and translational levels. The recombinant protein yields reached up to 0.9 % of total soluble protein. The transplastomic plants displayed a growth retardation and pale green leaf color compared to the wild type, but they developed normally and produced seeds. Southern blot analysis showed heteroplasmy of the plastids in the obtained plants due to recombination events between native and introduced regulatory plastid DNA elements. The recombinant protein from plant tissue was purified using metal affinity chromatography. Future research will be focused on determining the potential of the chloroplast-produced protein to induce neutralizing antibodies against *SPPV* strains.

Key words: *Sheeppox virus*; tobacco; *Nicotiana tabacum*; L1R protein; chloroplasts; transplastomic plants.

For citation: Beisenov D.K., Stanbekova G.E., Iskakov B.K. Transplastomic tobacco plants producing the hydrophilic domain of the sheep pox virus coat protein L1R. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(8):905-912. DOI 10.18699/VJ20.689

Транспластомные растения табака, продуцирующие гидрофильный домен белка оболочки L1R вируса оспы овец

Д.К. Бейсенов ✉, Г.Э. Станбекова, Б.К. Искаков

Институт молекулярной биологии и биохимии им. М.А. Айтхожина, Алматы, Казахстан
✉ e-mail: beisenov.d@gmail.com

Аннотация. Оспа овец имеет широкий географический ареал и представляет угрозу овцеводству во всем мире, так как заболевание высококонтагиозное и сопровождается большими экономическими потерями. В настоящее время для профилактики этого заболевания применяются вакцины на основе живых аттенуированных штаммов вируса. Подобные вакцины эффективны, однако потенциально опасны из-за возможной реверсии вируса к патогенному состоянию, поэтому весьма актуально создание безопасных рекомбинантных субъединичных вакцин против оспы овец. Известно, что хлоропласты растений в силу своей полиплоидности могут нарабатывать чужеродные белки в больших количествах. Целью данного исследования было получение трансластомных растений табака *Nicotiana tabacum*, синтезирующих один из кандидатных вакцинных белков вируса оспы овец L1R. Для проведения генетической трансформации хлоропластов создана конструкция, обеспечивающая интеграцию делеционного варианта гена *SPPV_56*, кодирующего N-концевую гидрофильную часть оболочечного белка L1R, в межгенную область *trnG*/*trnfM* хлоропластного генома табака путем гомологичной рекомбинации. Методом биобаллистики с помощью «генной пушки» получены линии табака, устойчивые к селективному антибиотику спектиномицину. ПЦР-анализ в присутствии ген-специфичных праймеров подтвердил интеграцию целевой вставки в растительный геном. Последующие нозерн- и вестерн-блот анализы препаратов РНК и белковых экстрактов из полученных растений показали экспрессию целевого гена на транскрипционном и трансляционном уровне. Содержание рекомбинантного белка составило ~0.9 %

от общего растворимого белка. Несмотря на задержку роста и более бледную окраску листьев по сравнению с растениями дикого типа, трансластомные растения нормально развивались и завязывали семена. Оценка гомопластидности методом Саузерна выявила гетерогенность пластидных геномов полученных растений, обусловленную генетической рекомбинацией между эндогенными и привнесенными в составе конструкции хлоропластными регуляторными ДНК-последовательностями. Методом металл-аффинной хроматографии была проведена очистка рекомбинантного белка из растительной ткани. В дальнейшем планируется изучить способность продуцируемого хлоропластами белка индуцировать вируснейтрализующие антитела против штаммов вируса оспы овец.
Ключевые слова: вирус оспы овец; *Sheeppox virus*; табак; *Nicotiana tabacum*; белок L1R; хлоропласты; трансластомные растения.

Introduction

Sheeppox virus (SPPV) belongs to the *Capripoxvirus* genus, a member of the Poxviridae family (Tulman et al., 2002). The highly contagious sheeppox disease causes significant economic damage to sheep breeding farms, due to high sheep mortality rate, especially among young animals. It also decrease the productivity of meat and wool, and increase the cost of veterinary and sanitary measures. The geographic distribution of sheep pox is very extensive. The disease is endemic in the Middle East, Central and South Asia, China, Central and North Africa. Outbreaks of sheep pox are regularly recorded in the CIS countries, including Russia and Kazakhstan.

Currently available live attenuated vaccines based on "NISKHI" strain are widely used for the specific prophylaxis of sheep pox in Russia and the CIS countries (Kurchenko et al., 1991). Live attenuated vaccines are potentially dangerous and theoretically capable of recombining to form virulent strains. As an alternative, recombinant vaccines containing highly immunogenic coat proteins are effective and safer immunizing drugs. Hepatitis B vaccine produced by incorporating the surface antigen of the virus into the genome of yeast cells is an example (McAleer et al., 1984).

Bacterial, yeast, animal, plant and other systems are currently used for the production of target recombinant proteins. The main disadvantage of prokaryotic systems is the absence of post-translational modification of proteins, while yeasts are characterized by excessive glycosylation of proteins, which is different from mammalian cells. Animal systems for the expression of recombinant proteins are extremely expensive and allow to obtain only small amounts of a pure product (Demain, Vaishnav, 2009).

Plant-based recombinant vaccine is an attractive alternative to other systems due to their low cost and absence of human and animal pathogens. Transgenic plants obtained by nuclear transformation usually produce low level of foreign protein (Shchelkunov et al., 2011). Multiplication of transgene copies in order to increase its expression level often leads to post-transcriptional silencing (Finnegan, McElroy, 1994).

Expression of transgenes in plastids has several important advantages over nuclear expression. The content of the recombinant protein in chloroplasts is several orders of magnitude higher compared to nuclear expression and possible to reach more than 70 % of the total soluble protein (TSP) (Oey et al., 2009). Other advantages are the absence of transgene silencing, protection of the target recombinant proteins in plastids from cellular proteases, integration of the transgene into the same intergenic region of plastid DNA, and the possibility of simultaneous expression of several transgenes by combining them into one operon. In contrast to nuclear transformation, transplastomic plants are safer for agrocenoses, since chloroplasts are not contained in pollen and the transfer of transgenes to closely related plant species is unlikely (Clarke, Daniell, 2011).

Works on Dengue viruses, poliomyelitis, the causative agent of tuberculosis, and smallpox vaccine are examples of successful production of antigens in transplastomic plants (Rigano et al., 2009; Daniell et al., 2019; Saba et al., 2019; van Eerde et al., 2019).

We used *SPPV_56* gene in this study, as it encodes an ortholog of the well-studied L1R protein of the vaccinia virus, which was used as a live smallpox vaccine in the 20th century. The L1R is a membrane protein of the infectious intracellular mature virion (IMV) and is required for the virus to enter the cell (Bisht et al., 2008). Bacterially synthesized shortened form of the L1R protein induced the production of virus-neutralizing antibodies to SPPV in immunized laboratory animals (Chervyakova et al., 2016) and gave us the reason to consider it as a candidate subunit vaccine.

The aim of this work is to obtain transplastomic tobacco plants expressing a deletion variant of the *SPPV_56* gene encoding a protein domain exposed on the outer side of the virion membrane, which we have chosen as a candidate subunit vaccine and will be designated hereinafter as shL1RΔ.

Research methods

Development of chloroplast transformation plasmids. We have cloned a 567 bp fragment of the *SPPV_56* gene (GenBank ID: NP_659632), encoding the N-terminal hydrophilic part of the shL1R protein, into the pET19b/SPPV_56Δ expression vector (Beisenov et al., 2014). Then the deletion variant of the SPPV gene was sequentially transferred into the pICH11599, pHK20 and pNT4 plasmids, kindly provided by prof. H. Warzecha (Germany). The resulting vector was designated pNT4/shL1RΔ. Thermo produced all enzymes used in this work. The target gene was under the transcriptional control of chloroplast elements: the P*rrn* promoter of the ribosomal operon and the T*rbcL* terminator of the gene for the large subunit of ribulose bisphosphate carboxylase. The vector includes the *aadA* marker gene encoding aminoglycoside adenylyltransferase, which confers antibiotic resistance to spectinomycin and streptomycin and allows selection of transformants. The *aadA* gene is located between the P*rrn* promoter and the plastid *psbA* gene terminator. The flanking sequences ensure the integration of the transgene into the *trnG*/*trnfM* intergenic region of the chloroplast genome by homologous recombination.

DNA sequencing was performed using a commercial Big Dye® Terminator v. 3.1 kit (Applied Biosystems) according to the manufacturer's protocol. Gene-specific 56-for (5′-gcat catatgggagcagccgctagtat) and 56-rev (5′-gcatgtcgacttatatata aaattgatatatccgtatcccga) primers were used for reading in both directions. DNA samples were analyzed on an ABI Prism 310 genetic analyzer (Applied Biosystems).

Chloroplast transformation. We used *Nicotiana tabacum* (cv. Petit Havana) leaves for transformation. Tobacco plants were grown under aseptic conditions *in vitro* on MS medium (Murashige, Skoog, 1962) containing 3 % sucrose and 0.7 % agar. Biolistic was performed using a PDS-1000/He Biolistic Particle Delivery System (Bio-Rad) gene gun according to a generally accepted protocol (Svab et al., 1990). Leaf explants were placed for regeneration on Petri dishes with MS medium containing 1 μg/ml BAP (6-benzylaminopurine), 0.1 μg/ml NAA (naphthylacetic acid), and 500 μg/ml of the spectinomycin (Sm). The dishes were incubated at 23 °C, with an illumination of 3000 lux and a light regime of 16/8 hours (day/night). Leaf segments were transferred to Petri dishes with fresh medium every two weeks. The shoots regenerated for several months were cut and rooted on MS medium without hormones and with Sm antibiotic.

DNA isolation. Total DNA were isolated from 100 mg of tobacco leaves using the DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's protocol.

Polymerase chain reaction (PCR) for the detection of the transgene in plants was carried out in the presence of Taq DNA polymerase (Thermo) and a pair of 56-for/56-rev primers. A pair of trnH-for/aadA-rev primers (5′-cacaatccactgccttgatcc; 5′-agaagaagatcgcttggcctc) were used to detect the recombination variant. 50 ng of total plant DNA was used as a template. The reaction was carried out in the following temperature regime: stage 1 – 3 min at 94 °C (1 cycle); stage 2 – 30 sec at 94 °C, 30 sec at 54 °C, 1 min at 72 °C (30 cycles); stage 3 – 5 min at 72 °C (1 cycle).

Western blotting. Protein preparations from plant leaves were isolated using the Trizol reagent (Sigma) according to the manufacturer's recommendations. Protein concentration was measured relative to known concentrations of bovine serum albumin (BSA) by M. Bradford method (Bradford, 1976).

We used 15 μg of each sample for the electrophoretic separation of plant proteins in a 12 % SDS-PAA gel according to the generally accepted method (Laemmli, 1970). Proteins were transferred to a PVDF membrane (Bio-Rad) after electrophoresis by semi-dry electroblotting in transfer buffer (102 mM glycine, 25 mM Tris-HCl, 20 % (v/v) ethanol) at a 0.8 mA/cm^2 current for 1 hour. The membrane was blocked in 5 % non-fat milk (Sigma) prepared in TBS buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl) for 1 hour. Polyclonal rabbit antibodies specific to the shL1RΔ protein (kindly provided by the Research Institute for Biological Safety Problems, Kazakhstan) or mouse antibodies to pentahistidine (5 PRIME) diluted in blocking buffer in 1:4000 ratio were used as primary antibodies to detect the protein. We used anti-mouse or anti-rabbit IgG conjugated with horseradish peroxidase (Santa Crus Biotechnology) and diluted in blocking buffer at 1:4000 ratio as secondary antibodies. Incubation with antibodies was carried out for 1 hour at room temperature. Antibodies were washed off four times for 5 min with TBST buffer (TBS containing 0.05 % Tween-20). Chemiluminescent Peroxidase Substrate-3 (Sigma) reagent was used as a substrate. The membranes were exposed on X-ray film (USA Scientific). Protein content in tobacco lines was determined densitometrically using the Image J 1.42 (NIH) program relative to known concentrations of purified shL1RΔ protein synthesized in bacteria. The size of the recombinant proteins was calculated using the GelAnalyzer 19.1 software (www.gelanalyzer.com).

Southern blotting. Total DNA isolated from transplastomic lines and wild-type plants was treated with *Eco*O109I restriction enzyme (Thermo), selected as a result of computer analysis of the nucleotide sequence of tobacco chloroplast DNA (GenBank ID: Z00044) using the SnapGene program (www.snapgene.com). The DNA fragments were transferred onto a positively charged nylon membrane (Macherey-Nagel) after electrophoresis in 0.8 % agarose gel. Hybridization was carried out at 42 °C overnight. The probe was a DIG-labeled PCR product obtained during the amplification of wild-type DNA using a pair of chl-dir/chl-rev primers (5′-cgacggaga gggggtccacc; 5′-gaagccccctttaccattctgtat). Probe labeling and detection of bound DNA fragments were performed using PCR-DIG Probe Synthesis Kit (Roche) and DIG Luminescent Detection Kit (Roche). The membranes were exposed on X-ray film (USA Scientific).

Northern blotting. RNA were isolated using the Trizol reagent (Sigma). 5 μg of RNA was separated by electrophoresis in a formaldehyde-containing 1.2 % agarose gel, transfered to a nylon membrane (Macherey-Nagel) and incubated with a probe at 50 °C overnight. A DIG-labeled PCR product obtained with the participation of the pNT4/shL1RΔ plasmid and the 56-for/56-rev primer pair was used as a probe. Probe labeling and detection of hybridization products were performed with the same reagents used in Southern blotting.

Isolation of recombinant protein from leaves. 1 g of leaves was ground with a pestle in a mortar with 8 ml of lysis buffer (50 mM NaH_2PO_4, 0.3 M NaCl, 2 mM imidazole, 1 % Triton X-100, 15 mM β-mercaptoethanol, 2 mM PMSF, pH 8.0). The lysate was centrifuged at 10,000 g for 20 min. 1 ml of Ni-NTA agarose suspension (5 PRIME) was added to the supernatant and the mixture was incubated with shaking for 1 hour on ice. Afterwards, the resin was washed twice with 8 ml of wash buffer (50 mM NaH_2PO_4, 0.3 M NaCl, 20 mM imidazole). The protein was washed out in seven steps using 7 ml of elution buffer (50 mM NaH_2PO_4, 0.3 M NaCl, 250 mM imidazole, pH 8.0). The fractions were pooled, dialyzed against potassium phosphate buffer (pH 7.0) and concentrated by ultrafiltration through 3,000 MWCO HY columns (Amicon).

Results

Genetic construct design and transformation of tobacco chloroplasts

We cloned the deletion variant of the *SPPV_56* gene (includes the first 567 bp out of 738 bp) into the pNT4 chloroplast vector in three steps to obtain a vector intended for the transformation of chloroplasts. At the first stage, the SPPV gene from the pET19b/SPPV_56Δ plasmid was transferred using *Nco*I/*BamH*I restriction sites into the pICH11599 vector for transient expression, that resulted in pICH11599/SPPV_56Δ plasmid

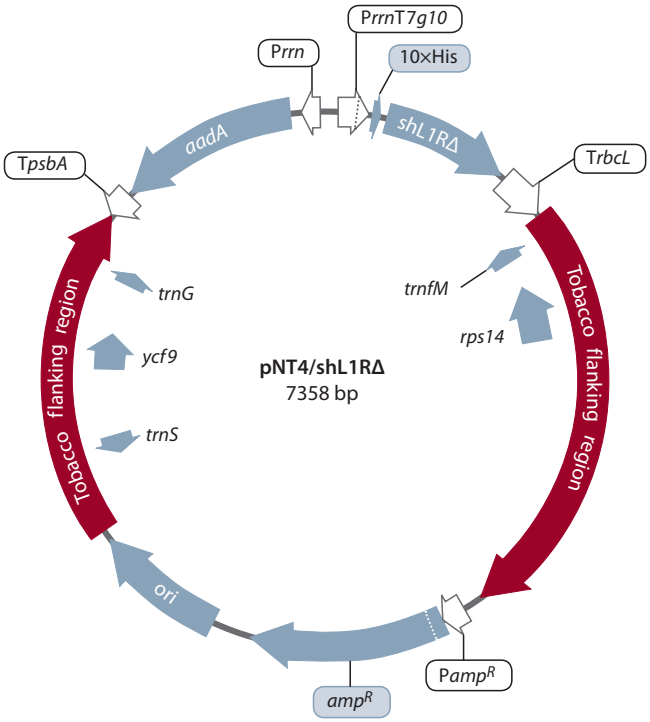


Fig. 1. Map of the pNT4/shL1RΔ plasmid.

trnG – glycine tRNA gene; T*psbA* – terminator of the plastid *psbA* gene; *aadA* – spectinomycin resistance gene; P*rrn* – plastid ribosomal operon promoter; P*rrn*T7*g10* – P*rrn* fused to the 5′-untranslated region from gene 10 of phage T7; 10×His – a sequence encoding 10 histidine residues; shL1RΔ is a deletion variant of the *SPPV_56* gene; T*rbcL* – terminator of the plastid *rbcL* gene; *trnfM* – formyl methionine tRNA gene; amp^R and Pamp^R – ampicillin resistance gene and its promoter, respectively; ori – replication start site; *ycf9* – gene coding protein Z of photosystem II reaction center; *trnS* – serine tRNA gene; *rps14* – ribosomal protein S14 gene.

construction. The presence of the *Xba*I site in the resulting plasmid made it possible to carry out the subsequent cloning of the SPPV gene digested with *Nco*I/*Xba*I restriction endonucleases into the intermediate pHK20 vector at the *Nde*I/*Xba*I sites. The insert and the vector were treated with *Nco*I and *Nde*I restriction enzymes. Protruding 5′-ends of the DNA were completed with the Klenow fragment in the presence of dNTPs. At the third stage, the *SPPV_56*Δ gene, the P*rrn* chloroplast promoter and the T*rbcL* terminator were transferred at the *Sac*I/*Hind*III restriction sites into the pNT4 chloroplast vector. Figure 1 shows a schematic representation of the pNT4/shL1RΔ vector, intended for chloroplasts transformation. Nucleotide sequencing showed absence of any mutations in the *SPPV_56*Δ gene after cloning and the resulting construct can be used in further work.

The process of obtaining transplastomic plants by the biolistic method consisted of several stages: bombardment of whole tobacco leaves with the pNT4/shL1RΔ plasmid immobilized on the surface of gold particles, cultivating them on a medium with hormones and Sm, obtaining calli, selecting regenerating shoots that are resistant to Sm. Figure 2 shows the stages of obtaining transplastomic tobacco plants. As a result, three Sm-resistant shoots were selected from separate segments of eight shot leaves.

Molecular and genetic analysis of the resulting plants

Plants were screened for the presence of the target gene using the PCR method. Total DNA preparations were isolated from the leaves of the tested plants and wild-type plant, which were analyzed using a pair of gene-specific 56-for/56-rev primers. The test showed that all three selected lines contained the *SPPV_56*Δ gene of the expected size (567 bp) (Fig. 3, *a*). The ability of the resulting lines to express the target gene was studied at the transcriptional and translational levels. Northern blotting of total cellular RNA preparations using a DIG-labeled probe to the *SPPV_56*Δ gene revealed the presence of two types of recombinant mRNA in all obtained lines (see Fig. 3, *b*). Along with the monocistronic transcript, a longer product was found, apparently due to ineffective transcription termination, which is generally typical for plastids (Zhou et al., 2007; Oey et al., 2009).

The ability of transplastomic lines to produce the recombinant shL1RΔ protein was assessed by immunoblotting using antibodies to bacterially synthesized shL1RΔ. The recombinant protein in plant extracts corresponds to the theoretically expected size of 23 kDa (see Fig. 3, *c*). We also identified a 46 kDa protein, which, presumably, is a dimeric form of shL1RΔ. Comparative densitometric analysis of protein bands relative to known amounts of purified bacterially synthesized shL1RΔ in three independent experiments showed that the level of recombinant protein in plants reaches ~0.9 % of the total soluble protein.

Assessment of the homoplasticity of transplastomic plants

To obtain homoplastic plants, each line was subjected to further selection in order to eliminate wild-type and select transformed plastids. Leaf segments of the lines were cultivated on a medium with hormones and antibiotic until the appearance of secondary regenerants. This procedure was carried out four times. Then the plants of the T_0 generation were planted in the ground for further analysis. Transplastomic lines transplanted into soil, showed signs of growth retardation and paler leaf color in comparison with the wild type plants (see Fig. 2, *e*). Despite this circumstance, all lines were fertile and formed viable seeds upon self-pollination.

The homoplasticity of the obtained lines was assessed by the restriction fragment length polymorphism using Southern blotting with labeled probe that covered the site of transgene insertion into plastid DNA between the *trnG* and *trnfM* genes with adjacent regions (Fig. 4, *a*). Analysis showed, the probe bound to one 3.2 kb DNA fragment of the expected length from wild-type plants (see Fig. 4, *c*). In transplastomic lines, in addition to the expected 5.3 kb fragment we revealed several additional fragments (marked in Fig. 4, *c* with asterisks). Possibly, plastids heterogeneity revealed in lines is caused by intermolecular post-transformation recombination between endogenous and plastid regulatory elements introduced into the structure: promoters, terminators, 5′-untranslated sequences. Such cases are described in a number of works (McCabe et al., 2008; Zhou et al., 2008; Gray et al., 2009). One of the recombination variants identified in this work (marked in Fig. 4, *c* with two asterisks and schematically shown in Fig. 4, *d*) serves as evidence of the rearrangements that have taken place. The fact of recombination between the natural and introduced plastid T*psbA* terminators was confirmed by

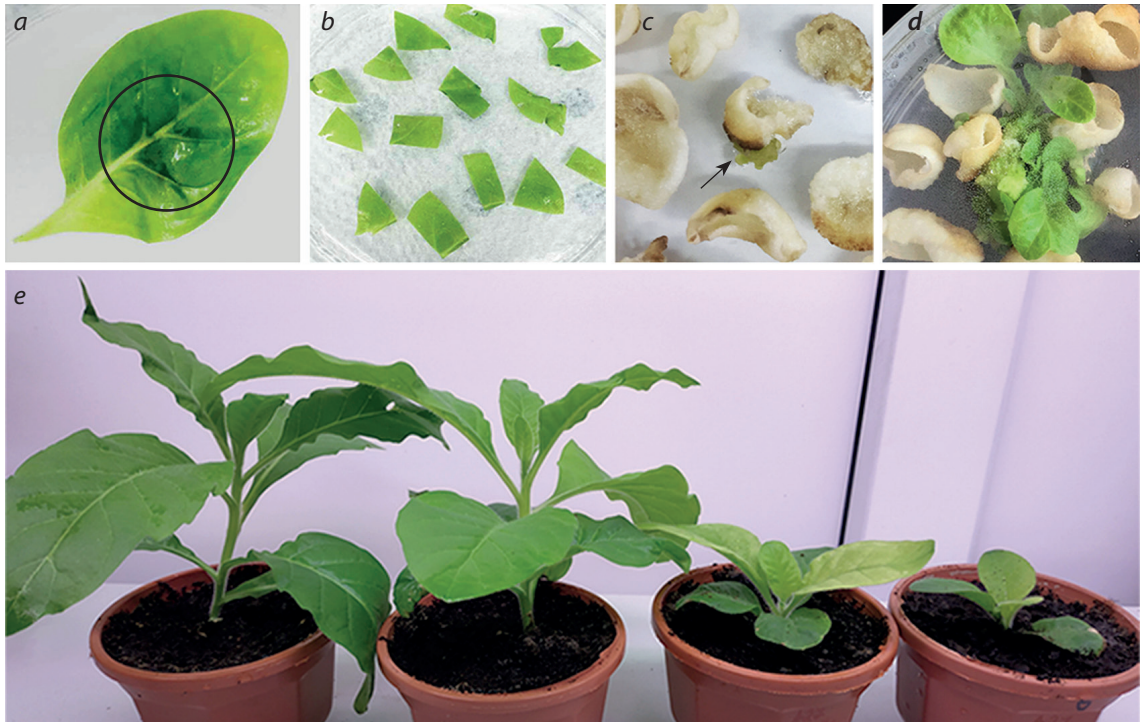


Fig. 2. Stages of creating transplastomic tobacco plants.

a – a leaf immediately after bombardment (microparticles penetration area is marked); *b* – leaf segments on the medium for regeneration; *c* – callus formation (marked by arrow); *d* – regeneration; *e* – plants planted in the soil (two wild-type plants on the left, two transplastomic lines on the right).

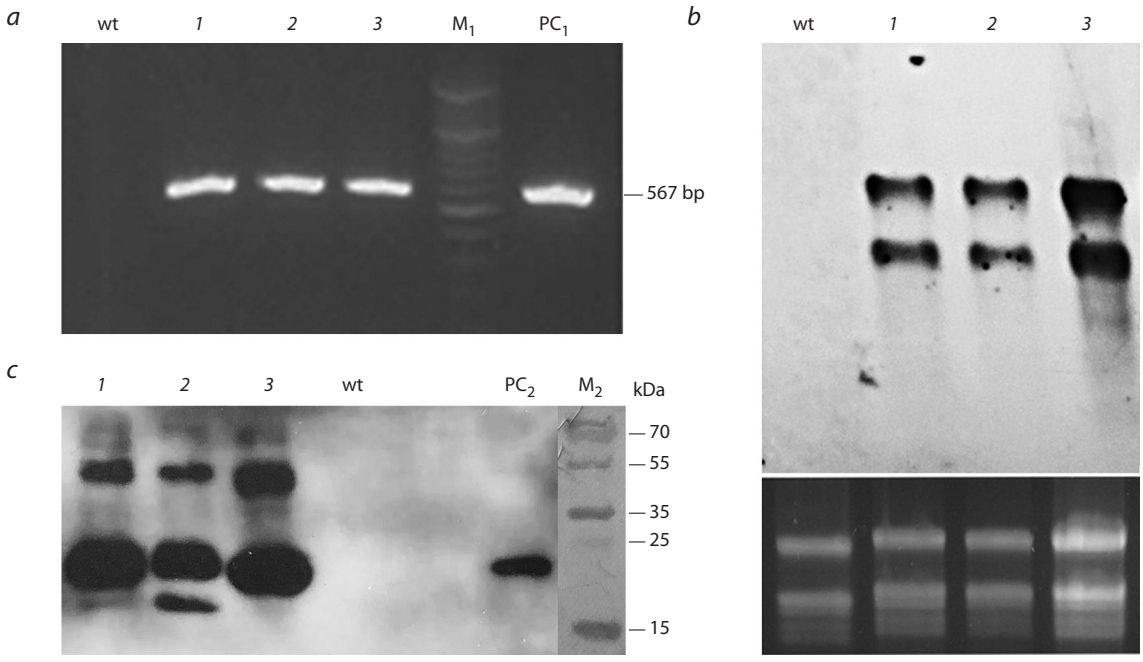


Fig. 3. Molecular and genetic analysis of transplastomic plants.

a – PCR analysis of total DNA from various lines with *SPPV_56Δ* gene specific primers; *b* – Northern blotting of total RNA with *SPPV_56Δ* probe (lower gel stained with ethidium bromide reflects the amount of analyzed RNA); *c* – Western blotting of protein extracts with antibodies against shL1RΔ.

wt – negative control (DNA from wild-type tobacco plant); *1–3* – analyzed plant lines, M_1 – DNA marker Gene Ruler 100 bp (Thermo); PC_1 – 10 ng pNT4/shL1RΔ; M_2 – protein marker PageRuler Plus (Thermo); PC_2 – 20 ng of shL1RΔ protein purified from bacteria.

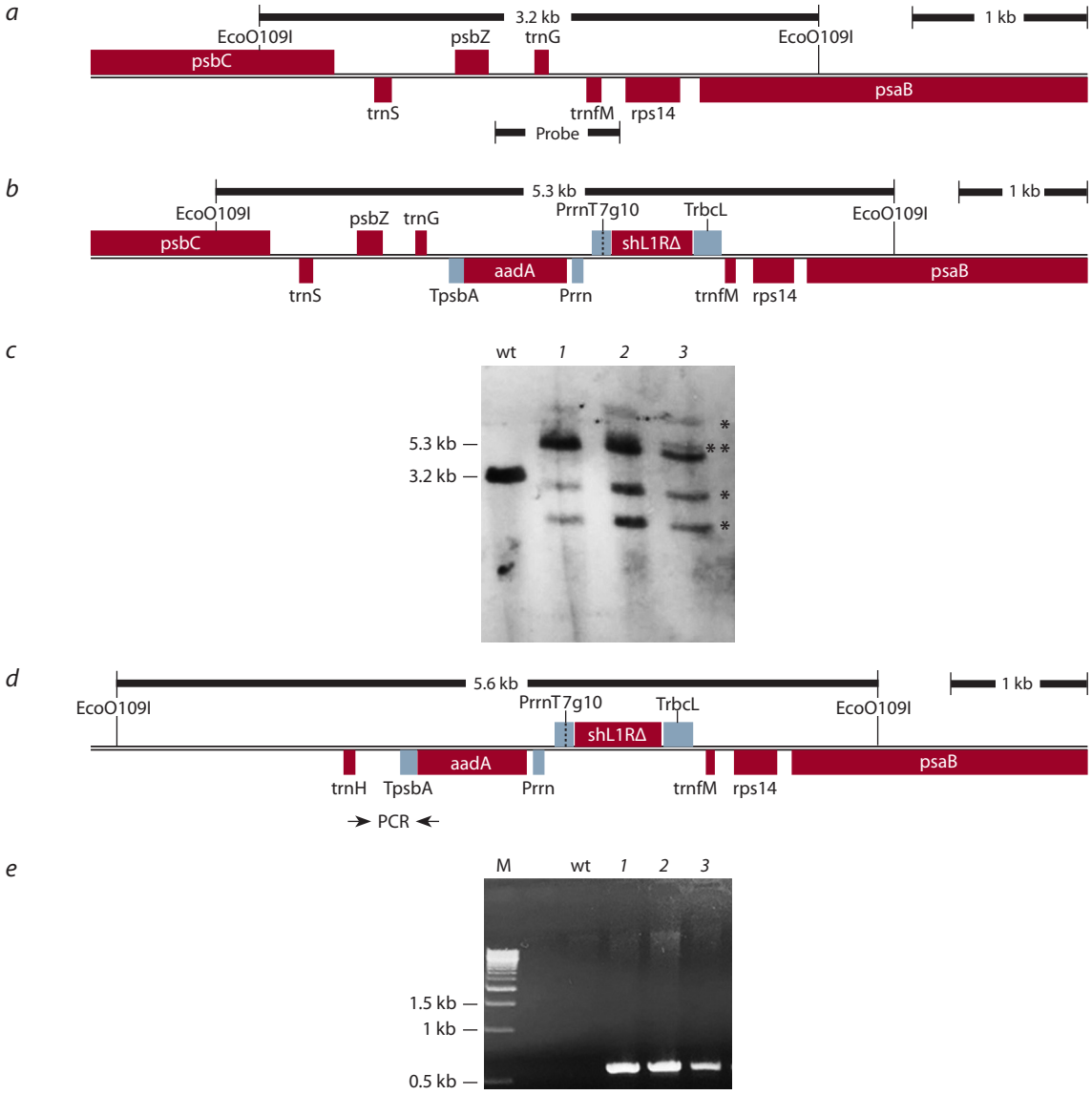


Fig. 4. Homoplasmy assessment of transplastomic plants.

a – the region of transgene integration into the plastid genome of wild-type plants; *b* – the same region in the transplastomic lines; *c* – Southern blot analysis with probe to the *SPPV_56Δ* gene; *d* – a recombination variant; *e* – PCR analysis of lines with a pair of trnH-for/aadA-rev primers.
wt – a wild-type plant; M – DNA marker Gene Ruler 1 kb (Thermo); *1–3* – analyzed lines.

the presence of the 594 bp DNA fragment of the expected size amplified during PCR analysis with the trnH-for/aadA-rev primers (see Fig. 4, *e*). We did not study the rest of the recombination variants.

Purification of recombinant protein from plant material

The presence of decahistidine at the N-terminus of the recombinant shL1RΔ protein facilitates its further purification by metal affinity chromatography on Ni-NTA agarose. Preliminary experiments showed that the addition of the non-ionic detergent Triton X-100 to the extraction buffer provided a higher yield of the target protein as compared to Tween-20 and SDS. Figure 5 shows the results of immunodetection of the shL1RΔ protein in purified fractions. Protein-containing fractions were pooled, then dialyzed against potassium phosphate buffer and concentrated by ultrafiltration. The yield of the recombinant protein purified from the leaves was 10.3 μg/g.

Discussion

The development of genetic engineering and biotechnology over the past decades has opened up wide opportunities for obtaining a new generation of vaccines based on highly immunogenic surface antigens of human and animal pathogens. The plastid and the transient plant expression systems are cheap source of recombinant proteins for medical and veterinary purposes.

In this work, we described the production of transplastomic plants producing one of the candidate vaccine proteins, namely the truncated form of the structural L1R protein of the sheeppox virus. Earlier, we obtained transgenic rapeseed

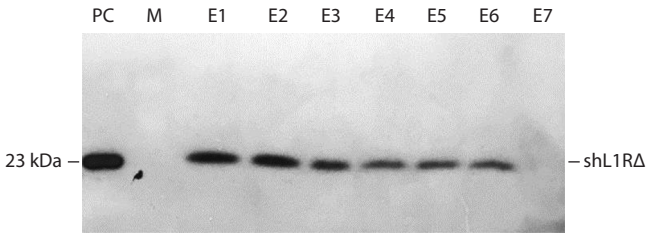


Fig. 5. Western blot analysis of fractions eluted from a Ni^{2+}-NTA agarose.

PC – 40 ng of shL1RΔ protein produced in bacteria; M – protein marker PageRuler Plus (Thermo); E1– E7 – eluted fractions.

plants with the nuclear localization of the same gene (Beisenov et al., 2019). The content of recombinant viral protein in rape plants was about 0.1 % of the TSP. In this study, we managed to significantly increase the expression of the target gene by transferring it to the chloroplast genome. The content of the recombinant protein was about 0.9 % of the TSP. Potentially, the content of the recombinant protein may be increased by inserting an artificially synthesized gene with a codon optimized for expression in chloroplasts, as demonstrated for antigens of the human papillomavirus (Lenzi et al., 2008; Daniell et al., 2019).

Alternative approaches to increasing the protein content are to increase the copy number of the gene by integration into the inverted repeat region of the plastid genome, and the addition of certain N-terminal peptides in the case of unstable recombinant proteins (Bock, 2014). The problem of heterogeneity of the plants obtained in this work can be solved by reorganizing the genetic structure intended for transformation. F. Zhou et al. changed the orientation of the target gene relative to the *aadA* gene. As a result, increased distance between the two P*rrn* promoters made possible to obtain stable homoplastic plants producing antigens of the human immunodeficiency virus at a 40 % of the TSP (Zhou et al., 2008). Nevertheless, the achieved level of synthesis allows us to isolate a sufficient amount of protein required for further immunological studies. We intend to study the ability of the recombinant viral protein shL1RΔ purified from plants to induce the production of virus neutralizing antibodies in laboratory animals.

Conclusion

As a result of our studies, we have shown the ability to synthesize the shortened structural protein shL1R of sheeppox virus in transplastomic tobacco plants. The recombinant protein can then be used to develop a subunit vaccine against sheep pox.

References

Beisenov D.K., Argimbaeva T.U., Stanbekova G.E., Iskakov B.K. Synthesis of the immunogenic domain of the L1R protein of sheep pox in rapeseed. *Veterinariya*, *Zootekhniya i Biotekhnologiya = Veterinary*, *Zootechnics and Biotechnology.* 2019;8:45-54. (in Russian)

Beisenov D., Stanbekova G., Nadirova L., Iskakov B. Sheep pox viral envelope protein L1RΔ synthesis in plants. *Vestnik KazNU. Seriya Biologicheskaya = KazNU Bulletin. Biology series.* 2014;60: 187-190. (in Russian)

Bisht H., Weisberg A.S., Moss B. Vaccinia virus L1 protein is required for cell entry and membrane fusion. *J. Virol.* 2008;82:8687-8694. DOI 10.1128/JVI.00852-08.

Bock R. Engineering chloroplasts for high-level foreign protein expression. *Methods Mol. Biol.* 2014;1132:93-106. DOI 10.1007/978-1-62703-995-6_5.

Bradford M. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 1976;72:248-254.

Chervyakova O.V., Zaitsev V.L., Iskakov B.K., Tailakova E.T., Strochkov V.M., Sultankulova K.T., Sandybayev N.T., Stanbekova G.E., Beisenov D.K., Abduraimov Y.O., Mambetaliyev M., Sansyzbay A.R., Kovalskaya N.Y., Nemchinov L.G., Hammond R.W. Recombinant sheep pox virus proteins elicit neutralizing antibodies. *Viruses.* 2016;8:159-171. DOI 10.3390/v8060159.

Clarke J.L., Daniell H. Plastid biotechnology for crop production: present status and future perspectives. *Plant Mol. Biol.* 2011;77:203. DOI 10.1007/s11103-011-9767-z.

Daniell H., Rai V., Xiao Y. Cold chain and virus-free oral polio booster vaccine made in lettuce chloroplasts confers protection against all three poliovirus serotypes. *Plant Biotechnol. J.* 2019;17:1357-1368. DOI 10.1111/pbi.13060.

Demain A.L., Vaishnav P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* 2009;27:297-306. DOI 10.1016/j.biotechadv.2009.01.008.

Finnegan J., McElroy D. Transgene inactivation: plants fight back! *Nat. Biotechnol.* 1994;12:883-887.

Gray B.N., Ahner B.A., Hanson M.R. Extensive homologous recombination between introduced and native regulatory plastid DNA elements in transplastomic plants. *Transgenic Res.* 2009;18:559-572. DOI 10.1007/s11248-009-9246-3.

Kurchenko F.P., Ivanyushchenkov V.N., Ufimtsev K.P. The effectiveness of dry culture vaccinia virus from the NISKHI strain against sheep pox. *Veterinariya = Veterinary Medicine.* 1991;10:21-24. (in Russian)

Laemmli U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 1970;227:680-685.

Lenzi P., Scotti N., Alagna F., Tornesello M.L., Pompa A., Vitale A., De Stradis A., Monti L., Grillo S., Buonaguro F.M., Maliga P., Cardi T. Translational fusion of chloroplast-expressed human papillomavirus type 16 L1 capsid protein enhances antigen accumulation in transplastomic tobacco. *Transgenic Res.* 2008;17:1091-1102. DOI 10.1007/s11248-008-9186-3.

McAleer W.J., Buynak E.B., Maigetter R.Z., Wampler D.E., Miller W.J., Hilleman M.R. Human hepatitis B vaccine from recombinant yeast. *Nature.* 1984;307:178-180. DOI 10.1038/307178a0.

McCabe M.S., Klaas M., Gonzalez-Rabade N., Poage M., Badillo-Corona J.A., Zhou F., Karcher D., Bock R., Gray J.C., Dix P.H. Plastid transformation of high-biomass tobacco variety Maryland Mammoth for production of human immunodeficiency virus type 1 (HIV-1) p24 antigen. *Plant Biotechnol. J.* 2008;6:914-929. DOI 10.1111/j.1467-7652.2008.00365.x.

Murashige T., Skoog F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* 1962;15:473-497.

Oey M., Lohse M., Kreikemeyer B., Bock R. Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. *Plant J.* 2009;57:436-445. DOI 10.1111/j.1365-313X.2008.03702.x.

Rigano M.M., Manna C., Giulini A., Pedrazzini E., Capobianchi M., Castilletti C., Di Caro A., Ippolito G., Beggio P., De Giuli Morghen C., Monti L., Vitale A., Cardi T. Transgenic chloroplasts are efficient sites for high-yield production of the vaccinia virus envelope protein A27L in plant cells. *Plant Biotechnol. J.* 2009;7:577-591. DOI 10.1111/j.1467-7652.2009.00425.x.

Saba K., Gottschamel J., Younus I., Syed T., Gull K., Lössl A.G., Mirza B., Waheed M.T. Chloroplast-based inducible expression of ESAT-6 antigen for development of a plant-based vaccine against tuberculosis. *J. Biotechnol.* 2019;305:1-10. DOI 10.1016/j.jbiotec.2019.08.016.

Shchelkunov S.N., Konstantinov Yu.M., Deineko E.V. Transplastome plants. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2011;15(4):808-817. (in Russian)

Svab Z., Hajdukiewicz P., Maliga P. Stable transformation of plastids in higher plants. *Proc. Natl. Acad. Sci. USA.* 1990;87:8526-8530. DOI 10.1073/pnas.87.21.8526.

Tulman E.R., Afonso C.L., Lu Z., Zsak L., Sur J.H., Sandybaev N.T., Kerembekova U.Z., Zaitsev V.L., Kutish G.F., Rock D.L. The genomes of sheeppox and goatpox viruses. *J. Virol.* 2002;76:6054-6061. DOI 10.1128/JVI.76.12.6054-6061.2002.

van Eerde A., Gottschamel J., Bock R., Hansen K.E.A., Munangándu H.M., Daniell H., Liu Clarke J. Production of tetravalent dengue virus envelope protein domain III based antigens in lettuce chloroplasts and immunologic analysis for future oral vaccine development. *Plant Biotechnol. J.* 2019;17:1408-1417. DOI 10.1111/pbi.13065.

Zhou F., Badillo-Corona J., Karcher D., Gonzalez-Rabade N., Piepenburg K., Borchers A.M., Maloney A.P., Kavanagh T.A., Gray J.C., Bock R. High-level expression of human immunodeficiency virus antigens from the tobacco and tomato plastid genomes. *Plant Biotechnol. J.* 2008;6:897-913. DOI 10.1111/j.1467-7652.2008.00356.x.

Zhou F., Karcher D., Bock R. Identification of a plastid intercistronic expression element (IEE) facilitating the expression of stable translatable monocistronic mRNAs from operons. *Plant J.* 2007;52:961-972. DOI 10.1111/j.1365-313X.2007.03261.x.

ORCID ID

D.K. Beisenov orcid.org/0000-0002-2116-7323
G.E. Stanbekova orcid.org/0000-0002-7819-6475
B.K. Iskakov orcid.org/0000-0002-5204-4377

Acknowledgements. The research was conducted under the scientific projects APO5132066 and APO5132532 with financial support from the Ministry of Education and Science of the Republic of Kazakhstan.
The authors are grateful to Dr Heribert Warzecha (Darmstadt Technical University, Germany) for providing the plasmids pICH11599, pHK20 and pNT4. We thank the Research Institute for Biological Safety Problems (Kazakhstan) for specific antibodies to shL1RΔ protein.

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

Received April 14, 2020. Revised September 24, 2020. Accepted October 26, 2020.

Вавиловский журнал генетики и селекции. 2020;24(8):913-921
DOI 10.18699/VJ20.690

Generation of donor organs in chimeric animals via blastocyst complementation

T.I. Babochkina ✉, L.A. Gerlinskaya, M.P. Moshkin

Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ e-mail: babochkinat@yahoo.com

Abstract. The lack of organs for transplantation is an important problem in medicine today. The growth of organs in chimeric animals may be the solution of this. The proposed technology is the interspecific blastocyst complementation method in combination with genomic editing for obtaining "free niches" and pluripotent stem cell production methods. The CRISPR/Cas9 method allows the so-called "free niches" to be obtained for blastocyst complementation. The technologies of producing induced pluripotent stem cells give us the opportunity to obtain human donor cells capable of populating a "free niche". Taken together, these technologies allow interspecific blastocyst complementation between humans and other animals, which makes it possible in the future to grow human organs for transplantations inside chimeric animals. However, in practice, in order to achieve successful interspecific blastocyst complementation, it is necessary to solve a number of problems: to improve methods for producing "chimeric competent" cells, to overcome specific interspecific barriers, to select compatible cell developmental stages for injection and the corresponding developmental stage of the host embryo, to prevent apoptosis of donor cells and to achieve effective proliferation of the human donor cells in the host animal. Also, it is very important to analyze the ethical aspects related to developing technologies of chimeric organisms with the participation of human cells. Today, many researchers are trying to solve these problems and also to establish new approaches in the creation of interspecific chimeric organisms in order to grow human organs for transplantation. In the present review we described the historical stages of the development of the blastocyst complementation method, examined in detail the technologies that underlie modern blastocyst complementation, and analyzed current progress that gives us the possibility to grow human organs in chimeric animals. We also considered the barriers and issues preventing the successful implementation of interspecific blastocyst complementation in practice, and discussed the further development of this method.
Key words: chimerism; interspecies chimera; embryo SC; iPSC; CRISPR/Cas9; organ generation.

For citation: Babochkina T.I., Gerlinskaya L.A., Moshkin M.P. Generation of donor organs in chimeric animals via blastocyst complementation. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020; 24(8):913-921. DOI 10.18699/VJ20.690

Использование метода бластоцистной комплементации для получения донорских органов в химерных животных

Т.И. Бабочкина ✉, Л.А. Герлинская, М.П. Мошкин

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ e-mail: babochkinat@yahoo.com

Аннотация. Сегодня актуальной проблемой в медицине является нехватка органов для трансплантаций. Одна из предполагаемых технологий получения этих органов – выращивание их из клеток человека в организме химерных животных с использованием метода межвидовой бластоцистной комплементации в комбинации с методами геномного редактирования и получения плюрипотентных стволовых клеток. Метод CRISPR/Cas9 позволяет создавать животных для бластоцистной комплементации с так называемыми свободными нишами. Совершенствование методов получения индуцированных плюрипотентных стволовых клеток дает возможность получать донорские клетки человека, способные заселять свободную нишу. Таким образом, с помощью современных технологий можно осуществить межвидовую бластоцистную комплементацию между человеком и другими животными, что в будущем позволит выращивать органы человека внутри химерных животных. Однако на практике для проведения успешной межвидовой бластоцистной комплементации необходимо решить ряд проблем: усовершенствовать методы получения «химер-компетентных клеток», преодолеть специфические межвидовые барьеры, подобрать совместимые стадии развития клеток для инъекции и соответствующего этапа развития эмбриона-реципиента, предотвратить

апоптоз донорских клеток, добиться эффективной колонизации донорскими клетками человека организма животного-реципиента. Кроме того, очень важно проанализировать и законодательно урегулировать этические аспекты, возникающие при разработке технологий, связанных с получением химерных организмов с участием клеток человека. Многочисленные исследования направлены на решение этих проблем, а также на поиски новых подходов в создании межвидовых химерных организмов с целью выращивания органов человека для трансплантаций. В настоящем обзоре описаны исторические этапы развития технологии бластоцистной комплементации, детально разобраны методы, лежащие в основе ее современного варианта, и проанализированы достижения, позволяющие приблизиться к возможности выращивания органов человека в химерных животных. Рассмотрены также барьеры и проблемы, мешающие успешному применению данного подхода на практике, и дальнейшие перспективы его развития.
Ключевые слова: химеризм; межвидовой химеризм; ЭС клетки; ИПСК; CRISPR/Cas9; органы для трансплантаций.

Chimerism: definitions and classifications

The first studies on generating chimeric animals were carried out in the 60s of the last century (Tarkowski, 1961; Mintz, 1965; McLaren, Bowman, 1969). Since then, a significant amount of scientific knowledge on chimerism has been accumulated, modern definitions have been formulated and various classifications of chimeras have been proposed.

Chimeric animals are composed of genetically different cells originating from two or more different zygotes (Tippett, 1983). There are different classifications of chimerism, depending on the number and type of donor cells and their distribution in chimeric organisms. Chimerism can be natural or artificial. Natural chimerism is represented by two forms: tetragametism and microchimerism. Tetragametism results from the fertilization of two separate eggs by two different spermatozoa, followed by the development of a single organism with mixed cell lines (Drexler et al., 2005). Microchimerism is a phenomenon that occurs when a small number of cells from another individual are present in a multicellular organism. Examples of natural microchimerism are twin chimerism (Chen K. et al., 2013) and feto-maternal microchimerism (Nelson et al., 1998). The artificial chimerism occurs for example as a result of organ or tissue transplantation or blood transfusions.

The chimerism can be partial or systemic depending on the degree of donor cells distribution in a chimeric organism (Suchy, Nakauchi, 2017). For example, during organ or tissue transplantation the distribution of donor cells is limited to a particular organ or tissue which results in partial chimerism. Systemic chimerism can be observed, for example, during the fusion of embryos at an early stage of development. As a result of such a fusion, an embryo with cell lines which distributed over different organs and tissues is formed, with these lines originating from two different zygotes.

Chimerism can be primary and secondary. Primary chimerism occurs in the early stages of embryogenesis, and secondary chimerism occurs after the onset of gastrulation (Mascetti, Pedersen, 2016a, b). Chimerism can be intraspecies and interspecies. The intraspecies chimeras consist of cell lines originating from different zygotes of the same species. Interspecies chimeras consist of cell lines originating from two or more zygotes of representatives of different species.

The methods underlying the development of the blastocyst complementation

The most popular methods to obtain chimera under laboratory conditions are cell aggregation (Tarkowski, 1961) and microinjection into the embryo (Gardner, 1968). Aggregation methods for producing chimeras are technically easier, do not require expensive micromanipulation equipment, and sometimes can work more efficiently than injection methods (Tachibana et al., 2012). However, in some cases, for example, when obtaining interspecies chimeras, the trophectoderm with donor cells can impede implantation, and in this case injection methods are preferred (MacLaren et al., 1992). In addition, the injection methods allow to control the number of injected cells.

In their study Okumura and colleagues compared the degree of distribution of rat cells in chimeric rat-mouse embryos by different methods: the 8-cell aggregation method, injection into an 8-cell embryo, and injection into a blastocyst. According to the study, the degree of chimerism was highest when researchers used the injection method into an 8-cell embryo, although the percentage of chimeric mice was higher when they injected cells into the blastocyst (Okumura et al., 2019).

The most common and promising method to generate human organs for transplantation in the organisms of interspecies chimeric animals is injection into the blastocyst – so called the blastocyst complementation method. Further in this review this method is considered first in its application to the rodents, then the development of techniques related to this method is described: obtaining "free niches" of animals and obtaining "chimera-competent" human cells. These techniques made it possible to perform the interspecies blastocyst complementation between humans and other animals.

The early version of the blastocyst complementation for obtaining rodent chimeras

Intraspecies chimeras. In 1993 the method of intraspecies blastocyst complementation was successfully demonstrated for the first time. The main idea of the method was that wild type mouse embryonic stem (ES) cells were injected into the blastocyst derived from *Rag2*$^{-/-}$ immunodeficient mouse with T and B lymphocytes deficiency. As a result, donor T and B lymphocytes were observed in chimeric animals (Chen J. et al., 1993). An important result of this

study was that donor ES cells were able to differentiate into T and B lymphocytes, using the vacant lymphoid T and B cell niche in an immunodeficient organism. It demonstrated the possibility of generating organs in the body of chimeric animals with so-called "free niches". Then, in 2007, blastocyst complementation was used to grow pancreatic epithelium in *Pdx1*$^{-/-}$ deficient mice with impaired pancreas development (Stanger et al., 2007). In 2012, successful intraspecies blastocyst complementation of ES cells from a healthy mouse into the blastocyst of a *Sall1*$^{-/-}$ deficient mouse with impaired renal development was demonstrated (Usui et al., 2012).

Interspecies chimeras. In 2010, for the first time viable interspecies chimeras with a developed rat pancreatic epithelium were obtained in the body of a *Pdx1*$^{-/-}$ deficient mouse by blastocyst complementation (Kobayashi et al., 2010). In this study, scientists successfully injected rat pluripotent ES cells into murine *Pdx1*$^{-/-}$ blastocysts that were genetically modified to impair pancreas development. In 2011, interspecies blastocyst complementation was used to inject rat ES cells into the blastocyst of a nude mouse without a thymus, and a chimeric mouse with a functioning thymus of rat origin was obtained (Isotani et al., 2011). Recently, it was reported about the successful generation of a mouse kidney in the chimeric organism of *Sall1*$^{-/-}$ rat by interspecies blastocyst complementation (Goto et al., 2019).

In 2017, a Nakauchi group demonstrated the successful transplantation of pancreatic tissue generated from pluripotent stem cells in *Pdx1*$^{-/-}$ deficient rats to diabetic mice (Yamaguchi et al., 2017). These results proved the possibility of using tissues generated in the body of interspecies chimeric animals for organ transplantation.

Further in the review, the following technologies underlying modern blastocyst complementation are discussed in details: obtaining animals with so-called "free niches" and obtaining "chimera-competent" cells for injection into the blastocyst.

Generation of animals with "free niches"

The animals with "free niches" in organogenesis, that is with the absence or partial development of certain organs or special cell lines, are necessary for obtaining chimeric animals by the method of blastocyst complementation. Such animals with "free niches" in organogenesis are possible to obtain by turning off the expression of genes involved in organogenesis. Certain types of stem cells in these animals lose the ability to specialize, proliferate or differentiate, that is, they cannot participate in organogenesis and the organ does not develop.

When donor cells with normal organogenesis are injected into the blastocyst of animals with "free niches", missing organs can be formed. For the generation of donor organs in chimeric organisms, it is necessary that the donor's cells have an advantage in the organogenesis of a certain tissue or organ, since these cells are introduced in small numbers, and they do not initially have a selective advantage. The creation of "free niches" allows donor cells to proliferate without competition with host cells in a chimeric organism and to form a given organ. A "free niche" can be created by the gene knockout method (Offield et al., 1996; Ohinata et al., 2005) or by methods of genome editing: zinc finger nucleases (ZFN), TALE-associated nucleases (transcription activator-like effector nucleases, TALEN) and CRISPR/Cas9 (clustered, regularly interspaced, short palindromic repeats (CRISPR)/CRISPR-associated system (Cas)).

The obtaining of knockout mice by injecting messenger RNA (mRNA) nuclease into mouse zygotes (based on the ZFN method) was demonstrated in 2010 (Carbery et al., 2010). The knockout mice were obtained in 2013 using the TALEN technology by injecting TALEN mRNA into the cytoplasm of zygotes (Sung et al., 2013). The CRISPR/Cas9 method was also first demonstrated in 2013 (Cong et al., 2013; Mali et al., 2013); this method today is the most popular in genetic engineering. In this method, targeted genome editing is carried out due to the complementary interaction between the non-coding synthetic RNA and the DNA of the target sites. This forms a complex of non-coding RNAs and Cas proteins which have nuclease activity. The pigs with a mutation in the genes were obtained by using the CRISPR/Cas9 method in the cells of pig embryo at the blastocyst stage *in vitro* in 2014 (Whitworth et al., 2014). Successful intraspecific neural blastocyst complementation was performed for the first time in 2018 in mice with "free niches" in the brain, including those obtained using the CRISPR/Cas9 method (Chang et al., 2018).

Types of "chimera-competent" cells for injection into a blastocyst

In order to obtain chimeric animals, ES cells and induced pluripotent stem cells (iPSCs) are used.

ES cells. Pluripotent cells isolated from the inner cell mass (ICM) and epiblast of the embryo are the most suitable candidates to generate organs in chimeric animals, since they are able to differentiate into the embryonic tissue. It turned out that the pluripotency degree of pluripotent cells of ICM and epiblast are different in mice. It was proposed to call "true" pluripotent cells obtained from ICM at an early pluripotency stage "naïve", and epiblast cells obtained at a later pluripotency stage – "primed" (Nichols, Smith, 2009; Hanna et al., 2010). In addition, it turned out that pluripotent cells isolated at the same developmental stage in different species differ in the degree of pluripotency. For example, mouse ES cells isolated from the ICM relate to the "naïve" status, while similar human cells relate to the "primed" status of pluripotency.

Pluripotent ES cells in the "naïve" status, isolated from the ICM of the blastocyst before the implantation stage, are of the most interest for obtaining chimeric animals since it turned out that cells in the "primed" status are not able to take part in the formation of chimeras when they are injected into the preimplantation blastocyst (Tesar et al., 2007).

Mouse ES cells were first obtained in 1981 (Evans, Kaufman, 1981). Mouse ES cells exhibit typical characteristics of pluripotency: they have the ability to form cells of ecto-

dermal, mesodermal, and endodermal origin (Martin, 1981), and they are involved in the formation of all tissues of the adult organism, when injected into the blastocyst (Bradley et al., 1984; Hayashi et al., 2017). And most importantly, mouse ES cells are involved in the formation of chimeras after injection into the blastocyst (Nichols, Smith, 2009; Betschinger et al., 2013).

In 1995, ES cells of *Rhesus macaque* were first obtained (Thomson et al., 1995). The same researchers obtained human ES cell lines from preimplantation human embryos for the first time in 1998. The *in vivo* pluripotency test demonstrated the ability of human ES cells to form teratomas with tissues of endodermal, mesodermal, and exodermal origin (Thomson et al., 1998). The differentiation of ES cells with the formation of embryoid bodies and differentiation into various cell types was shown *in vitro* for human ES cells (Wobus, Boheler, 2005). It is impossible to test for chimerism and to perform an accurate assessment of the pluripotency of human and primate ES cells due to ethical reasons. It appeared that although the human ES cells are similar in a number of characteristics to mouse ES cells (Wobus, Boheler, 2005; Huang et al., 2014), they differ significantly from them (Friel et al., 2005; Watanabe et al., 2007). It is assumed that human ES cells belong to the "primed", and mouse ES cells belong to the "naïve" status of pluripotency.

Pluripotent ES cells of humans and primates in "naïve" status. Human ES cells in "naïve" status were first obtained in 2010 by the method of ectopic induction of the factors Oct4, Klf4, and Klf2 in combination with LIF and the inhibitors GSK3β and ERK1/2 (Hanna et al., 2010). Then, the cultivation medium and cultivation conditions were optimized (Gafni et al., 2013). Attempts have also been made to obtain pluripotent cells in the "naïve" status in primates (Fang et al., 2014; Chen Y. et al., 2015; De Los Angeles et al., 2019). Today, numerous studies are aimed at obtaining pluripotent ES cells in "naïve" status and maintaining this status under culture conditions (Liu et al., 2017; Kilens et al., 2018). Due to ethical reasons, it is not possible to test the "naïve" status of these human pluripotent cells, but it is possible to determine the putative criteria by which these cells could be considered pluripotent in "naïve" status. Today, the so-called "naïve" factors of pluripotency are already described. One of these factors is KLF4, which is specific for mouse "naïve" pluripotent stem cells and for human preimplantation embryos (Guo et al., 2009; Dunn et al., 2014; Boroviak et al., 2016). In addition, cells in the "naïve" status are characterized by nuclear localization of TFE3 and a high level of mitochondrial respiration (Zhou et al., 2012; Betschinger et al., 2013). Other researchers have demonstrated that the level of transcription of transposons corresponds to the status of pluripotency; in addition, the induction of the "naive" cell status is accompanied by DNA hypomethylation (Theunissen et al., 2016; Wang, Li, 2017).

Obtaining "naïve" status in somatic cells. iPSC. Simultaneously with the study of the "naïve" status of pluripotent ES cells, the technologies for the production of iPSCs from somatic cells were actively developing. The iPSCs are a new type of pluripotent cells that can be obtained by reprogramming differentiated somatic cells. For the first time iPSCs from somatic cells were obtained by exogenous expression of transcription factors in 2006 (Takahashi, Yamanaka, 2006). The essence of the method is the transfection of an adult cell with four genes (*Oct4*, *Sox2*, *Klf4* and *c-Myc*), which encode transcription factors associated with the pluripotent status of embryonic cells. Researchers were able to obtain human iPSC cell lines that meet all the criteria for ES cells from human skin fibroblasts (Takahashi et al., 2007; Yu et al., 2007) and from human skin keratocytes (Aasen et al., 2008). Since ectopic expression of the *c-Myc* and *Klf4* genes is undesirable due to the high risk of forming malignant tumors, these genes were successfully replaced with the less dangerous genes *Nanog* and *Lin28* in 2007 (Okita et al., 2007; Yu et al., 2007).

The iPSC cells are very similar to ES cells: similar morphology and growth profile, and the same culture conditions (growth factors and signaling molecules). The iPSCs retain the normal karyotype during cultivation, have high telomerase activity, and differentiate *in vitro* into tissue cells of all three germ layers (Yu et al., 2007).

Capabilities and limitations of using "naïve" ES cells and iPSCs. The unique properties of ES cells and iPSCs make it possible to obtain "chimera-competent" cells for blastocyst complementation. When ES and iPSCs are injected into the blastocyst, these cells are included into development, leading to the formation of animals with a high degree of chimerism. The properties of ES cells and human iPSCs make them an exceptional source for obtaining tissues and organs in transplantation and create prospects for the development of new approaches for the treatment of incurable diseases. The technology for generation iPSCs also demonstrates the possibilities for generation autologous stem cells, which in the future will allow to solve the problem of immunological compatibility during transplantation of organs from chimeric animals to a patient. In addition, this technology makes it possible to obtain pluripotent stem cells from various types of somatic cells, thus avoiding the ethical issues associated with the use of living embryos.

However, there are some limitations. The cultured ES cells and iPSCs vary significantly in their pluripotent differentiation potential and gene expression profile (Yu et al., 2007). In the population of the obtained ES cells and iPSCs, undifferentiated cells remain which can give rise to a tumor or reactivation of viruses. It also remains a problem to obtain a large number of "chimera-competent" cells of high quality suitable for clinical use. In addition, heritable epigenetic disorders were found in cultured ES cells, which may be associated with the development of hereditary diseases and carcinogenesis (Allegrucci et al., 2007). Consequently, there is a necessity to standardize the condition for obtaining, cultivating, and assessing the pluripotent status of iPSCs and ES cells.

Application of the modern method of blastocyst complementation

Interspecies chimeras of humans and rodents. The availability of "chimera-competent" human cells, generating the animals with "free niches" in organogenesis and obtaining interspecies chimeras of animals by the method of blastocyst complementation made it possible to make attempts to create chimeric organisms between humans and other animals. In 2006 for the first time, human ES cells at the early stages of embryogenesis were injected into a mouse blastocyst; the obtained chimeras showed developmental abnormalities (James et al., 2006). In 2013, chimeric mice were obtained by injecting human iPSCs; however, for ethical reasons, the mouse embryos were sacrificed at an early stage of development (Gafni et al., 2013). Then, in 2014, chimeric animals were obtained by microinjection of "naïve" iPSCs obtained from *Rhesus macaque* fibroblasts into a mouse embryo at the blastocyst stage (Fang et al., 2014).

However, in the obtained interspecies chimeras, the degree of revealed chimerism was low, especially in comparison with the degree of chimerism in intraspecies chimeras among rodents. It is speculated that this might be due to the evolutionary distance between humans and other animals. Interestingly, attempts to obtain an interspecies human chimera were successful when human iPSCs were injected into a mouse embryo at a later stage of embryonic development – at the gastrula stage (Mascetti, Pedersen, 2016b). Thus, the ability to form chimeras depends on the coordination of the *in vitro* developmental stages of donor cells with the *in vivo* embryo developmental stages.

Interspecies chimeras of humans and large domestic animals. In 2017, chimeric embryos were obtained between a human and a pig, as well as between a human and a cow (Wu et al., 2017). In this study, the researchers used CRISPR/Cas9 genetic editing to create a "free niche" in combination with blastocyst complementation. Their results demonstrated that "naïve" human pluripotent stem cells proliferate in porcine and bovine preimplantation blastocysts, while their ability to proliferate is limited in porcine postimplantation blastocysts. Interestingly, with the use of so-called "intermediate human pluripotent stem cells", the degree of chimerism and the ability to proliferate into various cell types in post-implantation pig embryos was higher (Tsukiyama, Ohinata, 2014; Wu et al., 2017). Recently, the creation of a chimeric embryo between *Macaca fascicularis* and a pig was reported, functioning donor ES cells of the primate were detected in the tissues of the pig (Fu et al., 2020).

Artificial embryo

The creation of an artificial embryo is a promising alternative to the use of animal and human embryos for research purposes. Different researchers have demonstrated the creation of embryo-like formations on stem cell culture (Pera et al., 2015; Harrison et al., 2017). In 2017, the possibility of creating artificial embryos was demonstrated by the aggregation of trophoblastic stem cells and totipotent ES cells, which independently assemble into a blastocyst on a substrate of a three-dimensional extracellular matrix. Scientists have shown that the development of the embryo, its morphogenesis, structure and cellular composition follow the same development patterns as in a normal embryo (Harrison et al., 2017).

Then, in 2018, a fully-fledged blastocyst model was created, which was called the blastoid (Rivron et al., 2018). Recently, three main types of stem cells have been obtained from fibroblasts: epiblast cells, primitive endoderm cells, and trophectoderm cells. To obtain a certain type of these pluripotent cells, a combination of five transcription factors was selected: Gata3, Eomes, Tfap2c, Myc, and Esrrb. This achievement could lead to the creation *in vitro* of fully-fledged artificial embryos without the use of an egg and a sperm cell (Benchetrit et al., 2019). Advances in the creation of an artificial embryo demonstrate the possibility of using it to obtain chimeric organisms in the future.

The main problems hindering the development of technologies for generating organs in chimeric animals, and possible ways to solve them

Growing rat organs in a mouse organism and generation of man-pig, man-cow chimeras give us the possibility of creating xenogeneic organisms among various animal species and generating human organs in the future. The candidate animals for organ transplant growing considered are pigs, cows, sheep, and primates.

The development of technology for farming human organs in xenogeneic animals such as pigs is hindered by a number of factors. There is a risk of zoonosis and the risk of contamination of human organs with cells or proteins of the recipient animal (Rashid et al., 2014; Matsunari et al., 2020). One problem is that retroviruses integrated into the genome of chimeric animals can be transferred to humans when growing human organs. The consequences of the incorporation of animal retroviruses into the human genome cannot be predicted. There are fears that human organs derived from chimeric animals could be a source of danger.

In addition, there are a number of poorly identified and poorly understood biological factors associated with differences in the rate of embryonic development in different species (Barry et al., 2017). Understanding the mechanisms of these differences, the ability to modulate the time and developmental stage of donor cells *in vitro*, and the ability to influence the developmental stage *in vivo* would allow the synchronization of donor and host cells in a chimeric model. Recent studies have shown that the synchronization of developmental stages between donor cultured pluripotent ES cells and the recipient is a significant criterion for the successful formation of a chimera. For example, "naïve" mouse ES cells are involved in the formation of a chimera only when injected at the blastocyst stage, while "primed" mouse ES cells isolated from the epiblast are involved in the formation of a chimera when injected at the gastrula stage (Huang et al., 2012).

It is also interesting that attempts to obtain an interspecies human chimera were successful when the injection was car-

ried out at a later stage of embryonic development. Successful microinjection of human iPSCs into a mouse embryo at the gastrula stage was demonstrated in 2016, which confirms the hypothesis that the ability to form chimeras depends on the coordination of the *in vitro* stages of donor cells with the stage of *in vivo* host embryo development (Mascetti, Pedersen, 2016b).

One of the problems of generating interspecies human chimeras is the low percentage of donor cells in the chimeric organism. It is assumed that the negative results and low degree of chimerism in experiments on generating chimeras are associated with the apoptosis of cells. In 2016, it was demonstrated that expression of the anti-apoptotic gene *Bcl2* in "chimera-incompetent" epiblast stem cells in rat allows these cells to turn into "chimera-competent" cells and participate in the formation of all tissues in a chimeric rat-mouse embryo when injected into a mouse blastocyst (Masaki et al., 2016).

Very recently, it became possible to create human-mouse chimeric embryos in which the proportion of human cells for the first time was 4 %. In this study, «naïve» human PSCs obtained by the inhibition of mTOR protein kinase were microinjected into mouse blastocysts (Hu et al., 2020).

Another important problem to be solved for the successful cultivation of donor human organs in chimeric organisms is the problem of organ vascularization. Previous studies have demonstrated that vessels in chimeric organisms are formed from the cells of both donor and recipient (Kobayashi et al., 2010; Usui et al., 2012; Yamaguchi et al., 2017). For the successful transplantation of human organs grown in animals, it is necessary for the organ's circulatory system, like the organ, to be formed from human cells in order to minimize the xenogenic component during transplantation. Many researchers are working on this problem (Hamanaka et al., 2018; Matsunari et al., 2020). To solve all these problems, the factors influencing the success of the colonization of pluripotent donor cells into the organism of the recipient animal, and the mechanisms underlying the differentiation of these cells in the conditions of the "free niche" are still to be determined and investigated.

Besides biological, there are also ethical barriers. For example, one of the issues that can arise with interspecies human-animal chimeras is the production of gametes with the human genome in chimeric animals (Bourret et al., 2016; Farahany et al., 2018). Concerns are also raised by the likelihood of humanization of chimeric animals upon accidental differentiation of human cells in the brain tissues of the recipient (Shaw et al., 2015). In 2019, it was demonstrated that these issues can be solved by disabling the *Prdm14* and *Otx2* genes responsible for the formation of gametes and the brain in microinjected "chimera-competent" cells (Hashimoto et al., 2019).

Conclusion

Thus, in order to carry out successful blastocyst complementation and obtain an interspecies chimera between a human and another animal for the purpose of growing organs for transplantation, two key technologies need to be improved: (1) creation of animals with "free niches", and (2) ethical generation of pluripotent "chimera-competent" human cells capable of differentiating into a target organ or tissue in the body of a host animal. In addition, it is necessary to understand and overcome the biological barriers that cause the absence or low percentage of chimerism of pluripotent "chimera-competent" ES cells in the animal organism. It is also important to regulate emerging ethical issues at the legislative level. Despite all the difficulties, the technology of growing donor organs in chimeric organisms is very promising.

References

Aasen T., Raya A., Barrero M.J., Garreta E., Consiglio A., Gonzalez F., Vassena R., Bilić J., Pekarik V., Tiscornia G., Edel M., Boué S., Izpisúa Belmonte J.C. Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. *Nat. Biotechnol*. 2008;26(11):1276-1284. DOI 10.1038/nbt.1503.

Allegrucci C., Wu Y.Z., Thurston A., Denning C.N., Priddle H., Mummery C.L., Ward-van Oostwaard D., Andrews P.W., Stojkovic M., Smith N., Parkin T., Jones M.E., Warren G., Yu L., Brena R.M., Plass C., Young L.E. Restriction landmark genome scanning identifies culture-induced DNA methylation instability in the human embryonic stem cell epigenome. *Hum. Mol. Genet*. 2007;16(10):1253-1268. DOI 10.1093/hmg/ddm074.

Barry C., Schmitz M.T., Jiang P., Schwartz M.P., Duffin B.M., Swanson S., Bacher R., Bolin J.M., Elwell A.L., McIntosh B.E., Stewart R., Thomson J.A. Species-specific developmental timing is maintained by pluripotent stem cells *ex utero*. *Dev. Biol*. 2017;423(2):101-110. DOI 10.1016/j.ydbio.2017.02.002.

Benchetrit H., Jaber M., Zayat V., Sebban S., Pushett A., Makedonski K., Zakheim Z., Radwan A., Maoz N., Lasry R., Renous N., Inbar M., Ram O., Kaplan T., Buganim Y. Direct induction of the three pre-implantation blastocyst cell types from fibroblasts. *Cell Stem Cell.* 2019;24(6):983-994.e7. DOI 10.1016/j.stem.2019.03.018.

Betschinger J., Nichols J., Dietmann S., Corrin P.D., Paddison P.J., Smith A. Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell*. 2013;153(2):335-347. DOI 10.1016/j.cell.2013.03.012.

Boroviak K., Doe B., Banerjee R., Yang F., Bradley A. Chromosome engineering in zygotes with CRISPR/Cas9. *Genesis.* 2016;54(2):78-85. DOI 10.1002/dvg.22915.

Bourret R., Martinez E., Vialla F., Giquel C., Thonnat-Marin A., De Vos J. Human-animal chimeras: ethical issues about farming chimeric animals bearing human organs. *Stem Cell Res. Ther*. 2016;7(1):87. DOI 10.1186/s13287-016-0345-9.

Bradley A., Evans M., Kaufman M.H., Robertson E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*. 1984;309(5965):255-256. DOI 10.1038/309255a0.

Carbery I.D., Ji D., Harrington A., Brown V., Weinstein E.J., Liaw L., Cui X. Targeted genome modification in mice using zinc-finger nucleases. *Genetics.* 2010;186(2):451-459. DOI 10.1534/genetics.110.117002.

Chang A.N., Liang Z., Dai H.Q., Chapdelaine-Williams A.M., Andrews N., Bronson R.T., Schwer B., Alt F.W. Neural blastocyst complementation enables mouse forebrain organogenesis. *Nature.* 2018;563(7729):126-130. DOI 10.1038/s41586-018-0586-0.

Chen J., Lansford R., Stewart V., Young F., Alt F.W. RAG-2-deficient blastocyst complementation: an assay of gene function in lymphocyte development. *Proc. Natl. Acad. Sci. USA.* 1993;90(10):4528-4532. DOI 10.1073/pnas.90.10.4528.

Chen K., Chmait R.H., Vanderbilt D., Wu S., Randolph L. Chimerism in monochorionic dizygotic twins: case study and review. *Am. J. Med. Genet. A*. 2013;161A(7):1817-1824. DOI 10.1002/ajmg.a.35957.

Chen Y., Niu Y., Li Y., Ai Z., Kang Y., Shi H., Xiang Z., Yang Z., Tan T., Si W., Li W., Xia X., Zhou Q., Ji W., Li T. Generation of cynomolgus monkey chimeric fetuses using embryonic stem cells. *Cell Stem Cell*. 2015;17(1):116-124. DOI 10.1016/j.stem.2015.06.004.

Cong L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W., Marraffini L.A., Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science.* 2013;339(6121):819-823. DOI 10.1126/science.1231143.

De Los Angeles A., Elsworth J.D., Redmond D.E. ERK-independent African Green monkey pluripotent stem cells in a putative chimera-competent state. *Biochem. Biophys. Res. Commun*. 2019;510(1):78-84. DOI 10.1016/j.bbrc.2019.01.037.

Drexler C., Glock B., Vadon M., Staudacher E., Dauber E.M., Ulrich S., Reisacher B.K., Mayr W.R., Lanzer G., Wagner T. Tetragametic chimerism detected in a healthy woman with mixed-field agglutination reactions in ABO blood grouping. *Transfusion*. 2005;45(5):698-703. DOI 10.1111/j.1537-2995.2005.04304.x.

Dunn S.J., Martello G., Yordanov B., Emmott S., Smith A.G. Defining an essential transcription factor program for naïve pluripotency. *Science.* 2014;344(6188):1156-1160. DOI 10.1126/science.1248882.

Evans M.J., Kaufman M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature.* 1981;292(5819):154-156. DOI 10.1038/292154a.

Fang R., Liu K., Zhao Y., Li H., Zhu D., Du Y., Xiang C., Li X., Liu H., Miao Z., Zhang X., Shi Y., Yang W., Xu J., Deng H. Generation of naive induced pluripotent stem cells from rhesus monkey fibroblasts. *Cell Stem Cell*. 2014;15(4):488-497. DOI 10.1016/j.stem.2014.09.004.

Farahany N.A., Greely H.T., Hyman S., Koch C., Grady C., Pașca S.P., Sestan N., Arlotta P., Bernat J.L., Ting J., Lunshof J.E., Iyer E., Hyun I., Capestany B.H., Church G.M., Huang H., Song H. The ethics of experimenting with human brain tissue. *Nature*. 2018;556(7702):429-432. DOI 10.1038/d41586-018-04813-x.

Friel R., van der Sar S., Mee P.J. Embryonic stem cells: understanding their history, cell biology and signalling. *Adv. Drug. Deliv. Rev*. 2005;57(13):1894-1903. DOI 10.1016/j.addr.2005.08.002.

Fu R., Yu D., Ren J., Li C., Wang J., Feng G., Wang X., Wan H., Li T., Wang L., Zhang Y., Hai T., Li W., Zhou Q. Domesticated cynomolgus monkey embryonic stem cells allow the generation of neonatal interspecies chimeric pigs. *Protein Cell*. 2020;11(2):97-107. DOI 10.1007/s13238-019-00676-8.

Gafni O., Weinberger L., Mansour A.A., Manor Y.S., Chomsky E., Ben-Yosef D., Kalma Y., Viukov S., Maza I., Zviran A., Rais Y., Shipony Z., Mukamel Z., Krupalnik V., Zerbib M., Geula S., Caspi I., Schneir D., Shwartz T., Gilad S., Amann-Zalcenstein D., Benjamin S., Amit I., Tanay A., Massarwa R., Novershtern N., Hanna J.H. Derivation of novel human ground state naive pluripotent stem cells. *Nature*. 2013;504(7479):282-286. DOI 10.1038/nature12745.

Gardner R.L. Mouse chimeras obtained by the injection of cells into the blastocyst. *Nature*. 1968;220(5167):596-597. DOI 10.1038/220596a0.

Goto T., Hara H., Sanbo M., Masaki H., Sato H., Yamaguchi T., Hochi S., Kobayashi T., Nakauchi H., Hirabayashi M. Generation of pluripotent stem cell-derived mouse kidneys in Sall1-targeted anephric rats. *Nat. Commun.* 2019;10(1):451. DOI 10.1038/s41467-019-08394-9.

Guo G., Yang J., Nichols J., Hall J.S., Eyres I., Mansfield W., Smith A. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development.* 2009;136(7):1063-1069. DOI 10.1242/dev.030957.

Hamanaka S., Umino A., Sato H., Hayama T., Yanagida A., Mizuno N., Kobayashi T., Kasai M., Suchy F.P., Yamazaki S., Masaki H., Yamaguchi T., Nakauchi H. Generation of vascular endothelial cells and hematopoietic cells by blastocyst complementation. *Stem Cell Reports*. 2018;11(4):988-997. DOI 10.1016/j.stemcr.2018.08.015.

Hanna J., Cheng A.W., Saha K., Kim J., Lengner C.J., Soldner F., Cassady J.P., Muffat J., Carey B.W., Jaenisch R. Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc. Natl. Acad. Sci. USA*. 2010;107(20):9222-9227. DOI 10.1073/pnas.1004584107.

Harrison S.E., Sozen B., Christodoulou N., Kyprianou C., Zernicka-Goetz M. Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis *in vitro*. *Science*. 2017;356(6334). DOI 10.1126/science.aal1810.

Hashimoto H., Eto T., Yamamoto M., Yagoto M., Goto M., Kagawa T., Kojima K., Kawai K., Akimoto T., Takahashi R.I. Development of blastocyst complementation technology without contributions to gametes and the brain. *Exp. Anim*. 2019;68(3):361-370. DOI 10.1538/expanim.18-0173.

Hayashi K., Hikabe O., Obata Y., Hirao Y. Reconstitution of mouse oogenesis in a dish from pluripotent stem cells. *Nat. Protoc*. 2017;12(9):1733-1744. DOI 10.1038/nprot.2017.070.

Hu Z., Li H., Jiang H., Ren Y., Yu X., Qiu J., Stablewski A.B., Zhang B., Buck M.J., Feng J. Transient inhibition of mTOR in human pluripotent stem cells enables robust formation of mouse-human chimeric embryos. *Sci. Adv.* 2020;6(20):eaaz0298. DOI 10.1126/sciadv.aaz0298.

Huang Y., Liang P., Liu D., Huang J., Songyang Z. Telomere regulation in pluripotent stem cells. *Protein Cell*. 2014;5(3):194-202. DOI 10.1007/s13238-014-0028-1.

Huang Y., Osorno R., Tsakiridis A., Wilson V. *In vivo* differentiation potential of epiblast stem cells revealed by chimeric embryo formation. *Cell Rep*. 2012;2(6):1571-1578. DOI 10.1016/j.celrep.2012.10.022.

Isotani A., Hatayama H., Kaseda K., Ikawa M., Okabe M. Formation of a thymus from rat ES cells in xenogeneic nude mouse↔rat ES chimeras. *Genes Cells.* 2011;16(4):397-405. DOI 10.1111/j.1365-2443.2011.01495.x.

James D., Noggle S.A., Swigut T., Brivanlou A.H. Contribution of human embryonic stem cells to mouse blastocysts. *Dev. Biol*. 2006;295(1):90-102. DOI 10.1016/j.ydbio.2006.03.026.

Kilens S., Meistermann D., Moreno D., Chariau C., Gaignerie A., Reignier A., Lelièvre Y., Casanova M., Vallot C., Nedellec S., Flippe L., Firmin J., Song J., Charpentier E., Lammers J., Donnart A., Marec N., Deb W., Bihouée A., Le Caignec C., Pecqueur C., Redon R., Barrière P., Bourdon J., Pasque V., Soumillon M., Mikkelsen T.S., Rougeulle C., Fréour T., David L., Milieu Intérieur. Consortium. Parallel derivation of isogenic human primed and naive induced pluripotent stem cells. *Nat. Commun*. 2018;9(1):360. DOI 10.1038/s41467-017-02107-w.

Kobayashi T., Yamaguchi T., Hamanaka S., Kato-Itoh M., Yamazaki Y., Ibata M., Sato H., Lee Y.S., Usui J., Knisely A.S., Hirabayashi M., Nakauchi H. Generation of rat pancreas in mouse by interspecific blastocyst injection of pluripotent stem cells. *Cell*. 2010;142(5):787-799. DOI 10.1016/j.cell.2010.07.039.

Liu X., Nefzger C.M., Rossello F.J., Chen J., Knaupp A.S., Firas J., Ford E., Pflueger J., Paynter J.M., Chy H.S., O'Brien C.M., Huang C., Mishra K., Hodgson-Garms M., Jansz N., Williams S.M., Blewitt M.E., Nilsson S.K., Schittenhelm R.B., Laslett A.L., Lister R., Polo J.M. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. *Nat. Methods.* 2017;14(11):1055-1062. DOI 10.1038/nmeth.4436.

MacLaren L.A., Anderson G.B., BonDurant R.H., Edmondson A.J. Inter- and intraspecific placentae in sheep, goats and sheep-goat chimaeras. *J. Comp. Pathol*. 1992;106(3):279-297. DOI 10.1016/0021-9975(92)90056-z.

Mali P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E., Church G.M. RNA-guided human genome engineering

via Cas9. *Science*. 2013;339(6121):823-826. DOI 10.1126/science.1232033.

Martin G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA.* 1981;78(12):7634-7638. DOI 10.1073/pnas.78.12.7634.

Masaki H., Kato-Itoh M., Takahashi Y., Umino A., Sato H., Ito K., Yanagida A., Nishimura T., Yamaguchi T., Hirabayashi M., Era T., Loh K.M., Wu S.M., Weissman I.L., Nakauchi H. Inhibition of apoptosis overcomes stage-related compatibility barriers to chimera formation in mouse embryos. *Cell Stem Cell*. 2016;19(5):587-592. DOI 10.1016/j.stem.2016.10.013.

Mascetti V.L., Pedersen R.A. Contributions of mammalian chimeras to pluripotent stem cell research. *Cell Stem Cell.* 2016a;19(2):163-175. DOI 10.1016/j.stem.2016.07.018.

Mascetti V.L., Pedersen R.A. Human-mouse chimerism validates human stem cell pluripotency. *Cell Stem Cell.* 2016b;18(1):67-72. DOI 10.1016/j.stem.2015.11.017.

Matsunari H., Watanabe M., Hasegawa K., Uchikura A., Nakano K., Umeyama K., Masaki H., Hamanaka S., Yamaguchi T., Nagaya M., Nishinakamura R., Nakauchi H., Nagashima H. Compensation of disabled organogeneses in genetically modified pig fetuses by blastocyst complementation. *Stem Cell Reports*. 2020;14(1):21-33. DOI 10.1016/j.stemcr.2019.11.008.

McLaren A., Bowman P. Mouse chimaeras derived from fusion of embryos differing by nine genetic factors. *Nature.* 1969;224(5216):238-240. DOI 10.1038/224238a0.

Mintz B. Genetic mosaicism in adult mice of quadriparental lineage. *Science.* 1965;148(3674):1232-1233. DOI 10.1126/science.148.3674.1232.

Nelson J.L., Furst D.E., Maloney S., Gooley T., Evans P.C., Smith A., Bean M.A., Ober C., Bianchi D.W. Microchimerism and HLA-compatible relationships of pregnancy in scleroderma. *Lancet.* 1998;351(9102):559-562. DOI 10.1016/S0140-6736(97)08357-8.

Nichols J., Smith A. Naive and primed pluripotent states. *Cell Stem Cell.* 2009;4(6):487-492. DOI 10.1016/j.stem.2009.05.015.

Offield M.F., Jetton T.L., Labosky P.A., Ray M., Stein R.W., Magnuson M.A., Hogan B.L., Wright C.V. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development.* 1996;122(3):983-995.

Ohinata Y., Payer B., O'Carroll D., Ancelin K., Ono Y., Sano M., Barton S.C., Obukhanych T., Nussenzweig M., Tarakhovsky A., Saitou M., Surani M.A. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature.* 2005;436(7048):207-213. DOI 10.1038/nature03813.

Okita K., Ichisaka T., Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature.* 2007;448(7151):313-317. DOI 10.1038/nature05934.

Okumura H., Nakanishi A., Toyama S., Yamanoue M., Yamada K., Ukai A., Hashita T., Iwao T., Miyamoto T., Tagawa Y.I., Hirabayashi M., Miyoshi I., Matsunaga T. Contribution of rat embryonic stem cells to xenogeneic chimeras in blastocyst or 8-cell embryo injection and aggregation. *Xenotransplantation.* 2019;26(1):e12468. DOI 10.1111/xen.12468.

Pera M.F., de Wert G., Dondorp W., Lovell-Badge R., Mummery C.L., Munsie M., Tam P.P. What if stem cells turn into embryos in a dish? *Nat. Methods*. 2015;12(10):917-919. DOI 10.1038/nmeth.3586.

Rashid T., Kobayashi T., Nakauchi H. Revisiting the flight of Icarus: making human organs from PSCs with large animal chimeras. *Cell Stem Cell.* 2014;15(4):406-409. DOI 10.1016/j.stem.2014.09.013.

Rivron N.C., Frias-Aldeguer J., Vrij E.J., Boisset J.C., Korving J., Vivié J., Truckenmüller R.K., van Oudenaarden A., van Blitterswijk C.A., Geijsen N. Blastocyst-like structures generated solely from stem cells. *Nature*. 2018;557(7703):106-111. DOI 10.1038/s41586-018-0051-0.

Shaw D., Dondorp W., Geijsen N., de Wert G. Creating human organs in chimaera pigs: an ethical source of immunocompatible organs? *J. Med. Ethics*. 2015;41(12):970-974. DOI 10.1136/medethics-2014-102224.

Stanger B.Z., Tanaka A.J., Melton D.A. Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver. *Nature.* 2007;445(7130):886-891. DOI 10.1038/nature05537.

Suchy F., Nakauchi H. Lessons from interspecies mammalian chimeras. *Annu. Rev. Cell Dev. Biol.* 2017;33:203-217. DOI 10.1146/annurev-cellbio-100616-060654.

Sung Y.H., Baek I.J., Kim D.H., Jeon J., Lee J., Lee K., Jeong D., Kim J.S., Lee H.W. Knockout mice created by TALEN-mediated gene targeting. *Nat. Biotechnol.* 2013;31(1):23-24. DOI 10.1038/nbt.2477.

Tachibana M., Sparman M., Ramsey C., Ma H., Lee H.S., Penedo M.C., Mitalipov S. Generation of chimeric rhesus monkeys. *Cell*. 2012;148(1-2):285-295. DOI 10.1016/j.cell.2011.12.007.

Takahashi K., Tanabe K., Ohnuki M., Narita M., Ichisaka T., Tomoda K., Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*. 2007;131(5):861-872. DOI 10.1016/j.cell.2007.11.019.

Takahashi K., Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006;126(4):663-676. DOI 10.1016/j.cell.2006.07.024.

Tarkowski A.K. Mouse chimaeras developed from fused eggs. *Nature*. 1961;190:857-860. DOI 10.1038/190857a0.

Tesar P.J., Chenoweth J.G., Brook F.A., Davies T.J., Evans E.P., Mack D.L., Gardner R.L., McKay R.D. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature.* 2007;448(7150):196-199. DOI 10.1038/nature05972.

Theunissen T.W., Friedli M., He Y., Planet E., O'Neil R.C., Markoulaki S., Pontis J., Wang H., Iouranova A., Imbeault M., Duc J., Cohen M.A., Wert K.J., Castanon R., Zhang Z., Huang Y., Nery J.R., Drotar J., Lungjangwa T., Trono D., Ecker J.R., Jaenisch R. Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell.* 2016;19(4):502-515. DOI 10.1016/j.stem.2016.06.011.

Thomson J.A., Itskovitz-Eldor J., Shapiro S.S., Waknitz M.A., Swiergiel J.J., Marshall V.S., Jones J.M. Embryonic stem cell lines derived from human blastocysts. *Science.* 1998;282(5391):1145-1147. DOI 10.1126/science.282.5391.1145.

Thomson J.A., Kalishman J., Golos T.G., Durning M., Harris C.P., Becker R.A., Hearn J.P. Isolation of a primate embryonic stem cell line. *Proc. Natl. Acad. Sci. USA.* 1995;92(17):7844-7848. DOI 10.1073/pnas.92.17.7844.

Tippett P. Blood group chimeras. A review. *Vox Sang.* 1983;44(6):333-359. DOI 10.1111/j.1423-0410.1983.tb03657.x.

Tsukiyama T., Ohinata Y. A modified EpiSC culture condition containing a GSK3 inhibitor can support germline-competent pluripotency in mice. *PLoS One.* 2014;9(4):e95329. DOI 10.1371/journal.pone.0095329.

Usui J., Kobayashi T., Yamaguchi T., Knisely A.S., Nishinakamura R., Nakauchi H. Generation of kidney from pluripotent stem cells via blastocyst complementation. *Am. J. Pathol*. 2012;180(6):2417-2426. DOI 10.1016/j.ajpath.2012.03.007.

Wang R., Li T. DNA methylation is correlated with pluripotency of stem cells. *Curr. Stem Cell Res. Ther.* 2017;12(6):442-446. DOI 10.2174/1574888X11666161226145432.

Watanabe K., Ueno M., Kamiya D., Nishiyama A., Matsumura M., Wataya T., Takahashi J.B., Nishikawa S., Nishikawa S., Muguruma K., Sasai Y. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol*. 2007;25(6):681-686. DOI 10.1038/nbt1310.

Whitworth K.M., Lee K., Benne J.A., Beaton B.P., Spate L.D., Murphy S.L., Samuel M.S., Mao J., O'Gorman C., Walters E.M., Murphy C.N., Driver J., Mileham A., McLaren D., Wells K.D., Prather R.S. Use of the CRISPR/Cas9 system to produce genetically

engineered pigs from in vitro-derived oocytes and embryos. *Biol. Reprod*. 2014;91(3):78. DOI 10.1095/biolreprod.114.121723.

Wobus A.M., Boheler K.R. Embryonic stem cells: prospects for developmental biology and cell therapy. *Physiol. Rev.* 2005;85(2):635-678. DOI 10.1152/physrev.00054.2003.

Wu J., Platero-Luengo A., Sakurai M., Sugawara A., Gil M.A., Yamauchi T., Suzuki K., Bogliotti Y.S., Cuello C., Morales Valencia M., Okumura D., Luo J., Vilariño M., Parrilla I., Soto D.A., Martinez C.A., Hishida T., Sánchez-Bautista S., Martinez-Martinez M.L., Wang H., Nohalez A., Aizawa E., Martinez-Redondo P., Ocampo A., Reddy P., Roca J., Maga E.A., Esteban C.R., Berggren W.T., Nuñez Delicado E., Lajara J., Guillen I., Guillen P., Campistol J.M., Martinez E.A., Ross P.J., Izpisua Belmonte J.C. Interspecies chimerism with mammalian pluripotent stem cells. *Cell.* 2017;168(3):473-486. e415. DOI 10.1016/j.cell.2016.12.036.

Yamaguchi T., Sato H., Kato-Itoh M., Goto T., Hara H., Sanbo M., Mizuno N., Kobayashi T., Yanagida A., Umino A., Ota Y., Hamanaka S., Masaki H., Rashid S.T., Hirabayashi M., Nakauchi H. Interspecies organogenesis generates autologous functional islets. *Nature*. 2017;542(7640):191-196. DOI 10.1038/nature21070.

Yu J., Vodyanik M.A., Smuga-Otto K., Antosiewicz-Bourget J., Frane J.L., Tian S., Nie J., Jonsdottir G.A., Ruotti V., Stewart R., Slukvin I.I., Thomson J.A. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007;318(5858):1917-1920. DOI 10.1126/science.1151526.

Zhou W., Choi M., Margineantu D., Margaretha L., Hesson J., Cavanaugh C., Blau C.A., Horwitz M.S., Hockenbery D., Ware C., Ruohola-Baker H. HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J*. 2012;31(9):2103-2116. DOI 10.1038/emboj.2012.71.

ORCID ID

T.I. Babochkina orcid.org/0000-0003-3041-2789
L.A. Gerlinskaya orcid.org/0000-0002-8118-1362
M.P. Moshkin orcid.org/0000-0002-5388-2946

Acknowledgements. The research was supported by the Russian Science Foundation grant No. 20-14-00055, budget project No. 0324-2019-0041 and carried out using the equipment of the Center for Genetic Resources of Laboratory Animals, Federal Research Center ICG SB RAS, supported by the Ministry of Education and Science of Russia (Unique project identifier RFMEFI62119X0023).
The authors are grateful to A.G. Menzorov and S.A. Fedorova for valuable comments during the preparation of the manuscript.
Conflict of interest. The authors declare no conflict of interest.

Received April 10, 2020. Revised October 21, 2020. Accepted November 17, 2020.

Алфавитный указатель авторов статей, опубликованных в журнале в 2020 г.

Прием статей через электронную редакцию на сайте http://vavilov.elpub.ru/index.php/jour
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции»/"Vavilov Journal of Genetics and Breeding"
до 2011 г. выходил под названием «Информационный вестник ВОГиС»/
"The Herald of Vavilov Society for Geneticists and Breeding Scientists".

Регистрационное свидетельство ПИ № ФС77-45870 выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 20 июля 2011 г.

«Вавиловский журнал генетики и селекции» включен ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Российский индекс научного цитирования, ВИНИТИ, базы данных Emerging Sources Citation Index (Web of Science), Zoological Record (Web of Science), Scopus, Ebsco, DOAJ, Ulrich's Periodicals Directory, Google Scholar, Russian Science Citation Index на платформе Web of Science, каталог научных ресурсов открытого доступа ROAD.

Открытый доступ к полным текстам:
на сайте ИЦиГ СО РАН – bionet.nsc.ru/vogis/
платформе Elpub – vavilov.elpub.ru/index.php/jour
платформе Научной электронной библиотеки – elibrary.ru/title_about.asp?id=32440

Подписку на «Вавиловский журнал генетики и селекции» можно оформить в любом почтовом отделении России. Индекс издания 42153 по каталогу «Пресса России».

✉ e-mail: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.
Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218.
Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.
Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.

Подписано в печать 21.12.2020. Выход в свет 30.12.2020. Формат 60 × 84 $^{1}/_{8}$. Усл. печ. л. 13.72.
Уч.-изд. л. 16.5. Тираж 150 экз. (1-й завод 1–50 экз.). Заказ № 203. Цена свободная.

Отпечатано в типографии ФГУП «Издательство СО РАН», Морской проспект, 2, Новосибирск, 630090.