2026 • 30 • 3

Online ISSN 2500-3259

Сетевое издание

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

VAVILOV JOURNAL OF GENETICS AND BREEDING

Основан в 1997 г.
Периодичность 8 выпусков в год
doi 10.18699/vjgb-26-38

Учредители

Сибирское отделение Российской академии наук

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

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

2026 • 30 • 3

Online ISSN 2500-3259

Online edition

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

VAVILOV JOURNAL OF GENETICS AND BREEDING

Founded in 1997
Publication frequency: 8 issues a year
doi 10.18699/vjgb-26-38

СОДЕРЖАНИЕ • 2026 • 30 • 3

Генетика человека

Медицинская генетика

Биоинформатика и системная биология

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII • VAVILOV JOURNAL OF GENETICS AND BREEDING

CONTENTS • 2026 • 30 • 3

MOLECULAR AND CELL BIOLOGY
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):345-361

doi 10.18699/vjgb-26-39

Concept of natural genome reconstruction.
Part 5. Analysis of changes in the lifespan of old animals after reinfusion of bone marrow cells derived from old animals and treated with hDNA^{gr} in combination with recombinant human angiogenin

V.S. Ruzanova [1]*, L.U. Grivtsova [2]*, S.G. Oshikhmina [1], A.S. Proskurina [1], G.S. Ritter [1], E.V. Dolgova [1], S.S. Kirikovich [1], E.V. Levites [1], Y.R. Efremov [1], T.D. Dubatolova [3], M.I. Meschaninova [4], A.L. Mamaev [5], O.S. Taranov [6], S.V. Sidorov [7, 8], O.Y. Leplina [9], A.A. Ostanin [9], E.R. Chernykh [9], N.A. Kolchanov [1], A.S. Bryukhovetskiy [10]#, S.S. Bogachev [1]#

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] National Medical Research Radiological Centre of the Ministry of Health of the Russian Federation, Obninsk, Russia
[3] Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[4] Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[5] Laboratory Angiopharm LLC, Novosibirsk, Russia
[6] State Scientific Center of Virology and Biotechnology "Vector" of Rospotrebnadzor, Koltsovo, Novosibirsk region, Russia
[7] Novosibirsk State University, Novosibirsk, Russia
[8] City Clinical Hospital No. 1, Novosibirsk, Russia
[9] Research Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia
[10] Clinical Hospital "NeuroVita", Moscow, Russia
labmolbiol@mail.ru

Abstract. Two series of tests were performed, on mice and rats, to assess the lifespan of old animals reinfused with bone marrow cells from old animals treated with fragmented human DNA (hDNA^{gr}), recombinant human angiogenin, and both preparations together. Animals reinfused with untreated bone marrow cells from old animals or bone marrow cells from young animals were used as comparison groups. Using both outbred mice and CBA/Lac mice, no significant increase in the lifespan of animals reinfused with bone marrow cells treated with the hDNA^{gr} was found compared with the group of mice reinfused with untreated bone marrow cells. Using the CBA/Lac line, mice reinfused with bone marrow cells treated with angiogenin simultaneously died of the characteristic symptom complex at 10 months after treatment. Pathomorphological analysis suggests that the simultaneous death of mice occurred as a result of pathological disorders in the excretory systems of animals. Reinfusion of bone marrow cells from old animals treated with angiogenin and hDNA^{gr} and bone marrow cells taken from young animals significantly increases the lifespan of mice in groups. The combined use of two activators, angiogenin and hDNA^{gr}, increased the average lifespan of 30 % of experimental mice to 35 months compared to 28 months in the control. Using Wistar rats as model animals in the first experiment, a reliable increase in the lifespan of rats with reinfusion of bone marrow cells from old animals treated with the hDNA^{gr} preparation to 28 months was shown compared to the group that received untreated bone marrow cells from old animals, where the average lifespan of rats was 24 months. In the second similar experiment, no reliable difference in the lifespan of rats for the two groups was shown. Animals injected with bone marrow cells treated with angiogenin lived significantly longer than rats from the control group. The analysis of the amount of telomeric DNA in bone marrow cells of rats from the experimental and control groups 12 months after treatment showed that there was no significant increase in telomeric DNA. A molecular/cellular model of aging of the organism associated with the concept of "natural reconstruction of the genome" is considered.

Key words: telomere length; life expectancy; double-stranded DNA preparation (hDNA^{gr}); recombinant human angiogenin; pathomorphological analysis; reinfusion of bone marrow cells

For citation: Ruzanova V.S., Grivtsova L.U., Oshikhmina S.G., Proskurina A.S., Ritter G.S., Dolgova E.V., Kirikovich S.S., Levites E.V., Efremov Y.R., Dubatolova T.D., Meschaninova M.I., Mamaev A.L., Taranov O.S., Sidorov S.V., Leplina O.Y., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bryukhovetskiy A.S., Bogachev S.S. Concept of natural genome reconstruction. Part 5. Analysis of changes in the lifespan of old animals after reinfusion of bone marrow cells derived from old animals and treated with hDNA^{gr} in combination with recombinant human angiogenin. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):345-361. doi 10.18699/vjgb-26-39

Funding. This work was supported by the Ministry of Science and Higher Education of the Russian Federation for the Institute of Cytology and Genetics (state budget-funded project No. FWNR-2026-0025) and by A.A. Purtov, I.N. Zaitseva and LLC "ES.LAB DIAGNOSTIC".

*, # Equal contributed.

Концепция природной реконструкции генома.
Часть 5. Анализ изменения продолжительности жизни старых животных после реинфузии клеток костного мозга старых животных, обработанных hDNA^{gr} и ангиогенином рекомбинантным человеческим

В.С. Рузанова [1]*, Л.Ю. Гривцова [2]*, С.Г. Ошихмина [1], А.С. Проскурина [1], Г.С. Риттер [1], Е.В. Долгова [1], С.С. Кирикович [1], Е.В. Левитес [1], Я.Р. Ефремов [1], Т.Д. Дубатолова [3], М.И. Мещанинова [4], А.Л. Мамаев [5], О.С. Таранов [6], С.В. Сидоров [7, 8], О.Ю. Леплина [9], А.А. Останин [9], Е.Р. Черных [9], Н.А. Колчанов [1], А.С. Брюховецкий [10#], С.С. Богачев [1#]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Национальный медицинский исследовательский центр радиологии Министерства здравоохранения Российской Федерации, Обнинск, Россия
[3] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
[4] Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия
[5] ООО «Лаборатория Ангиофарм», Новосибирск, Россия
[6] Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора, р.п. Кольцово, Новосибирская область, Россия
[7] Новосибирский национальный государственный исследовательский университет, Новосибирск, Россия
[8] Городская клиническая больница № 1, Новосибирск, Россия
[9] Научно-исследовательский институт фундаментальной и клинической иммунологии, Новосибирск, Россия
[10] АО Клинический госпиталь «НейроВита», Москва, Россия
labmolbiol@mail.ru

Аннотация. Проведены две серии тестов (на мышах и крысах) по оценке продолжительности жизни старых животных, которым были реинфузированы клетки костного мозга старых животных, обработанные фрагментированной ДНК человека (hDNA^{gr}), ангиогенином рекомбинантным человеческим и двумя препаратами совместно. В качестве групп сравнения использовались животные, которым реинфузировали необработанные клетки костного мозга старых животных или клетки костного мозга от молодых особей. В случае как аутбредных мышей, так и мышей линии CBA/Lac не выявлено достоверного увеличения продолжительности жизни животных при реинфузии клеток костного мозга, обработанных препаратом hDNA^{gr}, по сравнению с группой мышей, которым реинфузировали необработанные клетки костного мозга. При использовании линии CBA/Lac мыши, которым реинфузировали клетки костного мозга, обработанные ангиогенином, на 10-й месяц после обработки одновременно пали от характерного симптомокомплекса. Патоморфологический анализ предполагает, что одновременная гибель мышей произошла в результате патологических нарушений в выделительных системах животных. Реинфузия клеток костного мозга от старых животных, обработанных ангиогенином и hDNA^{gr}, и клеток костного мозга, взятых от молодых животных, значительно увеличивает продолжительность жизни мышей в группах. Совместное применение двух активаторов, ангиогенина и hDNA^{gr}, увеличивало среднюю продолжительность жизни 30 % экспериментальных мышей до 35 мес. при 28 мес. в контроле. При использовании в качестве модельных животных крыс линии Вистар в первом эксперименте было показано достоверное увеличение жизни крыс при реинфузии клеток костного мозга старых животных, обработанных препаратом hDNA^{gr}, до 28 мес. по сравнению с группой, получавшей необработанные клетки костного мозга старых животных, где средняя продолжительность жизни крыс составила 24 мес. Во втором аналогичном эксперименте достоверной разницы в продолжительности жизни крыс для указанных двух групп показано не было. Животные, которым вводили клетки костного мозга, обработанные ангиогенином, прожили достоверно дольше, чем крысы из контрольной группы. Проведенный анализ количества теломерной ДНК в клетках костного мозга крыс экспериментальных и контрольной групп через 12 мес. после обработки свидетельствовал, что достоверного увеличения теломерной ДНК не произошло. Рассматривается молекулярная/клеточная модель старения организма, связанная с концепцией «природной реконструкции генома».

Ключевые слова: длина теломеры; продолжительность жизни; препарат двуцепочечной ДНК (hDNA^{gr}); ангиогенин рекомбинантный человеческий; патоморфологический анализ; реинфузия клеток костного мозга

Introduction

The problem of aging is a dominant issue in modern biology, since the primary focus of scientific thought and applied research is searching for ways to halt the progressive increase in the number of people suffering from the so-called “diseases of civilization”, which are inherently related to this natural biological process.

Aging and death are fundamental, intrinsic biological properties representing the functioning and evolution of all living organisms, including humans. From a thermodynamic perspective, aging is a critical temporal threshold between the ability of an organism to reduce its entropy (also a fundamental intrinsic property of living systems) and successive progression toward an irreversible stationary state (i. e., death) characterized by maximum entropy (Lefever, 2018; Tlidi et al., 2018a, b). At the level of a living organism, aging is a dynamic pathophysiological process of accumulating alterations at multiple biological levels,

leading to gradual decline in vital activity and, ultimately, death. The pace of this universal and irreversible process (within the framework of conventional existence) is governed by numerous external and internal factors. At its core, there lies "genetic aging" of the organism's DNA repair machinery in the form of stem cells of various lineages that sustain coordinated responses maintaining internal homeostasis. All other pathological processes result from the failure of stem cells to preserve homeostasis (Bowen, Atwood, 2004; Mikheev et al., 2023).

The numerous existing theories of aging are generally categorized into two groups: the "programmed" (adaptive) and "damage" theories.

According to the programmed theories, functioning of a living organism is presumed to be biologically programmed by nature only for the period of its active life cycle, which includes development (i. e., organismal growth) and reproductive capacity (the so-called biological usefulness), determined by species-specific population expediency (da Costa et al., 2016). In other words, aging-related changes are regulated by a kind of biological sensor whose primary function is to monitor the developmental schedule of the organism until it reaches sexual maturity and reproductive capacity (Kozlov, 1999).

The damage theories posit that aging is not an inevitable sequela of an organism's existence but rather results from accumulation of damage, or stochastic errors, in the genetic information storage and transmission systems. Over time, compensation for these impairments becomes infeasible, thus leading to death Kirkwood, 2005; Vijg, Campisi, 2008; Gems, Partridge, 2013; Mikheev et al., 2023)

Proponents of the programmed aging theories rely on the premise that aging is governed by genetic mechanisms orchestrating the evolution of living organisms. However, aging-related changes may also involve additional mechanisms not encoded by the genetic program, which exert non-programmed effects on the organism that also cause life termination. This effect may arise from stochastic cellular damage that alters the cell structure, function, and metabolism (and primarily, the membrane structure). This impact can also affect the genetic information as chromosomal DNA, which is believed to encode the aging program, thus potentially modulating its activation. Furthermore, toxic byproducts and free radicals are generated during normal cellular metabolic processes, their destructive activity being normally counterbalanced by cellular deactivation systems. Failure in these deactivation mechanisms, similar to external impacts, may cause damage to membranes and chromosomal DNA, which will also be accompanied by activation or modulation of aging mechanisms.

Hence, both the approach conceptualizing aging as a genetically encoded evolutionarily programmed process (programmed aging) and the alternative paradigm viewing cellular damage as non-genetically determined yet either modulating or constituting the aging mechanism (non-programmed or nonadaptive aging) share fundamental biological commonality in explaining the causes of involution, senescence, and life termination.

Therefore, when analyzing organismal aging, we deem it more correct to shift focus from philosophical considerations of the reasons behind it to examining functional, cellular, and molecular-level alterations in the organism being indicators of aging (senescence markers). Each of these alterations accompanies aging and is a detectable manifestation of the ongoing involutionary process. We deliberately avoid using the term "aging mechanism", since it means the initiation and development of events that are driven by the fundamental causes of senescence and currently remain beyond our scientific grasp.

There are many reviews classifying prominent detectable senescence markers. These reviews systematize the data characterizing indicators accompanying organism's aging: genomic instability, telomere attrition, epigenetic modifications, loss of proteostasis, impaired macroautophagy, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, altered intercellular communication, chronic inflammation, microbiota imbalance, and disrupted neuroendocrine regulation. Nearly all molecular and biochemical processes that can be classified according to specific feature and are affected by aging processes fall within these fundamental categories (Kenyon, 2010; López-Otín et al., 2013, 2023; Krauss, de Haan, 2016; Proshkina et al., 2020; Zhu et al., 2021; Mikheev et al., 2023).

One of the most remarkable findings of this study is the fact that extracellular double-stranded DNA (dsDNA) fragments delivered to hematopoietic stem cells (HSCs) via a natural biological mechanism induce processes ultimately increasing the telomeric DNA content. Experimental evidence indicates that it occurs in a telomerase-independent manner, presumably via the alternative lengthening of telomeres (ALT) mechanism. The third part of this research cycle (Ruzanova et al., 2025) characterizes the telomere structure and function, as well as the molecular mechanisms responsible for maintaining chromosomal telomere length.

Briefly, telomeres are genomic regions at the ends of linear chromosomes. In vertebrates, telomeric DNA consists of TTAGGG repeats bound by specialized proteins that form telomeric heterochromatin, thus modulating biological functions of telomeres. The terminal portion of single telomere strand, together with capping proteins, forms a complex protecting the recombinogenic structure against recognition by factors initiating the DNA damage repair mechanism. It is commonly believed that because of the conserved replication mechanism, DNA polymerase cannot completely replicate linear DNA templates, leading to progressive telomere shortening (Harley et al., 1990). Critically short telomeres cannot bind a sufficient amount of defense proteins, thus exposing double-stranded ends. This highly recombinogenic structure triggers the DNA damage repair

mechanism, which in turn activates cyclin-dependent kinase inhibitors p21 and p16 and arrests proliferation (Stein et al., 1999). Despite the critical shortening, these telomeres preserve a certain amount of proteins, which prevents fusion of telomeres in different chromosomes but does not halt the activated repair mechanism. Proliferative arrest becomes permanent in this situation. Cellular senescence, the primary trigger of organismal involution and numerous age-related diseases, is initiated and sustained (He, Sharpless, 2017). Telomere-associated repair foci arising from attacks of telomeric replication forks by the activated molecular repair machinery are among the senescence markers.

It was demonstrated that DNA repair foci associated with incurred damage also arise within telomeric heterochromatin at long telomeres in terminally differentiated cells (Di Micco et al., 2021). In this case, the mechanism initiating cellular senescence in non-dividing cells can be conceptualized as follows: in proliferating cells, telomere-binding proteins inhibit DNA repair *in cis*, thereby preventing chromosomal fusion. Consequently, the insensitivity of dividing cells to DNA damage repair within telomeric heterochromatin in shortening telomeres conflicts with the molecular repair machinery and the overall repair process. This conflict induces intercellular propagation of signals from the DNA damage response mechanism and leads to formation of repair foci within telomeric heterochromatin of long telomeres in terminally differentiated cells. Instead of being repaired, the repair foci formed in telomeres of these cells are accumulated and induce an aging-like phenotype. Following this logic, persistent activation of the DNA damage repair mechanism is a common causal event that underlies replicative cellular senescence driven by critically short telomeres as well as aging-like states induced by damaged telomeres in non-dividing cells (Rossiello et al., 2022).

Apoptosis or autophagy is induced in the case of severe telomere dysfunction, being accompanied by all the associated molecular events (Nassour et al., 2019). Senescent cells acquire a senescence-associated secretory phenotype, characterized by secretion of a set of pro-inflammatory cytokines and negatively affecting the extracellular matrix structure and viability of stem cells (Tchkonia et al., 2013).

Hence, the inability to restore the lost telomeric heterochromatin content is among the causal events behind cellular senescence and its propagation throughout the organism. The conflict arising from insensitivity to telomeric DNA damage repair mechanisms further underscores the predominant role of chromatin DNA integrity in initiation of aging. Activation of the DNA damage repair mechanism and accumulation of telomeric repair foci are also presumed to be associated with other aging-related processes such as mitochondrial dysfunction, altered nutrient sensing, impaired autophagy, loss of proteostasis, and epigenetic dysregulation. In this regard, a "telomere-centric" rationale has been proposed for explaining many features of aging (Chakravarti et al., 2021), where telomere shortening being considered a key characteristic. The numerous aging-associated diseases are linked to telomeric heterochromatin attrition (Rossiello et al., 2022).

V.S. Ruzanova et al. (2025) demonstrated that treatment of HSCs within bone marrow cells with fragmented dsDNA ($hDNA^{gr}$) in three model organisms in HSC descendants, where these cells had been naturally internalized, statistically significantly increased telomeric DNA content so that it became comparable to that in young animals. The study also demonstrated that this increase was unrelated to changes in telomerase activity.

Our work, under the "telomere-centric" concept of aging and its propagation to hematopoietic progenitors, proposes an experimental approach to assess the relationship between the lifespan of experimental animals and telomeric DNA content in HSCs and HSC descendants. The experimental design of the study involved reinfusion of a suspension of bone marrow cells treated *ex vivo* with preparations of fragmented genomic DNA ($hDNA^{gr}$), angiogenin, and two inducers administered simultaneously, into aged animals. Recombinant human angiogenin was selected as a comparator because its application had also been shown to increase telomeric DNA content in HSC descendant cells; however, unlike the DNA preparation, this rise was associated with activation of telomerase gene expression (Ruzanova et al., 2025). These treatments were expected to extend the lifespan of the experimental animals.

Materials and methods

Experimental animals. Sexually mature male Wistar rats (weight, 400–450 g) were used for the pilot experiments. At experiment initiation, the rats were aged 13.5 months. Eighteen inbred female mice (weight, 50–60 g) were also included in the study; they were aged 12 months at experiment initiation. The animals were procured from the breeding facility of the Research Center for Biomedical Technologies, Federal Medical and Biological Agency of the Russian Federation, and had valid veterinary certificates. Following a two-week quarantine period, the animals were housed in the vivarium of the A.F. Tsyb Medical Radiological Research Center until they reached the appropriate age for starting the experiment. The animals were housed in polypropylene cages (five animals per cage) and had *ad libitum* access to food and water. The manipulations involving laboratory animals were conducted in accordance with the State Standard GOST 33044-2014 "Principles of Good Laboratory Practice". Euthanasia was performed under ether anesthesia followed by cervical dislocation.

Male CBA/Lac mice aged 14 months and female Wistar rats aged 16 months, bred at the Vivarium of Conventional Animals core facility of the Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia), were used for the repeated experiments. The animals were housed in groups of 6–10 mice and 3–4 rats per cage, with *ad libitum* access

to food and water. All animal experiments were approved by the Animal Care and Use Committee of the Institute of Cytology and Genetics SB RAS. Mice were euthanized by cervical dislocation; rats were euthanized by CO_2 inhalation or decapitation. Bone marrow cells were isolated from 14-month-old and 2-month-old male CBA/Lac mice, as well as 15-month-old male and a 2.5-month-old female Wistar rats.

Isolation of bone marrow cells. For isolating bone marrow, mice were euthanized; femurs and tibias were removed; epiphyses were separated; the medullary cavity was washed with DMEM+2 % FBS. The cell suspension was passed through a 21 gauge needle several times to remove rosette-forming cells and filtered through a 40 μm mesh. The cells were pelleted by 10-min centrifugation at 400g and resuspended in a buffer containing 130 mM ammonium chloride for erythrocyte lysis during 3–5 min. Subsequently, the buffer was diluted tenfold with PBS, and the cells were centrifuged again. The resulting cell pellet was resuspended in DMEM medium, and cells were counted in a Goryaev chamber.

Treatment of bone marrow cells with inducers. In the pilot experiment, bone marrow cells isolated from animals were incubated in the presence of $hDNA^{gr}$ for 1 h at room temperature (20–22 °C). A total of 200 ng of $hDNA^{gr}$ was used for treating 1×10^6 bone marrow cells.

In the repeated experiment, bone marrow cells isolated from animals were incubated in the presence of inducers for 1 h in an atmosphere of 5 % CO_2 (95 % humidity, 37 °C): 500 μg $hDNA^{gr}$, or 500 ng of angiogenin, or 500 μg $hDNA^{gr}$ + 500 ng of angiogenin per 3×10^6 cells in 1 mL of serum-free DMEM.

$hDNA^{gr}$ preparation. Human genome DNA reconstructor ($hDNA^{gr}$) was isolated from placenta of healthy females. DNA was ultrasonically fragmented to 1–20 nucleosome monomers (200–2,000 bp), deproteinated using proteinase K, and isolated by phenol-chloroform extraction.

Angiogenin. Angiogenin was provided by Angiopharm Laboratory LLC (Novosibirsk, Russia). Angiogenin was labeled with Cy5 according to the manufacturer's protocol (Lumiprobe, Germany).

Intravenous administration of a bone marrow cell preparation. In the pilot experiment, the experimental mice and rats received a single dose of 1×10^6 bone marrow cells in 0.3 or 0.5 mL of 0.9 % sodium chloride solution administered into the tail vein. Control animals intravenously received 0.3 and 0.5 mL of 0.9 % sodium chloride solution, respectively.

In the repeated experiment, mice and rats were reinfused a single dose of 1×10^6 bone marrow cells in 0.2 or 0.5 mL of 0.9 % sodium chloride solution into the tail vein. Control animals were reinfused with 1×10^6 untreated bone marrow cells isolated from old and young animals.

Assessment of the effect of bone marrow cell preparation. Animals were examined daily throughout the entire study. General behavior and health status of the animals were monitored. External morphology was assessed and documented through photographs. Animals' body weight was measured. The morphology of spontaneous tumors was analyzed, and natural mortality among the experimental animals was documented.

Postmortem analysis of mouse organs. Organs and tumors were isolated from the animals and fixed in 4 % neutral paraformaldehyde. Samples of organs were dehydrated using increasing concentrations of ethanol, cleared in xylene, and embedded in paraffin. Paraffin sections up to 5 μm thick were stained with hematoxylin and eosin. Visualization and microphotography of the specimens were performed using an Axio Imager Z1 light microscope (Carl Zeiss Microscopy, Germany).

Preparation of blood smears. Blood smears were prepared using blood collected from the tail vein. The smears were fixed in methanol (OJSC Vekton, Russia) for 6–10 min, rinsed with water, dried, and stained using the Romanowsky–Giemsa protocol at pH 7.4. The prepared slides were examined under a Leica DV 4000V microscope (Germany) using transmitted light with immersion oil at a magnification of ×100.

Preparation of bone marrow smears. Bone marrow smears were prepared either by the conventional smear preparation procedure using a small amount of bone marrow cell suspension or by imprinting of animals' thoracic bone sections. Further preparation of bone marrow specimens was conducted using the same procedure as the one employed for blood smear preparation.

Analysis of changes in telomeric DNA content. Quantification of telomeric DNA content was carried out using bone marrow cells isolated from inducer-treated animals 12 months after reinfusion, as well as human bone marrow cells cultured for 15 days on methylcellulose after exposure to inducers. The bone marrow cells were embedded into 1 % low-melting-point agarose blocks (5×10^5 cells per block). Prior to analysis, the agarose blocks were stored in 0.5 M EDTA at 4 °C. Before electrophoresis, the blocks were rinsed in TE buffer and incubated in the presence of a lysis buffer (50 mM EDTA, 1 % sarkosyl (Serva, Germany), and 1 mg/mL proteinase K (Thermo Fisher Scientific, USA)) for 20 min at 50 °C. Next, the low-melting-point agarose blocks were secured in the wells of agarose block and subjected to electrophoretic separation using pulsed-field gel electrophoresis in the following mode: 3 s forward pulse; 1 s reverse pulse; RAM factor, 0.9.

Next, DNA was transferred onto a Hybond N membrane using the capillary transfer method in 20×SSC (Maniatis et al., 1984). DNA samples were UV-annealed to the membrane for 10 min and stored until hybridization.

The membrane with the crosslinked DNA was placed into 50 mL of prehybridization buffer containing 0.1 % SDS, 5×SSC, 5×Denhardt's solution, and 100 μg/mL total yeast RNA, and incubated at 37 °C for 1–3 h. The labeled

DNA specimen (^{32}P-labeled oligonucleotide G-probe – $(TTAGGG)_9$; C-probe – $(CCCTAA)_9$) was denatured by 10-min boiling and added to 50 mL of hybridization buffer containing 0.1 % SDS, 5×SSC, 5 % dextran sulfate 500,000, and 100 μg/mL total yeast RNA. The prehybridization solution was removed, and hybridization buffer containing the labeled probe was added to the membrane after mixing. Hybridization was conducted overnight at 37 °C under permanent stirring. After hybridization, the membrane was washed thrice (15 min each washing cycle) with a solution containing 0.1 % SDS and 0.1×SSC at 37 °C. The hybridization conditions (buffer system, temperature, and number of washing cycles) for short oligonucleotides were optimized empirically based on multiple experiments with radioactive phosphorus; the temperature typically ranged from 37 to 42 °C (Dolgova et al., 2012).

The membrane with the transferred specimens was exposed to a K-type screen. Radioisotope specimens were scanned using a PharosFX system. Images were analyzed using the Quantity One software by measuring spot density (intensity/mm^2) or using the GEL-Pro software.

Statistical analysis was conducted using the Statistica 8 (StatSoft, USA) and GraphPad Prism 8.0.1 software (GraphPad, USA). Survival analysis was performed by constructing Kaplan–Meier curves using the log-rank (Mantel–Cox) test. The significance of differences was assessed using the Mann–Whitney U-test. The observed differences were considered statistically significant at $p < 0.05$.

Results

The effect of reinfusion of hDNA^{gr}-treated bone marrow cells (HSCs) on the lifespan and general condition of experimental animals

Using the model characterizing changes in telomeric DNA content and the FISH data, we demonstrated that genetic information contained in extracellular DNA fragments can be incorporated into the recipient genome, in either an integrated, in complex with chromatin or a circular form. The findings attest to an increase in telomeric DNA content and the emergence of numerous sites on chromosomes hybridizing with DNA material that originally had an extrachromosomal localization (as evidenced experimentally they can be represented by circular structures encompassing chromosomal DNA strand or they form a complex with chromatin and coexisting in these forms for a certain period (Ruzanova et al., 2025)).

The telomeric DNA content is known to be a biomarker of lifespan (Rossiello et al., 2022). A series of experiments evaluated the lifespan of experimental mice and rats. Xenogeneic human DNA was employed, which was consistently increasing telomeric DNA content in mouse and rat cells during cloning experiments, suggesting that there were conditions conducive to lifespan extension in these animals. However, we were acutely aware that, along with telomere elongation, there can occur an uncontrolled DNA-level interplay between human DNA fragments and rodent chromosomes, potentially adversely affecting condition of the experimental animals. Moreover, our studies, including the present work, revealed that only approximately 1 % of the genome can be delivered into the cell. It implies that internalization of this DNA will be highly degenerate. Different genomic DNA will enter different cells, further exacerbating the potential sequelae of the uncontrolled interplay between xenogeneic extracellular human DNA and rodent chromosomal DNA. Collectively, these considerations suggest that such treatment is inherently unpredictable and may result in either the anticipated extension of animals' lifespan or an opposite unfavorable outcome.

The effect of reinfusion of inducer-treated bone marrow cells (HSCs) on the lifespan and general condition of mice

For conducting the pilot study of the effect of reinfusion of hDNA^{gr}-treated bone marrow cells (HSCs) on the lifespan of animals, we selected nine outbred female mice of identical weight, without manifestations of spontaneous subcutaneous tumors. Five mice in the control group were intravenously administered with 0.9 % sodium chloride solution; the study group comprising four mice were infused with hDNA^{gr}-treated bone marrow cells. The age of mice at experiment initiation was 12 months. The mean body weight of the experimental and control mice was 42.1 ± 2.4 g.

Figure 1*A* shows the dynamics of death among experimental animals. The mean lifespan of mice was approximately 16 months: 487 ± 30 days in the control group and 490 ± 31 days in the experimental group. Two months after experiment initiation, body weight gain was observed for mice in the experimental group, whereas control mice experienced a mean weight loss of ~ 20 % (Fig. 1*B*). The physical condition of mice in the experimental group significantly differed from that in the control group (Fig. 1*C*). Control mice were less active during the terminal stage of life, had a ruffled hair coat and partial absence of undercoat hair. In contrast, mice reinfused with hDNA^{gr}-treated bone marrow cells maintained activity and normal feeding behavior, they had a smooth hair coat with well-defined underfur hair.

In the experimental group of mice, one animal aged 16 months was found to have a neoplasm: a spontaneously developing tumor nodule in the pelvic region (Fig. 1*D*). The tumor was rapidly growing and, two weeks after its detection, reached a volume of 7.5 cm^3. The tumor-bearing animal was withdrawn from the experiment for macroscopic and histopathological examination.

The conducted histopathological analysis revealed that the mouse tumor parenchyma consisted of cuboidal epithelial cells with basophilic cytoplasm, which contained moderately polymorphic nuclei occupying approximately half of the cells and well-defined nucleoli. Glandular epithelium tending to form acinar or papillary structures was detected in some areas (Fig. 1*E*1). Mitotic figures and

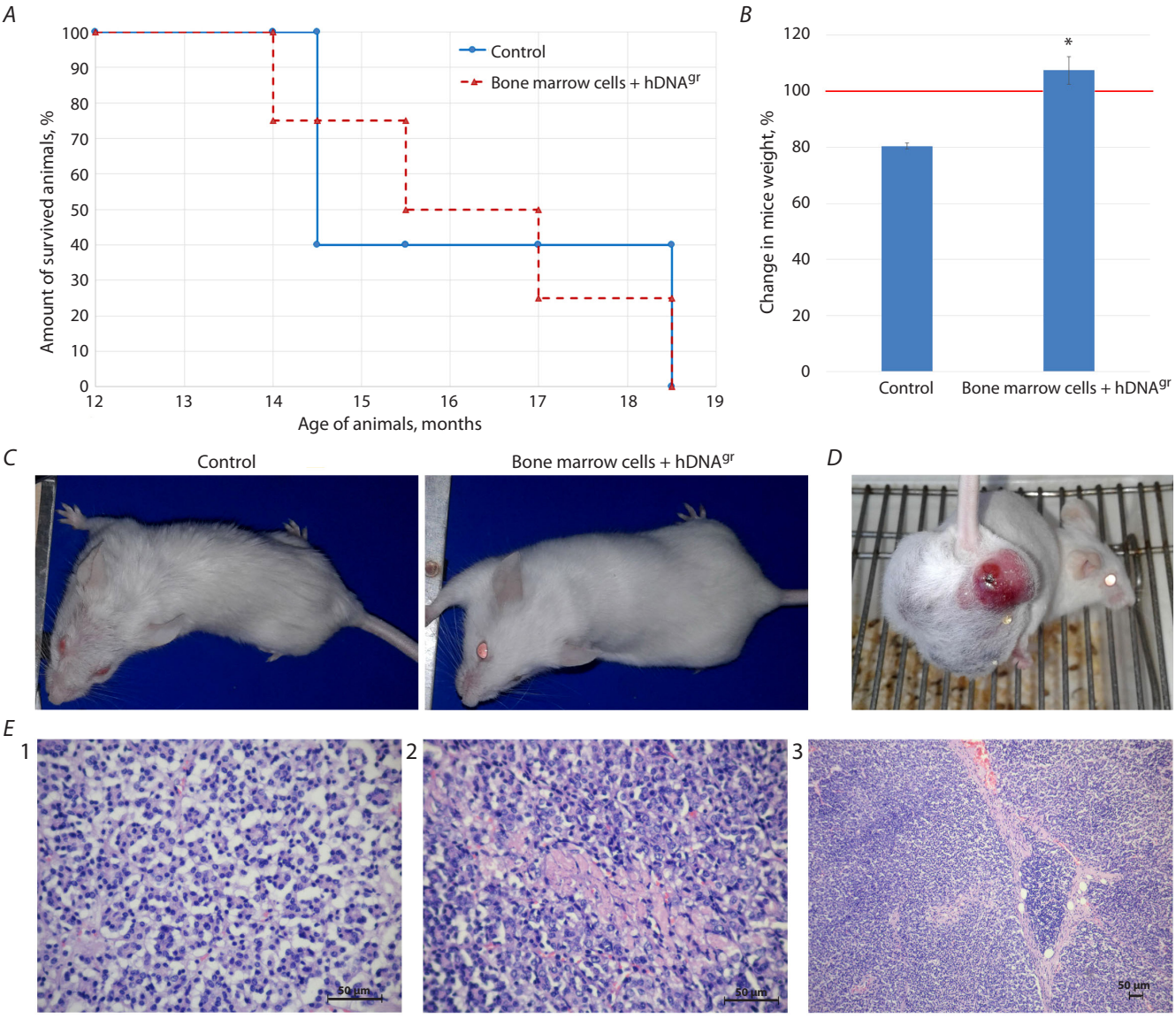


Fig. 1. The effect of hDNA^{gr} on the lifespan of mice.

A – the survival rate of animals; *B* – changes in body weight in 14-month mice in the control and experimental groups compared to the baseline weight at an age of 12 months at experiment initiation point taken as 100 % (shown with a red line). * Statistically significant differences compared to the control group, $p < 0.05$; Mann–Whitney U-test. *C* – comparative external appearance of animals in the control and experimental groups at an age of 16 months; *D* – external appearance of a tumor-bearing mouse in the experimental group at an age of 16 months; *E* – histopathological analysis of the tumor in a mouse in the experimental group at an age of 16 months: 1 – the histostructure of mouse mammary carcinoma. ×40 magnification. 2 – micronecrosis within the carcinoma parenchyma. ×40 magnification. 3 – connective tissue septum within the carcinoma parenchyma. ×10 magnification. Hematoxylin and eosin staining.

apoptotic bodies were rare. Small number of micronecrotic regions was revealed (Fig. 1*E*2). Chaotically distributed epithelial cells were found; a subtle lobular pattern of tissue was observed in some regions, being confined by thin connective tissue septa containing small blood vessels, with erythrocytes present in them. The connective tissue septa were thickened in some places to acquire a protruding structure (Fig. 1*E*3). These findings suggested spontaneous development of mammary carcinoma.

CBA/Lac mice were chosen to conduct the repeated experiment in mice. The animals were divided into five groups: (1) those reinfused with bone marrow cells from old animals; (2) those reinfused with bone marrow cells from young animals; (3) those reinfused with bone marrow cells from old animals treated with angiogenin; (4) those reinfused with bone marrow cells from old animals treated with hDNA^{gr}; and (5) those reinfused with bone marrow cells from old animals treated with angiogenin + hDNA^{gr}. Figure 2 shows an analysis of the lifespan of mice.

It was demonstrated that reinfusion of HSCs treated with hDNA^{gr} had no effect on the lifespan of mice compared to those reinfused with bone marrow cells from old animals

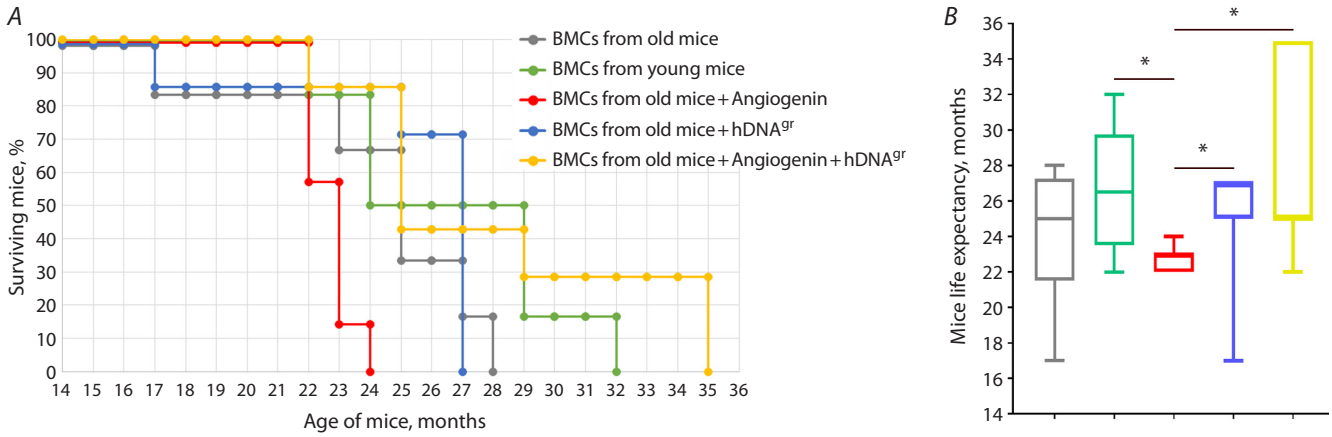


Fig. 2. Analysis of lifespan of experimental mice in groups.

A – the Kaplan–Meier curve; *B* – lifespan of mice. * Statistically significant intergroup differences in the lifespan of mice, $p < 0.05$, log-rank (Mantel–Cox) test.

(Fig. 2*B*). Like in the first experiment, this finding suggests that treatment with hDNA^{gr} has no effect on the lifespan of mice. Mice reinfused with angiogenin-treated bone marrow cells from old animals rapidly developed a symptom cluster eight months after treatment, resulting in weight loss (down to 17 g) and subsequent death within a short period (1 month) (Fig. 2).

The longest lifespan was observed in the groups of mice reinfused with bone marrow cells from young animals and from old animals treated with angiogenin + hDNA^{gr}, with one and two mice surviving up to 32 and 35 months, respectively (17 and 30 % of the total number of mice). Four pivotal results of the study using the mouse model have been obtained.

1. Treatment of bone marrow cells with xenogeneic dsDNA (hDNA^{gr}) has no effect on the lifespan of mice. We attribute the reasons for that to the complex pleiotropic effects of fragments of exogenous heterologous dsDNA after its internalization into hematopoietic stem cells (HSCs) on the "fundamental metabolic constants responsible for aging of an organism". Another fact important to consider is that a hematopoietic stem cell contains only 0.1–1.0 % of extracellular DNA, which constitutes a negligible portion of the genome. It means there is always a probability that telomeric repeat DNA is not delivered into the cell or its amount is insufficient for amplifying site-specific integration.

2. Reinfusion of bone marrow cells from mouse pups to old mice significantly increases the lifespan of mice in the group (up to 32 months).

3. Simultaneous administration of angiogenin and hDNA^{gr}, internalized by hematopoietic stem cells (HSCs) (either by the same cell, or by different cells, or in a mixed manner), significantly extends the lifespan of mice (up to 35 months), outperforming the results obtained for mice reinfused with bone marrow cells from mouse pups. It indicates that two independent HSC activators, belonging to two different classes of polymers, have a synergistic effect and favorably influence the "fundamental metabolic constants" of HSCs responsible for cell aging, being accompanied by extension of the lifespan of experimental mice. The observed phenomena need to be further studied experimentally.

4. Angiogenin administered in the monotherapy mode induced a specific symptom cluster in mice, being the reason for death of mice in this group. Histopathological analysis was conducted to understand what events had occurred in the bodies of mice in response to angiogenin processed cells infusion (Supplementary Materials 1 and 2)[1]. One mouse had no pathological changes in the examined tissues and organs. In two other animals, areas of atypical tissue in the liver and kidneys were identified. The key pathological changes were most prominent in the liver and kidneys. Hepatocyte dystrophy was observed in the liver, being accompanied by polymorphonuclear infiltration, corresponding to the morphologic pattern of acute hepatitis. Significant dystrophic changes in the tubular and glomerular epithelium were observed in the kidneys, corresponding to acute kidney injury. Both types of destructive pathology affect vital excretory systems and could have been responsible for animal death related to angiogenin exposure. However, histopathological analysis of an animal reinfused with bone marrow cells from old animals showed nearly identical results. Atypical tissue was also detected in its liver and kidneys. Parenchymal inflammatory changes were detected in the liver, similar to those observed in angiogenin-treated mice. The kidneys also had dystrophic changes in tubules and glomeruli. This fact suggests that all these changes may result from aging of the animals. Similar changes were observed in the liver and kidney parenchyma in two other groups of animals (those reinfused with bone marrow cells from young animals and hDNA^{gr}-treated cells), but there

[1] Supplementary Materials 1–3 are available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Ruz_Engl_30_3.pdf

were no signs of inflammation. This fact also suggests that both the detected pathological destruction of the liver and kidneys and other unidentified factors are the reasons behind death following exposure to angiogenin (Supplementary Material 1).

Simultaneously to histopathological analysis of mouse tissues and organs, we analyzed blood and bone marrow of animals withdrawn from the experiment at the premortem stage. Bone marrow cells were normal in all the experimental groups. Elevated lymphocyte counts in blood were observed in the groups of mice reinfused with bone marrow cells from old animals and angiogenin-treated cells. In the group of animals reinfused with bone marrow cells treated with angiogenin + hDNA^{gr}, erythrocytes had a pathological morphology (burr cells), being indicative of functional disturbances leading to pathological alteration of erythrocyte shape (Supplementary Material 2). However, these alterations did not affect the lifespan of animals in this group.

The effect of reinfusion of inducer-treated bone marrow cells (HSCs) on the lifespan of Wistar rats

A total of 35 animals were used in the first pilot experiment (JSC Clinical Hospital "NeuroVita", Moscow, Russia) to study the effect of hDNA^{gr} on the lifespan of female Wistar rats. The animals were divided into three groups: bone marrow donors; control group consisting of 10 animals that received an intravenous infusion of 0.9 % sodium chloride solution; and the experimental group consisting of 19 animals that were reinfused with hDNA^{gr}-treated bone marrow cells. At the time of experiment initiation, the rats were aged 13 months. The mean body weight of experimental and control rats was 460.0 ± 18.2 g.

The survival dynamics of the experimental animals are shown in Figure 3*A*. The mean lifespan of the Wistar rat population in the control group was 717 ± 38 days or 24 months; in the group of animals that were reinfused with hDNA^{gr}-treated bone marrow cells, it was 842 ± 25 days or 28 months, this difference being statistically significant ($p < 0.01$) (Fig. 3*B*). At an age of 24 months, the mean body weight of rats in the experimental group was significantly higher than that in the control group ($p < 0.05$) (Fig. 3*C*).

Monitoring the physical condition of animals throughout the entire experiment revealed that already by the age of 18 months, the experimental rats significantly differed from those in the control group: they were characterized by better skin and fur condition and increased motor activity (Fig. 3*D*). Feeding activity in the experimental rats was higher than that in control animals. By the age of 21 months (eight months post-treatment), noticeable changes in physical condition became evident. Control rats were characterized by sparse fur, impaired grooming, and decreased muscle tone. In contrast, experimental rats at this stage were more active, displayed an exploratory behavior, maintained muscle tone, and had normal hair and skin. At an age of 25 months (12 months post-treatment), experimental rats retained mobility and exploratory behavior; muscle tone was slightly reduced, and movement coordination was undisturbed. Slight underfur hair loss was observed. Skin was clean; the rats exhibited no pathologic grooming behavior. Control rats at this age were less mobile, had reduced muscle tone, and partially impaired movement coordination. Significant hair loss and disturbances in grooming behavior were observed. A control rat that had reached the age of 27 months was characterized by decreased body weight, sparse fur, low activity, and impaired movement coordination. In contrast, 11 rats from the experimental group survived to this age. They had a slightly reduced body weight and sparse fur. However, the rats remained active, with preserved movement coordination, and exhibited an exploratory behavior.

We conducted a histopathological examination of internal organ tissues harvested from two rats following therapy with hDNA^{gr}-treated bone marrow cells, which had been euthanized at an age of 30 and 32 months. A spontaneous neoplasm was detected in the abdominal cavity of the 30-month-old rat (Fig. 3*E*). Histopathological analysis of the upper pole of the right kidney revealed schwannoma, a benign tumor of neural origin, which appeared sporadically (Fig. 3*E*, *F*). The second rat in this group was euthanized at an age of 32 months. Histopathological analysis of the heart, lungs, liver, kidneys, spleen, brain, spinal cord, and red bone marrow revealed no signs of pathology.

A repeated experiment conducted using rats of the same line (Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia) involved comparative assessment to evaluate the effects of reinfusing bone marrow cells (HSCs) from old animals, treated with angiogenin, hDNA^{gr}, and angiogenin + hDNA^{gr}. Control groups included rats treated with bone marrow cells from young and old animals (Fig. 4). In this study, we performed Southern blot analysis of changes in telomeric DNA content in the bone marrow 12 months after reinfusion of bone marrow cells from old and young rats and bone marrow cells treated with angiogenin, hDNA^{gr}, and angiogenin + hDNA^{gr}. Additionally, in this part of the study, we analyzed using rat and human models which telomeric strand (the G or C one) was amplified.

The following conclusions were drawn from the experimental results: (1) Reinfusion of bone marrow cells from aged animals treated with hDNA^{gr} had no effect on lifespan extension compared to the control group (untreated bone marrow cells from aged animals) (Fig. 4). The average lifespan was approximately 25 months. (2) Reinfusion of bone marrow cells from young animals, as well as cells treated with angiogenin and angiogenin + hDNA^{gr} increased the lifespan of rats in these groups (Fig. 4*A*, *B*). In all the experimental groups, 40–60 % of animals developed spontaneous tumors (Fig. 4*C*). In control groups, this parameter was 20–40 %. The higher rate of tumor occurrence in groups of animals reinfused with cells treated with angiogenin and

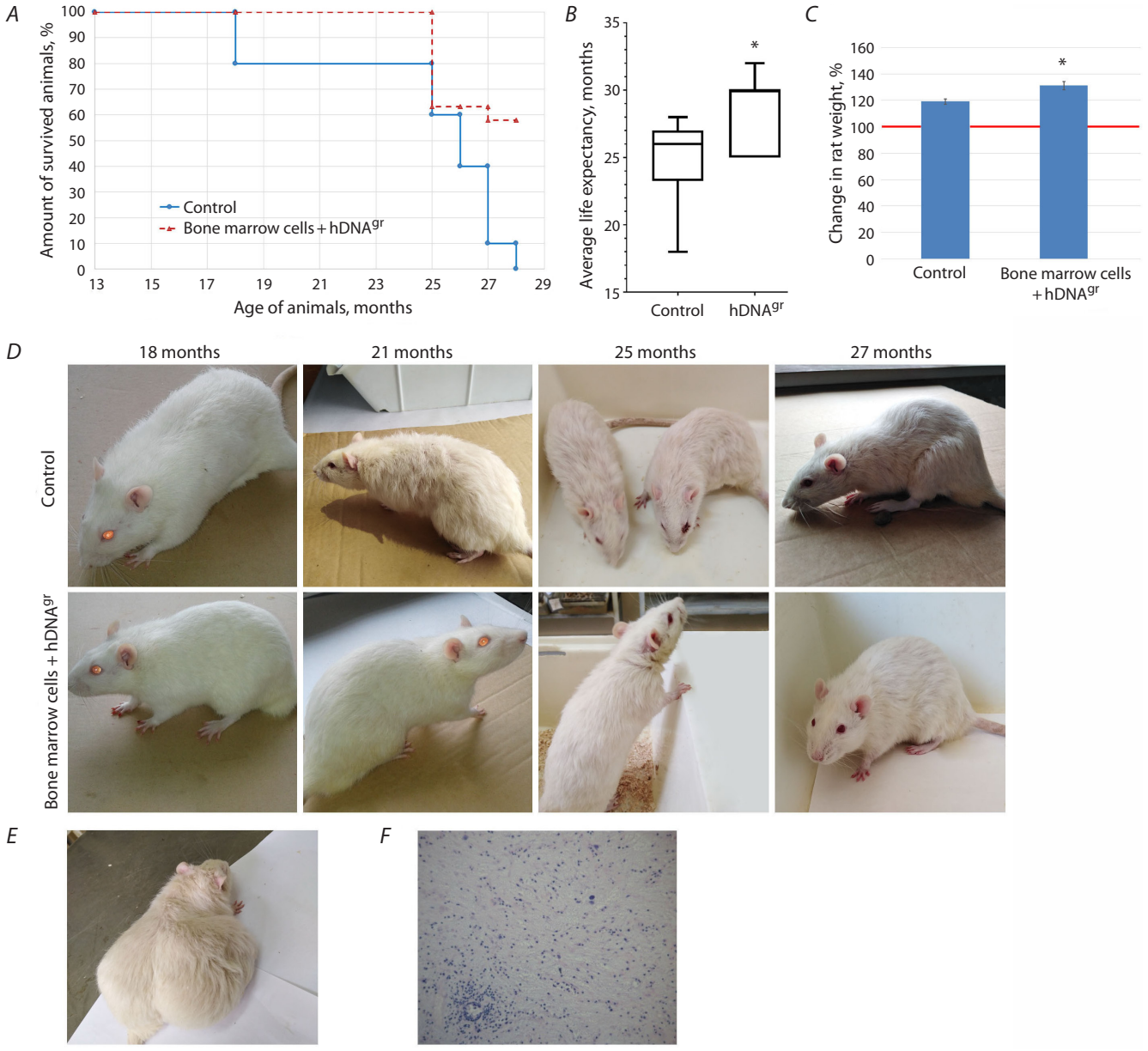


Fig. 3. The effect of DNA^{gr} on lifespan of Wistar rats.

A – the survival rate of animals; *B* – average lifespan of the animals. * Statistically significant differences compared to the control group, $p < 0.05$; log-rank (Mantel–Cox) test. *C* – changes in the body weight of rats in the control and experimental groups aged 24 months compared to the baseline weight at an age of 13 months at experiment initiation point taken as 100 % (shown with a red line). * Statistically significant differences compared to the control group, $p < 0.05$; Mann–Whitney U-test. *D* – comparative external appearance of animals in the control and experimental groups at different age; *E* – external appearance of a tumor-bearing rat in the experimental group at an age of 30 months; *F* – histopathological analysis of the tumor in a rat in the experimental group at an age of 30 months. Hematoxylin and eosin staining. ×20 magnification.

angiogenin + hDNA^{gr} can be attributed to the increased lifespan, which allowed more animals to develop neoplasms. Supplementary Material 3 summarizes the results of histopathological examinations of tumors in rats from different experimental groups. The findings indicate that rats in all the groups developed either spontaneous solid epithelial tumors, or squamous cell carcinomas, or glandular tumors that were identified as mammary gland tumors in some cases. All the tumors exhibited a low proliferative activity. We hypothesize that the developed neoplasms are related to animal housing rather than to tumorigenesis induced by hematopoietic stem cells. HSCs can differentiate into blood cells, but not into epithelial cells, which have been spontaneously transformed to tumors. Nevertheless, those are alarming findings requiring meticulous verification.

Furthermore, a high rate of pneumonia cases related to housing conditions was observed, which generally weakens the interpretative value of the results. Still, the findings from these pioneering studies provide a more meaningful basis for outlining the next level of research objectives.

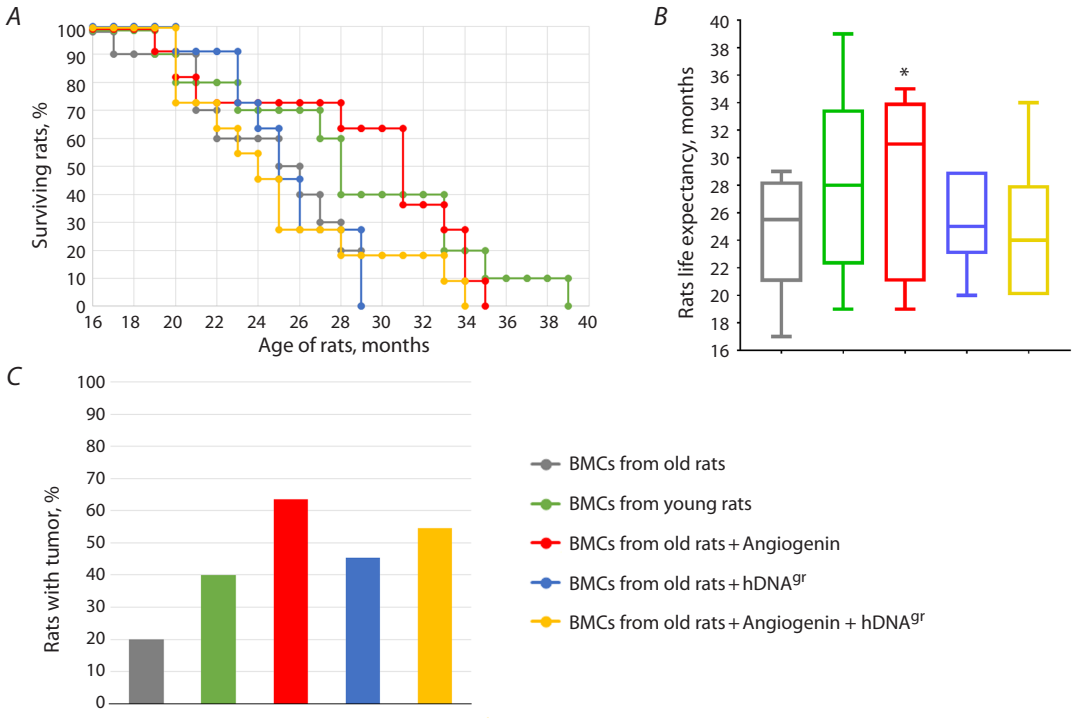


Fig. 4. Analysis of the lifespan of experimental rats in study group.

A – the Kaplan–Meier curve; *B* – lifespan of rats; *C* – the number of rats with tumors. * Statistically significant differences in the lifespan of rats compared to the control group, which received bone marrow cells from aged animals, $p < 0.01$, log-rank (Mantel-Cox) test.

Analysis of certain details of the telomere structure and comparative analysis of telomeric DNA content in bone marrow cells of experimental rats and in the specimen of human bone marrow cells

Certain structural details of telomeres in the bone marrow cells of experimental rats and in a specimen of human bone marrow cells were analyzed using specific probes labeled with ^{32}P targeting the G- and C-strand telomeric "tails". Additionally, a comparative analysis was performed for: (1) changes in telomeric DNA content in the bone marrow cells of rats reinfused with untreated bone marrow cells from old animals (control) and cells from old animals treated with hDNA^{gr}, angiogenin, and angiogenin + hDNA^{gr}, 12 months post-treatment; (2) changes in telomeric DNA content in human bone marrow cells in control samples and hDNA^{gr}-treated samples after 15-day culturing on methylcellulose.

Two independent experiments were performed to assess the telomeric DNA content in cell samples derived from rats reinfused with bone marrow cells from old rats treated with hDNA^{gr} (the first experiment) as well as angiogenin and angiogenin + hDNA^{gr} (the second experiment). Human bone marrow cells were also analyzed in the first experiment under methodological conditions identical to those used in the rat experiment.

Analysis of telomeric DNA content. Analysis of telomeric DNA content in bone marrow cells from control group rats and rats reinfused with hDNA^{gr}-treated bone marrow cells, 12 months post-treatment, and in specimens of human bone marrow cells (control and hDNA^{gr}-treated ones) on day 15 of cell culturing on methylcellulose.

Literature data demonstrate that telomeres have a long G tail. It means that hybridization with probes targeting different strands would generate a stronger hybridization signal when using C-strand probes.

The results obtained for rats fully confirmed the available data (Fig. 5*A–D*). However, the efficiency of hybridization using different probes (G/C, with ^{32}P-specific activity, and number of DNA hybridization probes being identical) did not differ significantly for human bone marrow cells. It implies that in the analyzed system, both strands are approximately of the same length. It also indicated that terminal reduction of telomeric heterochromatin had occurred, and telomeres had reached a critical length. We believe that it is most likely to be related to the patient's disease (the cryopreserved bone marrow cells used had been harvested from a 59-year-old patient with multiple myeloma) (Fig. 5*E–H*).

In the same experiment, we compared the changes in telomeric DNA content between control and experimental cell samples for both models (Fig. 5*D*, *H*).

In rats reinfused with hDNA^{gr}-treated bone marrow cells, no increase in telomeric DNA content (and telomere length) was observed 12 months post-reinfusion. These findings are consistent with the average lifespan data, which also showed no significant differences between the analyzed groups. They suggest that either hDNA^{gr} did not induce telomere elongation in bone marrow cells in this particular experiment or such cells were eliminated during

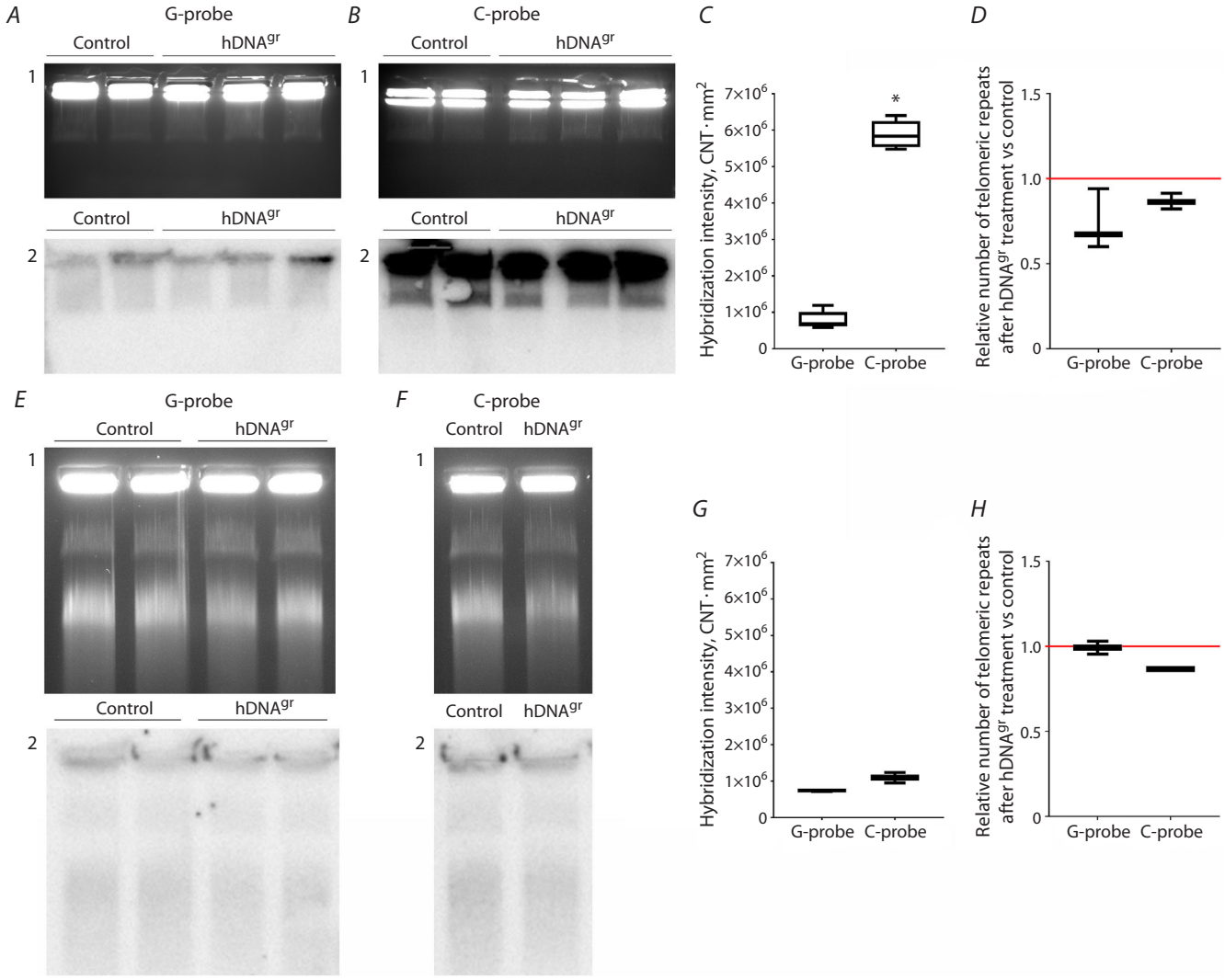


Fig. 5. Analysis of certain details of the telomeric structure in bone marrow cells of experimental rats and in human bone marrow cell specimens: *A–D* – the rat model; *E–H* – the human model.

A, *B*, *E*, and *F* – gel electrophoresis (1) and Southern blot analysis (2) of DNA samples after electrophoretic separation of lysed and deproteinized bone marrow cells embedded in low-melting-point agarose blocks. Hybridization was performed using ^{32}P-labeled complementary oligonucleotides containing nine telomeric hexanucleotide repeats corresponding to G- and C-strand telomeric sequences (*A*, *E* – G probe; *B*, *F* – C probe); *C*, *G* – assessment of the total telomeric repeat content in the G- and C-strands in bone marrow cells in the control and experimental groups. The (CNT*mm^2) values were compared using the GEL-Pro software. * Significant differences in hybridization intensity with the C probe with respect to hybridization intensity with the G probe, $p < 0.01$, Mann–Whitney U-test. *D*, *H* – estimation of telomeric repeat content in hDNA^{gr}-treated samples with respect to control ones. Relative values obtained by dividing the luminescence intensity of radioactive signals by the luminescence intensity of ethidium bromide for each lane, expressed in arbitrary units, were compared using the GEL-Pro software.

the 12-month follow-up period. Unfortunately, in the first pilot experiment, quantitative analysis using a radioactively labeled probe could not be performed because of some technical issues.

A similar result (no rise in telomeric DNA content) was obtained for the human model (Fig. 5*H*). We propose the following explanation in this case. For the control samples, amplification via the alternative lengthening of telomeres (ALT) mechanism requires freely accessible telomeric DNA rings to be present. These rings (t-circles) are formed after the detachment of a t-loop residing at the end of the long G tail in the telomere. According to the hybridization data, this long telomeric strand is absent in bone marrow cells derived from the analyzed sample, and t-circle formation is limited. This limitation explains why no difference in telomeric DNA content was observed in this particular experiment.

As mentioned previously, in the case of hDNA^{gr} inducer, alternative lengthening of telomeres (ALT) is the main mechanism for telomere lengthening; telomeric DNA is amplified on t-circles formed by extracellular fragments via the rolling-circle mechanism. We believe that like for the mouse model, the lack of increase in telomeric DNA content in these particular experiments is related to ambiguity in the internalization process. It was demonstrated that only 0.1–1.0 % of extracellular DNA, a negligible portion

of the genome, is internalized by a hematopoietic stem cell. It means that there will always be a risk that telomeric repeat DNA is not delivered into the cell or its amount is insufficient for amplification or site-specific integration. Some other explanations for this phenomenon, related to the biology of HSCs, are also possible; so further meticulous studies are needed (e. g., competitive elimination of a modified stem cells from the bone marrow by clones that have acquired clonal characteristics during age-related alterations).

Hence, in the first pilot experiment, hDNA^{gr}-treated bone marrow cells reinfused to experimental animals had a significant effect on their quality of life during the terminal period, as well as increased their average lifespan. Evidence supporting the validity of the proposed concept was obtained in the experiment. Both anticipated biological effects were observed: a positive trend in the overall condition of aged animals throughout their remaining lifespan and a statistically significant extension of the lifespan of experimental rats.

In the repeated experiment, lifespan extension was observed in neither the mouse model nor in rats in the groups of hDNA^{gr}-treated bone marrow cells. Moreover, molecular analysis revealed no changes in telomeric DNA content. In other words, these two parameters can be considered correlated.

As mentioned at the beginning of the article, the differences in the endpoints of two similar experiments can be attributed to the ambiguous behavior of the analyzed HSC system *vs* hDNA^{gr} because of the potential uncontrolled interplay between xenogeneic extracellular human DNA and rodent chromosomal DNA.

In the second series of hybridization experiments, we compared the telomeric DNA contents in the cells of control rats and rats reinfused with bone marrow cells derived from old rats and treated with angiogenin alone and with angiogenin + hDNA^{gr} 12 months post-treatment (Fig. 6). This analysis was needed because of the findings indicating that the lifespan of rats after treatment with angiogenin alone and in combination with hDNA^{gr} was significantly longer than that of both control animals and animals treated with hDNA^{gr} alone.

A comparative analysis of luminescence intensity of the signals from ethidium bromide and radioactive labeling was conducted. Importantly, all the electrophoresis and hybridization conditions were identical in the experiment, making certain quantification possible. In all the samples, the G-strand telomeric repeats (C probe) were shown to dominate over the C-strand repeats (G probe) (the length of DNA probe and specific activity of both probes being virtually identical). This finding is consistent with the available literature data (Fig. 6*A*2, *C*).

Previous hybridization experiment demonstrated that treatment of cells from old rats with hDNA^{gr} had no effect on the increase in telomeric DNA content in the bone marrow cells of experimental animals, and the average lifespan in this group did not differ from that in the control. Therefore, we have put forward a hypothesis that after treatment with angiogenin + hDNA^{gr}, the main effect is caused by angiogenin per se rather than by hDNA^{gr}, so the data for the "Angiogenin" and "Angiogenin + hDNA^{gr}" groups can be pooled. The analysis revealed that the relative content of telomeric repeats for C- and G-strands in total DNA in rat groups did not significantly change compared to control. However, the content of G-strand repeats was significantly higher than that of C-strand repeats (Fig. 6*C*). The telomeric DNA hybridization data indicate that the substantial rise in the lifespan of rats following treatment with angiogenin alone or in combination with hDNA^{gr} is not associated with an increase in telomeric DNA content (number of telomeric repeats), suggesting either that angiogenin does not activate endogenous telomerase activity (Ruzanova et al., 2025) or that other mechanisms are involved (e. g., competitive elimination of reinfused modified hematopoietic clones by bone marrow progenitors with clonal characteristics).

Discussion

The efforts to extend longevity have been accompanying humanity throughout the entire history of civilization. Currently, there is a well-established understanding of biomarkers characteristic of aging and inherent to aged organisms, along with various anti-aging interventions being implemented (López-Otín et al., 2013, 2023; Proshkina et al., 2020; Zhu et al., 2021). Nonetheless, modern anti-aging approaches and diverse medicines developed relying on the insights into molecular processes occurring within cells and organisms, rarely enable outliving the coveted 100-year limit. In our view, it stems from the fact that aging is an involutionary integrative state of the organism with unknown causes of degradation rather than being merely a response of a specific functional system to lifespan that can be "corrected" using therapeutic procedures or interventions. In this context, maintaining a healthy lifestyle is at least half of the success in achieving a long life. The genetic factor also plays a predominant role in the development of anti-aging strategies, and current scientific efforts in gerontology primarily pursuit this aspect. These therapies include cellular reprogramming using Yamanaka factors (Takahashi, Yamanaka, 2006; Ocampo et al., 2016; Gowing et al., 2017; Brooks, Robbins, 2018; Sogabe et al., 2018), utilization of microRNAs that target multiple genes within networks regulating fundamental aging pathways (Vaiserman et al., 2016), combinatorial gene therapy (*FGF21+αKloho+sTGFβR2*) (Davidsohn et al., 2019), and telomere length extension therapies, primarily through activation of endogenous telomerase or transduction of the telomerase gene (Aubert, Lansdorp, 2008; Bernardes de Jesus et al., 2012; Li et al., 2017; Hong, Yun, 2019).

As previously mentioned, telomere length is a fundamental factor determining cellular aging and the development of many diseases of civilization, and in particular when telomere shortening events occur in stem cells (Rossiello et al.,

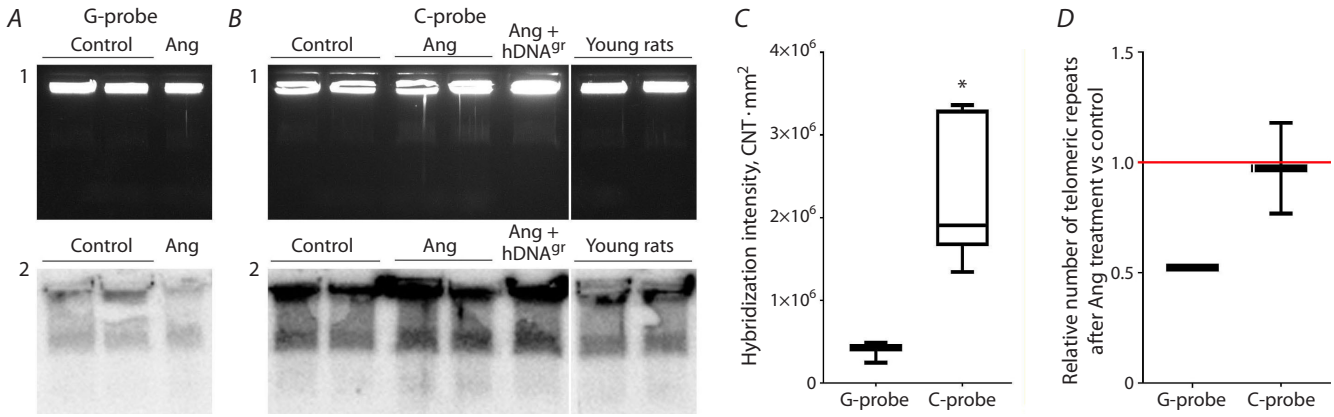


Fig. 6. Analysis of certain details of the telomeric structure in bone marrow cells of experimental rats 12 months post-treatment.

A, *B* – gel electrophoresis (1) and Southern blot analysis (2) of DNA samples after electrophoretic separation of lysed and deproteinized bone marrow cells embedded in low-melting-point agarose blocks. Hybridization was performed using ^{32}P-labeled complementary oligonucleotides containing nine telomeric hexanucleotide repeats corresponding to G- and C-strand telomeric sequences (*A* – G probe; *B* – C probe)). *C* – assessment of the difference in the total content of G- and C-strand telomeric repeats in bone marrow cells in the control and experimental groups. The (CNT · mm^2) values were compared using the GEL-Pro software. * Significant differences in hybridization intensity with the C probe with respect to hybridization intensity with the G probe, $p < 0.01$, Mann–Whitney U-test. *D* – estimation of the relative number of telomeric repeats in angiogenin-treated samples with respect to control ones. Relative values obtained by dividing the luminescence intensity of radioactive signals by the luminescence intensity of ethidium bromide for each lane, expressed in arbitrary units, were compared using the GEL-Pro software.

2022). In this context, understanding approaches enabling telomere length extension with systemic organism-wide effects, rather than solely in cultured cells, poses a non-trivial challenge for both biology and clinical medicine. Almost all the identified telomere elongation methods involve targeting the telomerase complex and its regulatory genes (Aubert, Lansdorp, 2008). We failed to identify clinical strategies capable of increasing telomere length through an equally important alternative mechanism, alternative lengthening of telomeres (Lundblad, 2002; Hande, 2004; Pickett et al., 2009; Nabetani, Ishikawa, 2011; Rovatsos et al., 2011; Doksani, 2019; Loe et al., 2020).

Our study assessed the feasibility of extending the lifespan using the technology embedded in the novel concept of natural genome reconstruction, which is associated with the potential for *in vivo* telomere lengthening in hematopoietic stem cells.

Animal studies indicate that reinfusion of bone marrow cells treated with inducers is relatively safe. No prominent immediate pathological changes were observed in experimental animals across both models.

The findings regarding the lifespan of rats reinfused with bone marrow cells (HSCs) activated using fragmented human DNA suggest that there is a fundamental feasibility of extending the lifespan of aged animals. This effect is plastic: a statistically significant increase in lifespan was observed in some cases, whereas in other cases treatment had no effect on the lifespan of experimental animals.

In the cases when no effect was observed, the detectable telomeric DNA content in rat bone marrow cells 12 months post-treatment did not differ from that in the control samples. Clonal competition can be a putative mechanism explaining the absence of increased telomeric DNA content in bone marrow cells one year after treatment with hDNA^{gr}. In this scenario, the rejuvenated clone with increased telomeric DNA content emerging after treatment as demonstrated in ref. (Ruzanova et al., 2025) could be outcompeted by dominant bone marrow residing clones that have acquired clonal characteristics. If this hypothesis is confirmed, it would be a crucial observation indicating that multiple initial interventions using modified bone marrow cells are needed to counteract the competitive expansion of dominant clones and achieve stable fixation of a trait in the bone marrow. Preliminary myeloablation is also possible, allowing reconstructed HSCs to be fixed in the vacated stromal niches.

Recombinant human angiogenin may exhibit systemic toxicity (a mouse model), which is mitigated when combining it with hDNA^{gr}. The synergistic use of angiogenin and hDNA^{gr} increases the lifespan in the group of mice.

The observed lifespan extension in rats after treatment with angiogenin is unrelated to telomere elongation, being determined by other properties of the factor, or clonal competition in the bone marrow, or a combination of both mechanisms.

The concept of the "programmed clonal hematopoiesis" and "programmed death" mechanisms

In their study, E.N. Proshkina et al. (2020) suggested aging to be a quasi-program, where the regulatory elements are not originally intended for its execution but are responsible for other processes such as cell growth, regeneration, stress response, and immunological protection. Therefore, it is

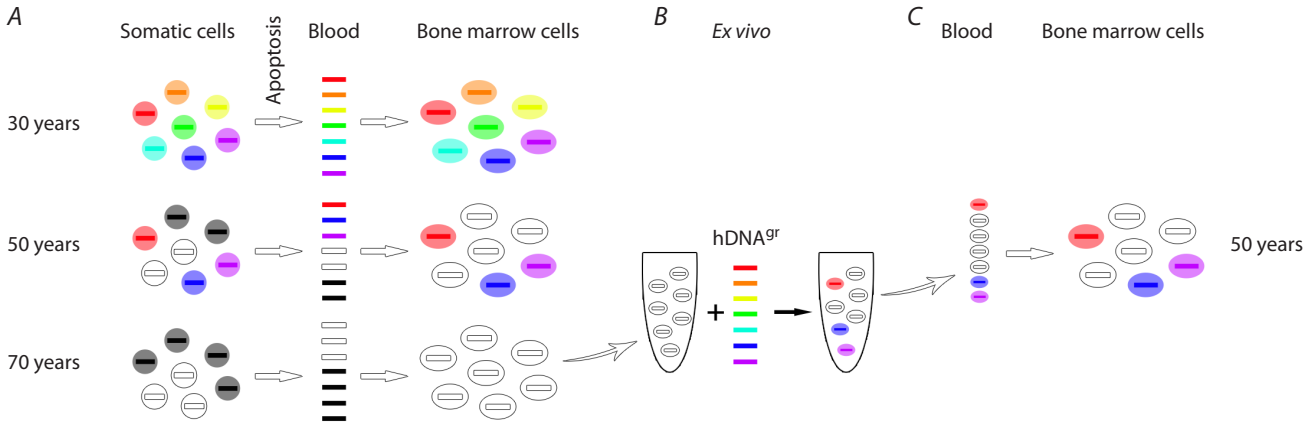


Fig. 7. The mechanistic diagram.

A – the emergence of mutations in somatic cells, their release into peripheral blood, and fixation within bone marrow cells. Formation of oligoclonal hematopoiesis; *B* – alterations in the HSC genome induced by fragments of extracellular dsDNA in the *ex vivo* system; *C* – reinfusion of cells with reconstructed genomes into the peripheral circulation and their fixation within the bone marrow. Recovery of polyclonal hematopoiesis.

inferred that the aging program does not exist, since there are no specific molecular or cellular subprograms that would assume the function of controlling the disintegration of the organism's system elements with age. That's exactly what it means, that the causes of aging are not understood yet, and it is only markers of aging that can be discussed.

Nevertheless, we have formulated a sufficiently adequate concept demonstrating that organismal aging is a program involving the dialectical unity of certain elements of its function and continuous interaction with the hostile environment, being related to the development of clonal hematopoiesis (this view is possibly characteristic of any stem cell system within the organism). This concept integrates all the revealed biological phenomena and related events. We also propose a method to overcome the sequelae of telomere shortening, fixation of pathological mutations, and their dominance in HSCs (as well as in stem cells of another origin), which could underlie future anti-aging therapy and treatment for diseases of civilization.

The putative "programmed clonal hematopoiesis" and "programmed death" mechanisms. At birth, humans have two fundamentally self-regulating mutually related systems: HSCs (possibly all other types of stem cells involved in regeneration processes in the organism) and the entire population of somatic cells.

- HSCs are responsible for regeneration in the body. They are hidden in the most secluded location, the bone marrow.
- Somatic cells interact with the hostile external environment.

The HSC population at birth is characterized by polyclonality, representing the full range of allelic variants required to perform regeneration function and maintain immune diversity.

The following features underlie the programmed clonal hematopoiesis and programmed aging mechanisms:

- HSCs internalize extracellular dsDNA fragments constituting approximately 0.02–1.0 % of the haploid genome.
- During terminal differentiation, HSCs reorganize the chromatin architecture via induction of single-strand breaks and relaxation of chromosomal DNA strands, thereby inducing a "recombinogenic situation" during which dsDNA fragments located within internal cellular compartments and delivered into the cell under various circumstances can be integrated.
- Having been internalized by a primitive hematopoietic cell, dsDNA fragments induce terminal differentiation of this cell. Hence, dsDNA fragments delivered into the cell both trigger a "recombinogenic situation" and participate in recombination process activated by them.
- dsDNA fragments constantly circulating in blood plasma are remnants of apoptotic cells; their chromatin eventually hydrolyses to a size of 1–20 nucleosome monomers.

The core of the mechanism. After natural death via apoptosis and subsequent secondary necrosis, organismal cells, with the entire range of mutations accumulated over their lifespan, release their fragmented dsDNA harboring numerous genetic abnormalities within various genes into the bloodstream.

This extracellular dsDNA is continuously transiently internalized by HSCs, where it induces commitment and is stochastically integrated into the genome, presumably at homologous regions, at specific time instants such as increased plasma DNA concentration following injury or ischemia. Gene conversion takes place, causing alterations in the genetics of HSCs: they lose their stem cell identity and acquire features characteristic of somatic cells that have responded to aggressive environmental stimuli. Over time, the HSC population becomes depleted as its pluripotent potential is exhausted, and the HSC system acquires oligoclonal features (Fig. 7).

This process reduces the DNA repair capacity of HSCs, and most importantly, results in loss of the original immune diversity, leading to physical aging of the organism.

It is possible to overcome the programmed clonal hematopoiesis and programmed aging mechanisms. It can be achieved by employing the same "weapon": regularly providing HSCs with an extracellular dsDNA substrate derived from healthy young individuals, hDNA^{gr} preparation being such a substrate. Healthy alleles will displace the accumulating mutant ones, and HSCs will maintain their pluripotent status.

References

Aubert G., Lansdorp P.M. Telomeres and aging. *Physiol Rev*. 2008; 88(2):557-579. doi 10.1152/physrev.00026.2007

Bernardes de Jesus B., Vera E., Schneeberger K., Tejera A.M., Ayuso E., Bosch F., Blasco M.A. Telomerase gene therapy in adult and old mice delays aging and increases longevity without increasing cancer. *EMBO Mol Med*. 2012;4(8):691-704. doi 10.1002/emmm.201200245

Bowen R.L., Atwood C.S. Living and dying for sex. A theory of aging based on the modulation of cell cycle signaling by reproductive hormones. *Gerontology*. 2004;50(5):265-290. doi 10.1159/000079125

Brooks R.W., Robbins P.D. Treating age-related diseases with somatic stem cells. In: Mettinger K., Rameshwar P., Kumar V. (Eds) Exosomes, Stem Cells and MicroRNA. Advances in Experimental Medicine and Biology. Vol. 1056. Springer, 2018;29-45. doi 10.1007/978-3-319-74470-4_3

Chakravarti D., LaBella K.A., DePinho R.A. Telomeres: history, health, and hallmarks of aging. *Cell*. 2021;184(2):306-322. doi 10.1016/j.cell.2020.12.028

da Costa J.P., Vitorino R., Silva G.M., Vogel C., Duarte A.C., Rocha-Santos T. A synopsis on aging – theories, mechanisms and future prospects. *Ageing Res Rev*. 2016;29:90-112. doi 10.1016/j.arr.2016.06.005

Davidsohn N., Pezzone M., Vernet A., Graveline A., Oliver D., Slomovic S., Punthambaker S., Sun X., Liao R., Bonventre J.V., Church G.M. A single combination gene therapy treats multiple age-related diseases. *Proc Natl Acad Sci USA*. 2019;116(47):23505-23511. doi 10.1073/pnas.1910073116

Di Micco R., Krizhanovsky V., Baker D., d'Adda di Fagagna F. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nat Rev Mol Cell Biol*. 2021;22(2):75-95. doi 10.1038/S41580-020-00314-w

Doksani Y. The response to DNA damage at telomeric repeats and its consequences for telomere function. *Genes*. 2019;10(4):318. doi 10.3390/genes10040318

Dolgova E.V., Nikolin V.P., Popova N.A., Proskurina A.S., Orishenko K.E., Alyamkina E.A., Efremov Y.R., … Taranov O.S., Rogachev V.A., Sidorov S.V., Bogachev S.S., Shurdov M.A. Internalization of exogenous DNA into internal compartments of murine bone marrow cells. *Russ J Genet Appl Res*. 2012;2:440-452. doi 10.1134/S2079059712060056

Gems D., Partridge L. Genetics of longevity in model organisms: debates and paradigm shifts. *Annu Rev Physiol*. 2013;75:621-644. doi 10.1146/annurev-physiol-030212-183712

Gowing G., Svendsen S., Svendsen C.N. Ex vivo gene therapy for the treatment of neurological disorders. *Prog Brain Res*. 2017;230:99-132. doi 10.1016/bs.pbr.2016.11.003

Hande M.P. DNA repair factors and telomere-chromosome integrity in mammalian cells. *Cytogenet Genome Res*. 2004;104(1-4):116-122. doi 10.1159/000077475

Harley C.B., Futcher A.B., Greider C.W. Telomeres shorten during ageing of human fibroblasts. *Nature*. 1990;345(6274):458-460. doi 10.1038/345458a0

He S., Sharpless N.E. Senescence in health and disease. *Cell*. 2017; 169(6):1000-1011. doi 10.1016/j.cell.2017.05.015

Hong J., Yun C.O. Telomere gene therapy: polarizing therapeutic goals for treatment of various diseases. *Cells*. 2019;8(5):392. doi 10.3390/cells8050392

Kenyon C.J. The genetics of ageing. *Nature*. 2010;464(7288):504-512. doi 10.1038/nature08980

Kirkwood T.B.L. Understanding the odd science of aging. *Cell*. 2005; 120(4):437-447. doi 10.1016/j.cell.2005.01.027

Kozlov A.A. Old age: social disunity or integrity? Theories and traditions of Western social psychology. *Mir Psikhologii: Nauchno-Metodicheskiy Zhurnal = World of Psychology: Research and Methodological Journal*. 1999;2:80-96 (in Russian)

Krauss S.R., de Haan G. Epigenetic perturbations in aging stem cells. *Mamm Genome*. 2016;27(7-8):396-406. doi 10.1007/s00335-016-9645-8

Lefever R. The rehabilitation of irreversible processes and dissipative structures' 50th anniversary. *Philos Trans A Math Phys Eng Sci.* 2018;376(2124):20170365. doi 10.1098/rsta.2017.0365

Li Y., Zhou G., Bruno I.G., Cooke J.P. Telomerase mRNA reverses senescence in progeria cells. *J Am Coll Cardiol*. 2017;70(6):804-805. doi 10.1016/j.jacc.2017.06.017

Loe T.K., Li J.S.Z., Zhang Y., Azeroglu B., Boddy M.N., Denchi E.L. Telomere length heterogeneity in ALT cells is maintained by PML-dependent localization of the BTR complex to telomeres. *Genes Dev*. 2020;34(9-10):650-662. doi 10.1101/gad.333963.119

López-Otín C., Blasco M.A., Partridge L., Serrano M., Kroemer G. The hallmarks of aging. *Cell*. 2013;153(6):1194-1217. doi 10.1016/j.cell.2013.05.039

López-Otín C., Blasco M.A., Partridge L., Serrano M., Kroemer G. Hallmarks of aging: an expanding universe. *Cell*. 2023;186(2):243-278. doi 10.1016/j.cell.2022.11.001

Lundblad V. Telomere maintenance without telomerase. *Oncogene*. 2002;21(4):522-531. doi 10.1038/sj.onc.1205079

Maniatis T., Fritch E., Sambrook D. Methods of Genetic Engineering. Molecular Cloning. Moscow: Mir Publ., 1984 (in Russian)

Mikheev R.K., Andreeva E.N., Grigoryan O.R., Sheremetyeva E.V., Absatarova Y.S., Odarchenko A.S., Opletaeva O.N. Molecular and cellular mechanisms of ageing: modern knowledge (literature review). *Problems of Endocrinology*. 2023;69(5):45-54. doi 10.14341/probl13278 (in Russian)

Nabetani A., Ishikawa F. Alternative lengthening of telomeres pathway: recombination-mediated telomere maintenance mechanism in human cells. *J Biochem*. 2011;149(1):5-14. doi 10.1093/jb/mvq119

Nassour J., Radford R., Correia A., Fusté J.M., Schoell B., Jauch A., Shaw R.J., Karlseder J. Autophagic cell death restricts chromosomal instability during replicative crisis. *Nature*. 2019;565:659-663. doi 10.1038/S41586-019-0885-0

Ocampo A., Reddy P., Martinez-Redondo P., Platero-Luengo A., Hatanaka F., Hishida T., Li M., … Nuñez Delicado E., Campistol J.M., Guillen I., Guillen P., Izpisua Belmonte J.C. In vivo amelioration of age-associated hallmarks by partial reprogramming. *Cell*. 2016; 167(7):1719-1733.e12. doi 10.1016/j.cell.2016.11.052

Pickett H.A., Cesare A.J., Johnston R.L., Neumann A.A., Reddel R.R. Control of telomere length by a trimming mechanism that involves generation of t-circles. *EMBO J*. 2009;28(7):799-809. doi 10.1038/emboj.2009.42

Proshkina E.N., Solovev I.A., Shaposhnikov M.V., Moskalev A.A. Key molecular mechanisms of aging, biomarkers, and potential interventions. *Mol Biol*. 2020;54(6):777-811. doi 10.1134/S0026893320060096

Rossiello F., Jurk D., Passos J.F., d'Adda di Fagagna F. Telomere dysfunction in ageing and age-related diseases. *Nat Cell Biol*. 2022; 24(2):135-147. doi 10.1038/s41556-022-00842-x

Rovatsos M.T., Marchal J.A., Romero-Fernández I., Fernández F.J., Giagia-Athanosopoulou E.B., Sánchez A. Rapid, independent, and extensive amplification of telomeric repeats in pericentromeric regions in karyotypes of arvicoline rodents. *Chromosome Res*. 2011; 19(7):869-882. doi 10.1007/S10577-011-9242-3

Ruzanova V.S., Oshikhmina S.G., Ritter G.S., Dolgova E.V., Kirikovich S.S., Levites E.V., Efremov Y.R., Karamysheva T.V., Bogomolov A.G., Meschaninova M.I., Mamaev A.L., Taranov O.S., Sidorov S.V., Nikonov S.D., Leplina O.Y., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Proskurina A.S., Bogachev S.S. Concept of natural genome reconstruction. Part 3. Analysis of changes in the amount of telomeric DNA in colony cells as a new amplified feature that arose during the processing of hematopoietic bone marrow stem cells. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025; 29(4):479-495. doi 10.18699/vjgb-25-52

Sogabe Y., Seno H., Yamamoto T., Yamada Y. Unveiling epigenetic regulation in cancer, aging, and rejuvenation with in vivo reprogramming technology. *Cancer Sci*. 2018;109(9):2641-2650. doi 10.1111/cas.13731

Stein G.H., Drullinger L.F., Soulard A., Dulić V. Differential roles for cyclin-dependent kinase inhibitors p21 and p16 in the mechanisms of senescence and differentiation in human fibroblasts. *Mol Cell Biol*. 1999;19(3):2109-2117. doi 10.1128/mcb.19.3.2109

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

Tchkonia T., Zhu Y., van Deursen J., Campisi J., Kirkland J.L. Cellular senescence and the senescent secretory phenotype: therapeutic opportunities. *J Clin Invest*. 2013;123(3):966-972. doi 10.1172/jci64098

Tlidi M., Clerc M.G., Panajotov K. Dissipative structures in matter out of equilibrium: from chemistry, photonics and biology, the legacy of Ilya Prigogine (part 1). *Philos Trans A Math Phys Eng Sci*. 2018a; 376(2124):20180114. doi 10.1098/rsta.2018.0114

Tlidi M., Clerc M.G., Panajotov K. Dissipative structures in matter out of equilibrium: from chemistry, photonics and biology, the legacy of Ilya Prigogine (part 2). *Philos Trans A Math Phys Eng Sci*. 2018b; 376(2135):20180276. doi 10.1098/rsta.2018.0276

Vaiserman A.M., Lushchak O.V., Koliada A.K. Anti-aging pharmacology: promises and pitfalls. *Ageing Res Rev*. 2016;31:9-35. doi 10.1016/j.arr.2016.08.004

Vijg J., Campisi J. Puzzles, promises and a cure for ageing. *Nature*. 2008;454(7208):1065-1071. doi 10.1038/nature07216

Zhu X., Chen Z., Shen W., Huang G., Sedivy J.M., Wang H., Ju Z. Inflammation, epigenetics, and metabolism converge to cell senescence and ageing: the regulation and intervention. *Signal Transduct Target Ther*. 2021;6(1):245. doi 10.1038/S41392-021-00646-9

Conflict of interest. The authors declare no conflict of interest.
Received July 5, 2024. Revised March 12, 2025. Accepted June 23, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(2):362-371

doi 10.18699/vjgb-26-40

Generation of the RCPCMi014-A and RCPCMi014-B lines from T-lymphocytes of a healthy donor and verification of their monoclonal origin by TCR V(D)J rearrangement analysis

D.K. Sherman [1&], M.E. Bogomiakova [1, 2&], E.A. Kastueva [1], I.V. Zvyagin [3], V.K. Ruppel[3, 4], E.V. Barsova [3, 4], A.Y. Gorbachev [1], N.A. Kulemin [1], A.A. Barinova [1], E.A. Zerkalenkova [5, 6], A.N. Bogomazova [1, 2], M.A. Lagarkova [1]

[1] Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia
[2] Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia
[3] Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia
[4] LLC "MiLaboratori", Moscow, Russia
[5] Dmitry Rogachev National Medical Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia
[6] Pirogov Russian National Research Medical University, Moscow, Russia
dar.sher.man0@gmail.com

Abstract. The generation of induced pluripotent stem cells (iPSCs) from peripheral blood T-lymphocytes offers a promising alternative to skin fibroblasts. This approach combines minimally invasive sample collection with rapid isolation of enriched target cell population and provides a lower baseline mutational burden, as T-lymphocytes are shielded from chronic ultraviolet radiation – a major driver of somatic mutations in skin fibroblasts. Objective verification of monoclonality of the resulting iPSC lines is possible due to V(D)J rearrangements in T-cell receptor (TCR) genes, which are formed as a result of somatic recombination in the thymus during the natural maturation of T-lymphocytes. These rearrangements remain unchanged during reprogramming and serve as a unique marker, allowing for reliable identification of monoclonal lines and exclusion of contaminated or polyclonal lines. Reliable verification of the clonal origin of iPSCs can be crucial in experiments that place high demands on genetic homogeneity. This work is dedicated to the generation and comprehensive characterization of monoclonal iPSC lines derived from CD3+ T-lymphocytes of a healthy donor using episomal reprogramming. The resulting lines, RCPCMi014-A (PBM022E5) and RCPCMi014-B (PBM022E7), meet the accepted quality criteria for iPSCs: they demonstrate expression of key pluripotency markers (OCT4, SOX2, SSEA-4, TRA-1-81), the ability to differentiate into derivatives of all three germ layers, and possess a normal karyotype. Both lines showed a high-efficiency capacity to differentiate into definitive endoderm, as assessed by CXCR4 expression, highlighting their potential for developing protocols to generate pancreatic β-cells. The obtained iPSC lines can be used for fundamental research into the mechanisms of pluripotency and differentiation, as well as for creating isogenic iPSC lines with introduced mutations for modeling rare hereditary diseases.

Key words: induced pluripotent stem cells; T-lymphocytes; reprogramming; T-cell receptor; cell clone; V(D)J recombination; definitive endoderm

For citation: Sherman D.K., Bogomiakova M.E., Kastueva E.A., Zvyagin I.V., Ruppel V.K., Barsova E.V., Gorbachev A.Y., Kulemin N.A., Barinova A.A., Zerkalenkova E.A., Bogomazova A.N., Lagarkova M.A. Generation of the RCPCMi014-A and RCPCMi014-B lines from T-lymphocytes of a healthy donor and verification of their monoclonal origin by TCR V(D)J rearrangement analysis. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):362-371. doi 10.18699/vjgb-26-40

Funding. The study was carried out with the financial support of the Ministry of Science and Higher Education of the Russian Federation (The Federal Scientific and Technical Program for the Development of Genetic Technologies 2019–2030, project No. 075-15-2025-518).

& Equal contributed.

Получение линий индуцированных плюрипотентных стволовых клеток RCPCMi014-A и RCPCMi014-B из Т-лимфоцитов здорового донора и верификация их моноклонального происхождения по V(D)J-перестройкам генов TCR

Д.К. Шерман [1&], М.Е. Богомякова [1, 2&], Э.А. Кастуева [1], И.В. Звягин [3], В.К. Руппель [3, 4], Е.В. Барсова [3, 4], А.Ю. Горбачев [1], Н.А. Кулемин [1], А.А. Баринова [1], Е.А. Зеркаленкова [5, 6], А.Н. Богомазова [1, 2], М.А. Лагарькова [1]

[1] Федеральный научно-клинический центр физико-химической медицины им. академика Ю.М. Лопухина Федерального медико-биологического агентства России, Москва, Россия
[2] Центр «Генетическое репрограммирование и генная терапия» Федерального научно-клинического центра физико-химической медицины им. академика Ю.М. Лопухина Федерального медико-биологического агентства России, Москва, Россия
[3] Государственный научный центр Российской Федерации Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук, Москва, Россия
[4] ООО «Майлаборатори», Москва, Россия
[5] Национальный медицинский исследовательский центр детской гематологии, онкологии и иммунологии им. Дмитрия Рогачева Министерства здравоохранения Российской Федерации, Москва, Россия
[6] Российский национальный исследовательский медицинский университет им. Н.И. Пирогова Министерства здравоохранения Российской Федерации, Москва, Россия

dar.sher.man0@gmail.com

Аннотация. Получение индуцированных плюрипотентных стволовых клеток из Т-лимфоцитов периферической крови представляет собой перспективную альтернативу использованию фибробластов кожи. Этот подход сочетает минимально инвазивный забор биоматериала совместно с быстрым получением обогащенной популяции целевых клеток. Дополнительным преимуществом является относительно низкий мутационный груз Т-лимфоцитов по сравнению с традиционным источником соматических клеток для репрограммирования – фибробластами кожи, ввиду того что Т-лимфоциты защищены от хронического воздействия ультрафиолетового излучения – одного из ключевых факторов накопления соматических мутаций. Объективная верификация моноклональности полученных линий индуцированных плюрипотентных стволовых клеток возможна благодаря наличию V(D)J-перестроек генов Т-клеточного рецептора, формирующихся в результате соматической рекомбинации в тимусе в процессе естественного созревания Т-лимфоцитов. Эти перестройки остаются неизменными при репрограммировании и служат уникальным маркером, позволяющим достоверно идентифицировать моноклональные линии и исключать контаминированные линии или линии поликлонального происхождения. Надежная верификация клонального происхождения индуцированных плюрипотентных стволовых клеток может быть важной в экспериментах, предъявляющих высокие требования к генетической однородности. Данная работа посвящена созданию и комплексной характеристике моноклональных линий индуцированных плюрипотентных стволовых клеток, полученных из CD3+ Т-лимфоцитов здорового донора методом эписомного репрограммирования. Полученные линии RCPCMi014-A (PBM022E5) и RCPCMi014-B (PBM022E7) соответствуют принятым критериям качества для индуцированных плюрипотентных стволовых клеток: они демонстрируют экспрессию ключевых маркеров плюрипотентности (OCT4, SOX2, SSEA-4, TRA-1-81), способность к дифференцировке в производные всех трех зародышевых листков и обладают нормальным кариотипом. Обе линии показали способность к дифференцировке в дефинитивную эндодерму с высокой эффективностью, оцененной по экспрессии CXCR4, что определяет их перспективность для разработки протоколов генерации β-клеток поджелудочной железы. Полученные линии индуцированных плюрипотентных стволовых клеток могут быть использованы для фундаментальных исследований механизмов плюрипотентности и дифференцировки, а также для создания изогенных линий индуцированных плюрипотентных стволовых клеток с внесенными мутациями при моделировании редких наследственных заболеваний.

Ключевые слова: индуцированные плюрипотентные стволовые клетки; Т-лимфоциты; репрограммирование; Т-клеточный рецептор; клон клеток; V(D)J-рекомбинация; дефинитивная эндодерма

Introduction

The discovery of induced pluripotent stem cells (iPSCs) in 2006 opened up new possibilities in regenerative medicine and biomedical research (Yamanaka, 2020). These cells have the ability to differentiate into all three germ layers and hold great potential for disease modeling *in vitro*. This allows researchers to study tissue-specific pathological processes at a cellular level (Summers et al., 2024). Typically, iPSC models are created from patients with inherited diseases (Grigor'eva et al., 2023), which helps to replicate individual disease characteristics. Healthy donor lines serve as important reference controls in these studies. They are especially valuable for studying orphan diseases, which have a small patient population, making it difficult to access cellular models directly. In these cases, creating isogenic pairs through editing healthy donor iPSCs is a promising

solution (Khomyakova et al., 2023; Fedorenko et al., 2024). It is important to note that iPSCs from healthy donors represent an essential resource for standardizing differentiation protocols, creating "universal" cell products, and studying basic reprogramming mechanisms (Cerneckis et al., 2024).

According to established practice, each morphologically homogeneous iPSC colony is considered to be a clone originating from a single reprogrammed precursor cell (Takahashi et al., 2007). Following this practice, individual colonies are cultured independently and are regarded as distinct cell lines. However, direct confirmation of monoclonality is rarely performed, and the lack of verification methods can lead to polyclonal lines going undetected, especially when colonies form due to the close proximity of cells. Cell line authentication, which includes confirmation of genetic identity and the absence of contamination, is a crucial laboratory procedure for ensuring research reproducibility, minimizing financial losses, and maintaining data reliability (Harbut et al., 2024).

In this context, one of the unique advantages of iPSCs derived from T-lymphocytes is the ability to verify monoclonal origin through analysis of genomic rearrangements that occur during V(D)J recombination during the formation of T-cell receptor (TCR) gene chains (Kishino et al., 2014). These clone-specific rearrangements, which are preserved after reprogramming, serve as a molecular "barcode" for the original T-cell. Modern approaches to V(D)J analysis, such as multiplex PCR (Seki et al., 2010) and targeted high-throughput sequencing (Nishimura et al., 2013), enable direct detection of polyclonal origins, which is crucial for ensuring the reproducibility and reliability of biomedical research results. The combination of this objective clonality verification with minimally invasive biomaterial collection makes T-lymphocyte-derived iPSCs an attractive strategy for both basic research and disease modeling applications.

In this study, iPSC lines were generated from healthy donor T-lymphocytes using episomal reprogramming. For the first time, polyclonal populations were excluded by analyzing V(D)J rearrangements at three genomic loci, allowing for reliable identification of monoclonal lines. The iPSC lines RCPCMi014-A (PBM022E5) and RCPCMi014-B (PBM022E7) with confirmed monoclonal origin demonstrated full compliance with quality criteria, including maintenance of a normal karyotype, expression of key markers (OCT4, SOX2, SSEA-4, TRA-1-81), and the capacity to differentiate into derivatives of all three germ layers. Specifically, they showed high efficiency in directed differentiation into definitive endoderm.

Materials and methods

Ethical considerations and collection of biological materials. The study was approved by the Research Ethics Committee of the Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Protocol No. 1, dated June 1, 2021. The biological material (peripheral venous blood) was obtained from a 35-year-old male volunteer donor. Blood sampling was performed by qualified medical staff of the clinic at the Lopukhin FRCC PCM Russia, in accordance with all standard medical protocols. Prior to the procedure, the donor was informed about the research objectives and provided voluntary informed consent for the use of his biological material for research purposes.

Exome sequencing and bioinformatic analysis. Genomic DNA was extracted from frozen whole blood samples using the ExtractDNA Blood & Cells kit (Evrogen). The sequencing library was prepared with the MGIEasy Universal DNA Library Prep Set (MGI), and exome regions were enriched using the VAHTS Target Capture Core Exome Panel (Vazyme), following the manufacturer's protocols. The resulting library underwent paired-end sequencing on the MGISEQ-2000 platform (MGI), achieving coverage of ~80x.

Primary bioinformatic processing included quality control of raw reads using FastQC, with subsequent filtering of low-quality reads and adapter sequences. Read alignment to the human reference genome GRCh37/hg19 was performed using the BWA-MEM algorithm. Alignment quality and enrichment efficiency were assessed using SAMtools, mosdepth, and Picard Tools. Detection of single nucleotide variants and small insertions/deletions (SNV/InDel) was conducted with the DeepVariant software package.

Isolation and activation of T-lymphocytes. Peripheral blood mononuclear cells (PBMCs) were isolated from venous blood by density gradient centrifugation using NycoPrep™ reagent (Axis-Shield) according to the standard protocol (Grigor'eva et al., 2024). CD3+ T-cells were subsequently isolated from PBMCs using magnetic-activated cell sorting with the Human CD3+ Cell Separation Kit (RWD), following the manufacturer's protocol. T-lymphocyte activation was performed using a modified two-step protocol (Maslennikova et al., 2022). During the first step, the cells were incubated with Dynabeads™ Human T-Activator CD3/CD28 magnetic particles (Gibco). After 24 hours, beads were removed using a magnetic stand, and the cells were resuspended in T-cell medium consisting of RPMI-1640 (PanEco) supplemented with 10 % heat-inactivated HI-FBS (BioFroxx) and 100 IU/mL IL-2 (Biotech NPK) at a concentration of 0.5×10^6 cells/mL for an additional 24 hours of culture.

Reprogramming of T-cells and iPSC culture. T-cell reprogramming was performed via transfection with episomal vectors (Okita et al., 2013) expressing reprogramming factors SOX2, OCT4, KLF4, L-MYC, LIN28, a dominant-negative form of mouse p53 (mp53DD), and the episomal maintenance factor EBNA1 (Addgene #41813-14, 41855-57). Transfection was carried out using the Neon Transfection System (ThermoFisher Scientific) with standard parameters: 1,650 V, 10 ms, 3 pulses. The pCE-GFP plasmid (Addgene #41858) served as a

transfection efficiency control. Twenty-four hours post-electroporation, the number of viable, DAPI-negative activated blast cells – characterized by increased size (FSC) and granularity (SSC) – was quantified using a NovoCyte flow cytometer (ACEA Biosciences). Cell suspensions were plated on 35-mm Petri dishes pre-coated with 1:50 diluted Matrigel (Corning) at a density of 10–12 × 10^3 viable blasts/cm^2 in T-cell medium. The following day, an equal volume of reprogramming medium was added, consisting of ReproTeSR medium (Stemcell Technologies), 100 ng/mLFGF2 (produced in-house), 10 μM ROCK inhibitor Y-27632, 0.5 μM PD0325901, and 2 μM SB431542 (all from DC Chemicals). Reprogramming medium was subsequently replaced every other day. On day 16, cells were treated with modified iPSC medium composed of a 4:1 mixture of GibberS-8 (PanEco) and mTeSR1 (Stemcell Technologies), with daily medium changes. On day 26, colonies were mechanically isolated using 0.1 % dispase solution (Invitrogen). The resulting iPSC clones were maintained under feeder-free conditions according to a previously described protocol (Goliusova et al., 2025).

Analysis of transgene integration by real-time qPCR. To assess potential integration of episomal vectors into the genome, iPSC lines were cultured for at least 10 passages. Genomic DNA was subsequently extracted using the M-Sorb kit (Syntol). Transgene integration analysis was performed by real-time quantitative PCR on a CFX96 Touch Real-Time PCR Detection System (BioRad) using 5X qPCRmix-HS SYBR (Evrogen) and specific primers (Supplementary Table S1)[1]. The amplification protocol consisted of the following steps: initial denaturation at 95 °C for 4 min; 40 cycles of 95 °C for 10 s and 60 °C for 40 s (with fluorescence detection); and a final melting curve analysis: 95 °C – 5 s, 65 °C – 5 s (with continuous fluorescence acquisition), 95 °C – 50 s).

Analysis of V(D)J rearrangement sequences. To determine the clonal composition of iPSC lines, we analyzed sequences of three genomic loci formed through V(D)J recombination that determine the mature α/δ-, β- and γ-chain TCR genes. High-throughput sequencing libraries were prepared from 42 ng of genomic DNA (equivalent to ~6,500 diploid genomes) using the Human IG/TCR DNA Multiplex Kit 7G (MyLaboratory) according to the manufacturer's protocol. Sequencing was performed on the Illumina MiSeq platform in paired-end mode with 150 bp read length. A minimum of 90,000 paired-end reads were obtained for each sample.

Bioinformatic analysis was conducted using the MiXCR software package (https://mixcr.com). A minimum of 83 % of reads in each sample were successfully identified as containing sequences unambiguously attributable to one of the investigated loci through alignment with reference sequences of the corresponding loci using MiXCR software. To determine clonality, we examined the distribution of read counts and predicted functionality of identified V(D)J rearrangement sequences in each of the three loci. Rearrangements supported by at least 100 independent reads were included in the analysis.

Mycoplasma screening. The absence of mycoplasma contamination was confirmed by PCR using primers targeting the 16S ribosomal RNA gene (Table S1), as previously described (Goliusova et al., 2024).

RNA isolation and RT-PCR. Total RNA was isolated from 1 million cells using Extract RNA reagent (Evrogen) according to the manufacturer's protocol. Reverse transcription was performed using the MMLV RT kit (Evrogen) and random decanucleotide primers. RT-PCR was conducted using specific primers (Table S1). The amplification protocol was: 95 °C for 3 min; 32 cycles of 95 °C for 15 s, 60 °C for 15 s, 72 °C for 30 s; final elongation: 72 °C for 5 min.

Karyotyping. Karyotyping was performed according to a previously described method at a resolution of 400 bands (Goliusova et al., 2024).

STR analysis for cell line authentication. STR profiling was conducted using the commercial COrDIS Plus kit (Gordiz), which analyzes 19 autosomal loci and the sex marker Amelogenin.

Analysis of pluripotency marker expression by flow cytometry. Flow cytometry was performed according to previously described methodology (Khomyakova et al., 2023). Antibodies used in the study are listed in Table S2.

Immunocytochemical staining. Immunocytochemical staining was carried out following a previously published protocol (Fedorenko et al., 2024). Antibodies used in the study are listed in Table S2. Visualization was performed using an Olympus IX53F inverted fluorescence microscope. Image processing was conducted using GIMP-2.10 software.

Functional assessment of pluripotency by spontaneous differentiation. To confirm the pluripotent status of iPSCs, a spontaneous differentiation test was performed to assess their ability to form derivatives of the three germ layers, following previously described protocol (Goliusova et al., 2024). After 15–20 days of adherent culture, embryoid bodies were fixed with 4 % PFA and stained for specific markers of the three germ layers (Table S2).

Differentiation into definitive endoderm. Differentiation into definitive endoderm was performed using the STEMdiff™ Definitive Endoderm Kit (STEMCELL Technologies) according to the manufacturer's protocol. Differentiation efficiency was evaluated based on the expression of the CXCR4 marker using antibodies listed in Table S2.

Results and discussion

This study focuses on the generation and detailed characterization of iPSC lines derived from T-lymphocytes of a healthy donor. Whole-exome sequencing of the donor revealed no pathogenic or likely pathogenic variants in genes associated with oncological diseases, according to the PanelApp database (Martin et al., 2019) and ACMG

[1] Supplementary Tables S1–S3 and Figure S1 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Sher_Engl_30_3.pdf

guidelines (Richards et al., 2015). The reprogramming scheme is presented in Figure 1*a*. CD3+ T-lymphocytes isolated from peripheral blood were activated using CD3/CD28 Dynabeads™ and transfected with a cocktail of episomal plasmids encoding *OCT4*, *SOX2*, *KLF4*, *c-MYC*, *LIN28*, *EBNA1*, and *mp53DD*. Twenty-four hours post-transfection, cells were plated on Matrigel-coated dishes and cultured in medium supplemented with a combination of small molecules: Y-27632 (ROCK inhibitor), SB431542 (TGF-β inhibitor), and PD0325901 (MEK inhibitor), which enhances cell survival and reprogramming efficiency (Watanabe et al., 2019). Three weeks after transfection, the formation of compact colonies morphologically resembling embryonic stem cell colonies was observed.

The overall strategy for iPSC clone selection and validation is presented in Fig. 1*b*. Episomal reprogramming of T-lymphocytes yielded 14 primary iPSC colonies (PBM022E1-E14) selected based on morphological criteria. Each colony was isolated and cultured as a separate cell line, consistent with standard practice in iPSC generation (Takahashi et al., 2007). Theoretically, morphologically homogeneous colonies could form from closely located reprogrammed cells, potentially resulting in polyclonal lines. While some studies demonstrate that such lines can maintain functional stability (Willmann et al., 2013), others show that polyclonality may manifest as differences in differentiation capacity (Yun et al., 2025) and proliferation rate (Mills et al., 2013), which could reduce experimental reproducibility.

Among the 14 primary iPSC clones obtained through reprogramming, 10 demonstrated stable cultivation for over 80 days (minimum 10 passages) and were selected for further functional characterization. Transgene integration analysis by qPCR revealed that four clones (PBM022E5, PBM022E7, PBM022E9, PBM022E14) completely lost episomal vectors (Fig. 1*c*). The iPSC line RCPCMi007-A-1, generated using non-integrating Sendai virus vectors (Bogomiakova et al., 2021), served as a negative control free of transgene integrations. The *B2M* gene, encoding beta-2-microglobulin protein, was used as a reference single-copy gene located on an autosomal chromosome. These findings align with literature reports on episomal plasmid elimination in other iPSC lines generated using this method (Grigor'eva et al., 2024; Podvysotskaya et al., 2025).

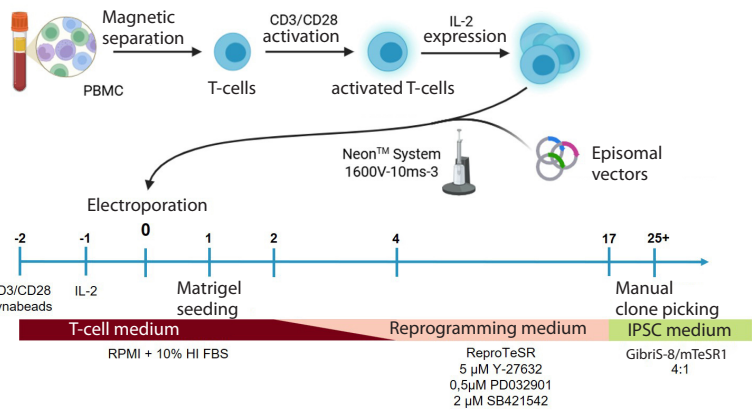
Compared to other somatic cells, mature T-lymphocytes enable objective confirmation of the monoclonality of derived iPSCs through analysis of genomic loci sequences generated by V(D)J recombination during T-cell maturation in the thymus (Alt et al., 1992). The random combination of V, D, and J segments, along with nucleotide insertions and deletions at their junctions, creates a CDR3 region with a sequence unique to each original T-cell precursor. This sequence is inherited by all cellular descendants and persists after reprogramming, enabling assessment of the clonal diversity in iPSC lines derived from mature T-lymphocytes.

Given the predominance of αβT-cells in the peripheral blood of a healthy adult donor, the obtained iPSC clones most likely originated from mature T-lymphocytes with functional TCR α- and β-chain genes (Kreslavsky et al., 2008, 2010). The criterion for monoclonality was defined as the presence of at least one productive V(D)J rearrangement pair in the *TRA* and *TRB* loci. The presence of multiple productive rearrangements and/or a total of more than two rearrangements in either locus would indicate polyclonal origin. Rearrangements in the *TRG* locus would not affect the interpretation of clonality based on the *TRA* and *TRB* loci, provided no rearrangements were detected in the δ-chain gene (located within the *TRAD* locus) and the total number of *TRG* rearrangements did not exceed two (Sherwood et al., 2011; Mahe et al., 2018).

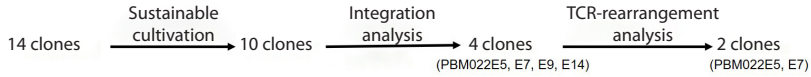
V(D)J repertoire analysis confirmed that all clones originated from αβT-cells, as productive rearrangements of the TCR δ-chain gene were absent. Two clones – PBM022E5 and PBM022E7 – exhibited a rearrangement pattern typical of monoclonal origin, characterized by one productive sequence each for the TCR α- and β-chain genes (Fig. 1*d*, highlighted in green). This profile allows these lines to be considered with high confidence as descendants of individual parental T-cells. In contrast, clone PBM022E14 was classified as polyclonal due to the presence of three productive β-chain and two α-chain TCR rearrangements (Fig. 1*d*). In total, four V(D)J rearrangements were identified in both the *TRA* and *TRB* loci of this clone (Table S3). Clone PBM022E9 displayed an asymmetric profile: one productive β-chain sequence was detected alongside three α-chain variants, suggesting potential polyclonal origin. Analysis of shared V(D)J rearrangements across all four investigated lines revealed identical TCR α- and β-chain sequences in clones PBM022E9 and PBM022E14 (Fig. 1*d* and Table S3, highlighted in red), indicating possible technical contamination.

Thus, analysis of genomic regions formed during V(D)J recombination in T-lymphocyte maturation represents an effective tool for verifying the clonality of iPSCs derived from mature T-lymphocytes. The unique CDR3 region sequence, generated through random combination of V, D, and J segments, serves as a reliable "genetic barcode" for each original cell, providing irrefutable evidence of monoclonal origin. The stability of this marker, which remains unchanged after reprogramming, enables its use for long-term clonal identification, including in differentiated derivatives. This approach becomes particularly valuable when working with unique models, such as iPSCs from monozygotic twins, where traditional DNA polymorphism-based genotyping methods lose informativeness due to the near-complete genetic identity of donors (Vlasov et al., 2021). Additionally, the method exhibits high sensitivity to contamination, enabling detection of cross-contamination

a



b



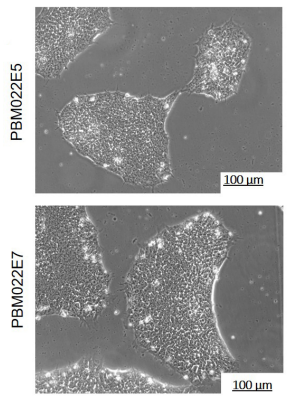
c

	E5	E7	E9	E14	negative control
B2M	23,8	23,0	22,8	23,1	23,2
OCT4	31,4	30,8	29,8	30,4	31,1
KLF4	29,3	29,1	28,2	28,8	29,0
SOX2	36,8	37,7	37,3	38,2	38,4
LIN28	27,8	27,7	26,9	27,5	28,0
c-MYC	38,2	39,1	37,9	39,1	36,5
mp53DD	36,1	33,8	35,2	34,7	34,0
EBNA1	35,0	33,5	35,5	34,9	33,9

d

		V-segment	CDR3 a/a sequence	J-segment
Monoclonal origin	PBM022E5	TRAV20	CAVQAASNNNDMRF	TRAJ43
		TRAV35	SRLGEE_ADGLTF	TRAJ45
		TRBV3-1	CASRTGTGGSPLHF	TRBJ1-6
		TRBV24-1	CATSDLSRG_MSLGETQYF	TRBJ2-5
		TRGV7	CATWDRLSYKKLF	TRGJ2
		TRGV9	CAPI_KKLF	TRGJ2
	PBM022E7	TRAV17	CATDAGGTSYGKLTF	TRAJ52
		TRAV38-2DV8	CAYINSG_GSEKLVF	TRAJ57
		TRBV5-1	CASSRKDSNTEAFF	TRBJ1-1
		TRGV3	CATWDRL_LHKKLF	TRGJ1
		TRGV10	CAAWD*NLFF	TRGJ2
Polyclonal origin	PBM022E9	TRAV13-1	CAALKVWQFYF	TRAJ49
		TRAV4	CLVGDMG_SNYQLIW	TRAJ33
		TRAV13-1	CAASIME_SQGNLIF	TRAJ42
		TRBV2	CASRLDDQAQHF	TRBJ1-5
		TRBV6-5	CASSTKD_VDTEAFF	TRBJ1-1
		TRGV8	CATWDSFADYKKLF	TRGJ2
		TRGV10	CAAWD_GKKLF	TRGJ1
		TRGV2	CATWD_KGKLF	TRGJ2
	PBM022E14	TRAV1-1	CAVNHYGQNFVF	TRAJ26
		TRAV13-1	CAALKVWQFYF	TRAJ49
		TRAV13-2	CAENQE_SYIPTF	TRAJ6
		TRAV13-1	CAASIME_SQGNLIF	TRAJ42
		TRBV7-7	CASTSQPGQRKEQPPTGELFF	TRBJ2-2
		TRBV18	CASSPPGRMIDTQYF	TRBJ2-3
		TRBV2	CASRLDDQAQHF	TRBJ1-5
		TRBV7-3	RASSLSSE_ELQETQYF	TRBJ2-5
		TRGV5	CATWGRKNYYKKLF	TRGJ2
		TRGV8	CATWG_YKKLF	TRGJ2
		TRGV2	CATWD_KGKLF	TRGJ2

e



f

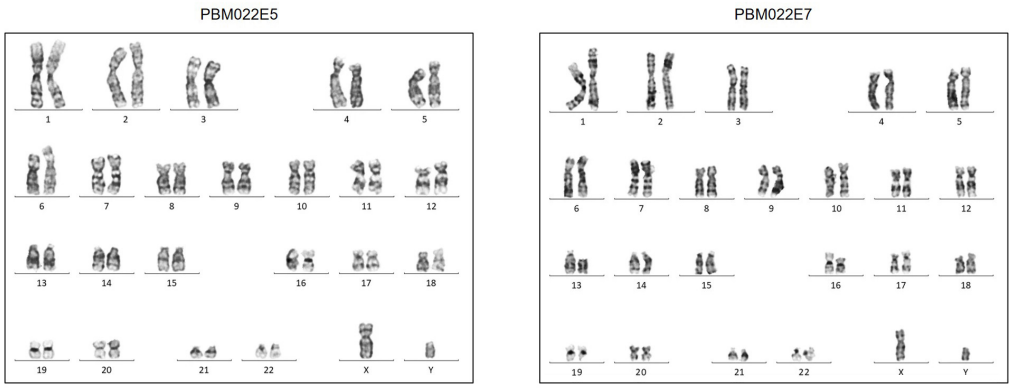


Fig. 1. T-cell reprogramming: *a* – schematic interpretation of T-cell activation and reprogramming; *b* – strategy for iPSC clone selection and validation; *c* – transgene integration analysis by Ct values: E5, E7, E9, E14 represent lines PBM022E5, PBM022E7, PBM022E9, PBM022E14, respectively; negative control: iPSC line RCPCMi007-A-1 (Bogomiakova et al., 2021). Comparable Ct values with the negative control indicate the absence of residual transgenes in the investigated clones; *d* – analysis of V(D)J rearrangement sequences: underline indicates frameshift disruption; green highlights, presumably functional chains without frameshift disruption; red highlights, V(D)J rearrangement variants shared between clones PBM022E9 and PBM022E14; *e* – morphology of the PBM022E5 and PBM022E7 lines; *f* – karyotype of the PBM022E5 and PBM022E7 lines.

between lines through the presence of identical V(D)J sequences.

However, the method has certain limitations. Firstly, it is applicable only to T-cell origins and cannot be used for iPSCs derived from other somatic cell types, such as fibroblasts, which lack analogous cell-specific markers. Secondly, result interpretation requires establishing clear threshold values (in our work – 100 reads) to distinguish technical errors from true polyclonality, as well as specialized sequencing equipment and bioinformatics expertise for data analysis. Despite these limitations, V(D)J profiling represents a powerful tool for standardizing cellular models and enhancing research reproducibility in cell biology.

The obtained iPSC lines PBM022E5 (RCPCMi014-A) and PBM022E7 (RCPCMi014-B) were registered in the Human Pluripotent Stem Cell Registry (hPSCreg, https://hpscreg.eu). Information about these cell lines can be found at the following links: https://hpscreg.eu/cell-line/RCPCMi014-A and https://hpscreg.eu/cell-line/RCPCMi014-B. Both lines demonstrated typical pluripotent stem cell morphology (Fig. 1*e*), normal karyotype (46, XY) (Fig. 1*f*), and confirmed genetic identity with the donor through STR profiling (data available from the authors upon request). All lines were tested for mycoplasma contamination, confirmed by negative PCR results (Fig. 2*a*). Expression of key pluripotency markers was confirmed by several methods. Immunocytochemical analysis revealed nuclear expression of OCT4 and SOX2, as well as surface expression of TRA-1-81 and SSEA-4 (Fig. 2*d*). RT-PCR results showed high expression levels of the *OCT4*, *SALL4*, and *DPPA5* genes (Fig. 2*b*). Flow cytometry data confirmed high percentages of positive cells: over 99 % expressed OCT4, over 99 % expressed SSEA-4, and over 92 % expressed TRA-1-81 (Fig. 2*c*). The capacity for differentiation into derivatives of all three germ layers was assessed through spontaneous *in vitro* differentiation. Expression of pan-cytokeratin confirmed ectoderm formation, troponin T confirmed it for mesoderm (including cardiomyocytes), and HNF4α confirmed it for endoderm (Fig. 2*d*).

The next aspect of characterization was the assessment of the lines' capacity to differentiate into definitive endoderm – a critical step for subsequent generation of

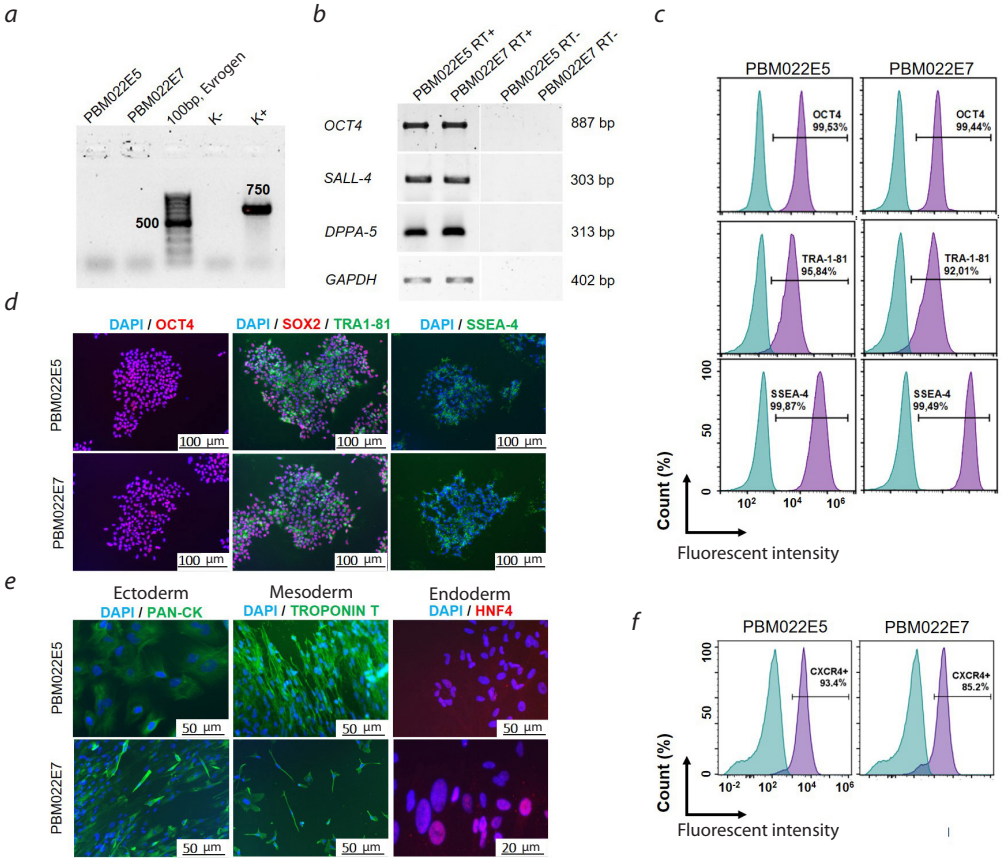


Fig. 2. Characterization of iPSC lines RCPCMi014-A (PBM022E5) and RCPCMi014-B (PBM022E7): *a* – PCR analysis for mycoplasma contamination; *b* – RT-PCR analysis of pluripotency marker expression; *c* – flow cytometry analysis of pluripotency markers; *d* – immunocytochemical staining of pluripotency markers: OCT4 (red), SOX2 (red), TRA-1-81 (green), SSEA-4 (green), nuclei stained with DAPI (blue); *e* – immunocytochemical staining of three germ layer markers: ectoderm PAN-CK (green), mesoderm TROPONIN T (green), endoderm HNF-4 (red); *f* – flow cytometry analysis of the definitive endoderm early marker CXCR4.

Cell line passports for RCPCMi014-A (PBM022E5) and RCPCMi014-B (PBM022E7)

Unique identifier	RCPCMi014-A, RCPCMi014-B
Alternative cell line names	PBM022E5, PBM022E7
Institution	Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia
Ethical approval	The study was approved by the Research Ethics Committee of the Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia, Protocol No. 1 dated June 1, 2021
Cell type	iPSC
Origin	Human
Additional information on the origin of the cell line	Age: 35 Gender: M Ethnicity: Caucasian race
Original cell type	T-cells
Date of biomaterial collection	March 26, 2024
Reprogramming method	Non-integrating episomal vectors
Reprogramming method	OCT4, SOX2, KLF4, LIN28, L-MYC, EBNA1 and mp53DD
Clonality	Clonal
Disease	No
Gene/loci	–
Morphology	Human pluripotent cell-like monolayer colonies
Pluripotency	Confirmed in tests for the formation of embryoid bodies, spontaneous differentiation into derivatives of three germ layers, as well as the expression of specific markers
Karyotype	46, XY
Check for contamination	Mycoplasma was not detected
Scope of application	*In vitro* reference model of healthy donor
Cultivation method	Non-feeder Matrigel layer
Growth medium	80 % GibriS-8, 20 % mTeSR1, 50 μg/ml Penicillin-Streptomycin or 10 μg/ml Gentamicin
Temperature, °C	37
CO_2 concentration, %	5
O_2 concentration, %	20
Passage method	0.05 % Trypsin-EDTA
Split ratio	1:4–1:6
Cryopreservation	90 % FBS, 10 % DMSO, 5 μM ROCK inhibitor Y-27632 (Stemcell Technologies)
Storage conditions	Liquid nitrogen
Registry ID	https://hpscreg.eu/cell-line/RCPCMi014-A https://hpscreg.eu/cell-line/RCPCMi014-B
Date of certification/registration	August 21, 2024

pancreatic β-cells. Our data reveal significant variability in differentiation potential among different lines: only one-third of iPSC lines demonstrate high efficiency in definitive endoderm formation (Fig. S1). Flow cytometry analysis showed that lines PBM022E5 and PBM022E7 exhibit high and stable capacity for definitive endoderm differentiation, as evidenced by CXCR4 marker expression in over 85 % of cells (Fig. 2*f*). These results confirm the potential utility of these lines for developing and optimizing protocols for directed differentiation into insulin-producing cells. Cell line passports for RCPCMi014-A (PBM022E5) and RCPCMi014-B (PBM022E7) are presented in the Table.

Conclusion

In this study, monoclonal iPSC lines reprogrammed from T-lymphocytes of a healthy donor were generated and characterized. Analysis of TCR gene V(D)J rearrangements provided reliable verification of the lines' monoclonal origin, which is critically important for models requiring high genetic homogeneity. The obtained lines RCPCMi014-A (PBM022E5) and RCPCMi014-B (PBM022E7) were shown to meet all quality criteria for iPSCs and demonstrate high efficiency in differentiating into definitive endoderm, making them promising for the generation of pancreatic β-cells. The established cell lines can serve as reliable reference models both for fundamental research into the mechanisms of pluripotency and differentiation and for applied tasks, including the creation of isogenic pairs for modeling rare and other diseases, as well as for optimizing methodological approaches in regenerative medicine.

References

Alt F.W., Oltz E.M., Young F., Gorman J., Taccioli G., Chen J. VDJ recombination. *Immunol Today*. 1992;13(8):306-314. doi 10.1016/0167-5699(92)90043-7

Bogomiakova M.E., Sekretova E.K., Eremeev A.V., Shuvalova L.D., Bobrovsky P.A., Zerkalenkova E.A., Lebedeva O.S., Lagarkova M.A. Derivation of induced pluripotent stem cells line (RCPCMi007-A-1) with inactivation of the beta-2-microglobulin gene by CRISPR/Cas9 genome editing. *Stem Cell Res.* 2021;55:102451. doi 10.1016/J.SCR.2021.102451

Cerneckis J., Cai H., Shi Y. Induced pluripotent stem cells (iPSCs): molecular mechanisms of induction and applications. *Signal Transduct Target Ther.* 2024;9(1):112. doi 10.1038/S41392-024-01809-0

Fedorenko A.V., Khomyakova E.A., Surdina A.V., Sekretova E.K., Limanskaya T.V., Belikova L.D., Volovikov E.A., … Fedotov D.A., Zerkalenkova E.A., Lagarkova M.A., Lebedev I.N., Bogomazova A.N. Design of iPSC-based cell model to study the functions of the *UBE2A* gene. *Genes and Cells*. 2024;19(2):297-313. doi 10.17816/GC623799 (in Russian)

Goliusova D.V., Lebedeva O.S., Sharikova M.Y., Kopylova I.V., Teryakova M.V., Lavrenteva K.A., Zerkalenkova E.A., Bogomazova A.N., Lagarkova M.A. Derivation of RCPCMi011-A induced pluripotent stem cell line from fibroblasts of a patient with restrictive cardiomyopathy caused by c.7416_7418delGAA mutation in the FLNC gene. *Russ J Dev Biol.* 2024;55(6):347-355. doi 10.1134/S1062360425700055

Goliusova D.V., Bogomolova A.P., Davidenko A.V., Lavrenteva K.A., Sharikova M.Y., Zerkalenkova E.A., Vassina E.M., Bogomazova A.N., Lagarkova M.A., Katrukha I.A., Lebedeva O.S. Metabolic culture medium enhances maturation of human iPSC-derived cardiomyocytes via cardiac troponin I isoform induction. *Int J Mol Sci.* 2025;26(15):7248. doi 10.3390/IJMS26157248/S1

Grigor'eva E.V., Kopytova A.E., Yarkova E.S., Pavlova S.V., Sorogina D.A., Malakhova A.A., Malankhanova T.B., Baydakova G.V., Zakharova E.Y., Medvedev S.P., Pchelina S.N., Zakian S.M. Biochemical characteristics of iPSC-derived dopaminergic neurons from N370S *GBA* variant carriers with and without Parkinson's disease. *Int J Mol Sci.* 2023;24(5). doi 10.3390/IJMS24054437

Grigor'eva E.V., Malakhova A.A., Yarkova E.S., Minina J.M., Vyatkin Y.V., Nadtochy J.A., Khabarova E.A., Rzaev J.A., Medvedev S.P., Zakian S.M. Generation and characterization of two induced pluripotent stem cell lines (ICGi052-A and ICGi052-B) from a patient with frontotemporal dementia with parkinsonism-17 associated with the pathological variant c.2013T>G in the *MAPT* gene. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Gen Breed*. 2024;28(7):679-687. doi 10.18699/VJGB-24-76

Harbut E., Makris Y., Pertsemlidis A., Bleris L. The history, landscape, and outlook of human cell line authentication and security. *SLAS Discov.* 2024;29(8):100194. doi 10.1016/J.SLASD.2024.100194

Kishino Y., Seki T., Fujita J., Yuasa S., Tohyama S., Kunitomi A., Tabei R., Nakajima K., Okada M., Hirano A., Kanazawa H., Fukuda K. Derivation of transgene-free human induced pluripotent stem cells from human peripheral T cells in defined culture conditions. *PLoS One*. 2014;9(5):e97397. doi 10.1371/journal.pone.0097397

Khomyakova E.A., Fedorenko A.V., Surdina A.V., Volovikov E.A., Belikova L.D., Zerkalenkova E.A., Lagarkova M.A., Bogomazova A.N. Derivation of induced pluripotent stem cells line (RCPCMi009-A-1) with knockout of the *UBE2A* gene by using CRISPR/Cas9 genome editing. *Russ J Dev Biol.* 2023;54(6):365-373. doi 10.1134/S1062360423060048

Kreslavsky T., Garbe A., Krueger A., von Boehmer H. T cell receptor-instructed alphabeta versus gammadelta lineage commitment revealed by single-cell analysis. *J Exp Med.* 2008;205(5):1173-1186. doi 10.1084/jem.20072425

Kreslavsky T., Gleimer M., von Boehmer H. Alphabeta versus gammadelta lineage choice at the first TCR-controlled checkpoint. *Curr Opin Immunol.* 2010;22(2):185-192. doi 10.1016/j.coi.2009.12.006

Mahe E., Pugh T., Kamel-Reid S. T cell clonality assessment: past, present and future. *J Clin Pathol.* 2018;71(3):195-200. doi 10.1136/jclinpath-2017-204761

Martin A.R., Williams E., Foulger R.E., Leigh S., Daugherty L.C., Niblock O., Leong I.U.S., … Kasperaviciute D., Smedley D., Caulfield M., Rendon A., McDonagh E.M. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet.* 2019;51(11):1560-1565. doi 10.1038/S41588-019-0528-2

Maslennikova A., Kruglova N., Kalinichenko S., Komkov D., Shepelev M., Golubev D., Siniavin A., Vzorov A., Filatov A., Mazurov D. Engineering T-cell resistance to HIV-1 infection via knock-in of peptides from the heptad repeat 2 domain of gp41. *mBio*. 2022;13(1):e0358921. doi 10.1128/mbio.03589-21

Mills J.A., Wang K., Paluru P., Ying L., Lu L., Galvão A.M., Xu D., … Mostoslavsky G., Kotton D.N., French D.L., Weiss M.J., Gadue P. Clonal genetic and hematopoietic heterogeneity among human-induced pluripotent stem cell lines. *Blood*. 2013;122(12):2047-2051. doi 10.1182/blood-2013-02-484444

Nishimura T., Kaneko S., Kawana-Tachikawa A., Tajima Y., Goto H., Zhu D., Nakayama-Hosoya K., … Iwamoto A., Koseki H., Nakanishi M., Eto K., Nakauchi H. Generation of rejuvenated antigen-specific T cells by reprogramming to pluripotency and redifferentiation. *Cell Stem Cell*. 2013;12(1):114-126. doi 10.1016/j.stem.2012.11.002

Okita K., Yamakawa T., Matsumura Y., Sato Y., Amano N., Watanabe A., Goshima N., Yamanaka S. An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells*. 2013;31(3):458-466. doi 10.1002/stem.1293

Podvysotskaya V.S., Grigor'eva E. V., Malakhova A.A., Minina J.M., Vyatkin Y. V., Khabarova E.A., Rzaev J.A., Medvedev S.P., Kovalenko L. V., Zakian S.M. Generation and characterisation of seven induced pluripotent stem cell lines from two patients with Parkinson's disease carrying the pathological variant c.1087G>T of the *LGR4* gene. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Gen Breed*. 2025;29(1):15-25. doi 10.18699/VJGB-25-03

Richards S., Aziz N., Bale S., Bick D., Das S., Gastier-Foster J., Grody W.W., Hegde M., Lyon E., Spector E., Voelkerding K., Rehm H.L. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Gen Med.* 2015;17(5):405-424. doi 10.1038/gim.2015.30

Seki T., Yuasa S., Oda M., Egashira T., Yae K., Kusumoto D., Nakata H., … Okada Y., Seimiya H., Fusaki N., Hasegawa M., Fukuda K. Generation of induced pluripotent stem cells from human terminally differentiated circulating t cells. *Cell Stem Cell*. 2010;7(1):11-14. doi 10.1016/j.stem.2010.06.003

Sherwood A., Desmarais C., Livingston R., Andriesen J., Haussler M., Carlson C., Robins H. Deep sequencing of the human TCRγ and TCRβ repertoires suggests that TCRβ rearranges after αβ and γδ T cell commitment. *Sci Transl Med.* 2011;3(90):90ra61. doi 10.1126/scitranslmed.3002536

Summers R.A., Fagiani F., Rowitch D.H., Absinta M., Reich D.S. Novel human iPSC models of neuroinflammation in neurodegenerative disease and regenerative medicine. *Trends Immunol.* 2024;45(10):799-813. doi 10.1016/j.it.2024.08.004

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):861-172. doi 10.1016/j.cell.2007.11.019

Vlasov I.N., Alieva A.K., Novosadova E.V., Arsenyeva E.L., Rosinskaya A.V., Partevian S.A., Grivennikov I.A., Shadrina M.I. Transcriptome analysis of induced pluripotent stem cells and neuronal progenitor cells, derived from discordant monozygotic twins with Parkinson's disease. *Cells*. 2021;10(12):3478. doi 10.3390/cells10123478

Watanabe T., Yamazaki S., Yoneda N., Shinohara H., Tomioka I., Higuchi Y., Yagoto M., Ema M., Suemizu H., Kawai K., Sasaki E. Highly efficient induction of primate iPS cells by combining RNA transfection and chemical compounds. *Genes Cells*. 2019;24(7):473-484. doi 10.1111/gtc.12702

Willmann C.A., Hemeda H., Pieper L.A., Lenz M., Qin J., Joussen S., Sontag S., Wanek P., Denecke B., Schüler H.M., Zenke M., Wagner W. To clone or not to clone? Induced pluripotent stem cells can be generated in bulk culture. *PLoS One*. 2013;8(5):e65324. doi 10.1371/journal.pone.0065324

Yamanaka S. Pluripotent stem cell-based cell therapy – promise and challenges. *Cell Stem Cell*. 2020;27(4):523-531. doi 10.1016/j.stem.2020.09.014

Yun J., So J., Jeong S., Jang J., Han S., Jeon J., Lee K., Jang H.R., Lee J. Transcriptome and epigenome dynamics of the clonal heterogeneity of human induced pluripotent stem cells for cardiac differentiation. *Cell Mol Life Sci.* 2025;82(1):2. doi 10.1007/S00018-024-05493-9

Conflict of interest. The authors declare no conflict of interest.
Received October 3, 2025. Revised November 26, 2025. Accepted November 27, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):372-380

doi 10.18699/vjgb-26-41

Development of a molecular marker for the *Run8* gene for the selection of barley genotypes resistant to smut

E.A. Orlova [1], N.P. Bechtold [1], Yu.N. Grigoriev [1], O.Yu. Shoeva [2], A.Yu. Glagoleva [2], T.V. Kukoeva [2]

[1] Siberian Research Institute of Plant Production and Breeding – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Krasnoobsk, Novosibirsk region, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
Orlova.Lena10@yandex.ru

Abstract. Loose smut of barley, caused by the basidiomycete *Ustilago nuda* (Jens.) Roster, occurs in all regions of the world where this crop is grown. This seed-borne disease causes significant losses in grain production. Selection for resistance to loose smut based on the use of donors with resistance genes is an ecologically and economically safe way to constrain the negative impact of the pathogen on barley. The introduction of molecular genetic approaches into the breeding process makes it possible to control the transfer of resistance genes to hybrid material. The *Run8* gene controls resistance to many isolates of loose smut, including in the West Siberian region of Russia. The objective of the current study is to develop a molecular marker for *Run8* for the selection of barley genotypes resistant to loose smut from hybrid populations. By comparing the nucleotide sequences of the *Run8* gene available from the barley pangenome database, an insertion/deletion of six nucleotide pairs in the coding region of the gene was identified. Based on the identified polymorphism, a molecular marker *Hor7050* was developed, which allows differentiating the alleles of *Run8*. The developed marker was tested on hybrid lines (F_5–F_6) obtained from crossing cultivar Elf, which is a donor of resistance to loose smut and carries, according to the originators, the *Run8* gene, with cultivar Tanai, which has practical resistance to the pathogen. Using the developed marker, 18 hybrids carrying *Run8* of Elf were selected from 84 hybrids; however, the phytopathological assessment showed that eight of the selected lines were susceptible to the disease. To clarify the genotype of 18 selected lines, an additional analysis was carried out using the microsatellite marker *EBmac0541* linked to *Run6*. A relationship was established between the presence of the allele of this marker from Elf and resistance to the disease. It is possible that Elf, in addition to *Run8*, carries *Run6*, which is effective against race 1 of the causative agent of loose smut. Additional studies are required to clarify the presence of *Run6* in the Elf variety. In addition to resistance, the selected lines were characterized by productivity traits. According to the two-year analysis, three productive resistant lines were identified, with *Run8* – 32, 65 and 79, significantly exceeding the control Elf in yield. The selected lines were transferred to breeding nurseries for their further evaluation by economically important traits.

Key words: barley; loose smut; resistance genes; markers; lines

For citation: Orlova E.A., Bechtold N.P., Grigoriev Yu.N., Shoeva O.Yu., Glagoleva A.Yu., Kukoeva T.V. Development of a molecular marker for the *Run8* gene for the selection of barley genotypes resistant to smut. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):372-380. doi 10.18699/vjgb-26-41

Funding. The development of molecular markers for the *Run6* and *Run8* genes was supported by the Russian Science Foundation (project No. 25-26-20151) and the Ministry of Science and Innovation Policy of the Novosibirsk Region (project No. 30-2025-000964 dated May 21, 2025).

Acknowledgements. The evaluation of barley hybrid lines for resistance to loose smut was carried out with financial support from the state budget-funded project of ICG SB RAS (project No. FWNR-2022-0018). The authors thank G.V. Generalova (ICG SB RAS) for assistance.

Разработка молекулярного маркера к гену *Run8* для отбора устойчивых к пыльной головне генотипов ячменя

Е.А. Орлова [1], Н.П. Бехтольд [1], Ю.Н. Григорьев [1], О.Ю. Шоева [2], А.Ю. Глаголева [2], Т.В. Кукоева [2]

[1] Сибирский научно-исследовательский институт растениеводства и селекции – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, р.п. Краснообск, Новосибирская область, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
Orlova.Lena10@yandex.ru

Аннотация. Пыльная головня ячменя, вызываемая базидиомицетом *Ustilago nuda* (Jens.) Roster, встречается во всех регионах, где возделывается эта культура. Селекция на резистентность к пыльной головне, основанная на использовании доноров, обладающих генами устойчивости, представляет экологически и экономически

безопасный способ, сдерживающий отрицательное действие патогена на ячмень. Внедрение в селекционный процесс молекулярно-генетических подходов позволяет контролировать передачу генов резистентности в гибридный материал. Ген *Run8* контролирует устойчивость ко многим изолятам пыльной головни и служит источником высокой устойчивости к этому возбудителю. Цель представленной работы – разработка молекулярного маркера к гену *Run8* для отбора из гибридных популяций генотипов ячменя, устойчивых к пыльной головне. С помощью сравнения доступных из базы данных пангенома ячменя нуклеотидных последовательностей гена *Run8* определена инсерция/делеция шести пар нуклеотидов в кодирующей области гена. На основе выявленного полиморфизма разработан молекулярный маркер *Hor7050*, позволяющий дифференцировать аллели гена *Run8*. Маркер был протестирован на гибридных линиях F_5–F_6, полученных от скрещивания сорта Эльф – донора резистентности к пыльной головне – и несущего, по данным оригинаторов, ген *Run8*, с сортом Танай, обладающим практической устойчивостью к возбудителю. С помощью разработанного маркера из 84 гибридов были отобраны 18, несущих ген *Run8* сорта Эльф. Однако проведенная фитопатологическая оценка показала, что из выделенных линий восемь восприимчивы к заболеванию. Для уточнения генотипа 18 отобранных линий был проведен дополнительный анализ с помощью микросателлитного маркера *EBmac0541*, сцепленного с геном *Run6*. Была установлена взаимосвязь между наличием аллеля данного маркера от сорта Эльф и устойчивостью к заболеванию. Возможно, сорт Эльф, кроме гена *Run8*, несет ген *Run6*, эффективный против расы 1 возбудителя пыльной головни. Для уточнения наличия гена *Run6* у сорта Эльф требуются дополнительные исследования. Помимо устойчивости, отобранные линии были охарактеризованы по признакам продуктивности. По результатам двухлетнего анализа выделены три продуктивные резистентные линии с геном *Run8* – 32, 65 и 79, достоверно превышающие по урожайности контрольный сорт Эльф.

Ключевые слова: ячмень; пыльная головня; гены устойчивости; маркеры; линии

Introduction

Barley is one of the main forage cereals. Its high plasticity and short growing season allow it to be cultivated in a wide range of soil and climatic zones. Modern barley breeding is aimed at developing high-yielding varieties resistant to biotic and abiotic stress factors. Diseases caused by phytopathogenic fungi, with loose smut being the most widespread and harmful, hinder the achievement of high and stable yields.

Loose smut of barley, caused by the basidiomycete *Ustilago nuda* (Jens.) Roster occurs in all regions of the world where this crop is grown (Nielsen, Thomas, 1996). This seed-borne disease causes significant losses in grain production. When infected by the pathogen, a smut sorus containing teliospores forms instead of a spike. Thus, the damage from loose smut is proportional to the percentage of infected spikes in the field (Druzhin, Krupnov, 2008). However, economic losses are significantly higher. According to many authors, the pathogen has a suppressive effect at all stages of plant development. As a result of seedling death, field germination rate decreases; loose smut negatively affects the number of fertile tillers per plant, plant height, main spike length, number of grains per spike, and 1,000 grain weight (Orlova et al., 2015; Usoltsev, 2018). Most spring barley varieties included in the State Register of Varieties and Hybrids approved for use in the Russian Federation (https://gossortrf.ru/registry), are susceptible to loose smut to varying degrees. Control of the disease involves mandatory seed treatment with systemic fungicides. Therefore, cultivation of resistant varieties is preferable from an economic standpoint. On top of that, the value of resistant varieties is that they can be used in low-input organic farming systems to produce environmentally safe products without additional protective measures.

Throughout the years, many researchers have confirmed that genetic resistance of spring barley to loose smut can be controlled by individual dominant or several recessive genes (Krivchenko, 1984; Menzies et al., 2010). The loose smut resistance gene *Run* was first identified by J.E. Livingston (1942) in the variety Trebi. Later, he also identified a weak resistance gene *Run2* in Missuri variety. W.P. Skoropad and L.P. Johnson (1952) identified two independent dominant genes *Run3* and *Run6* in Jet variety. The latter was used as the base for the Keystone variety carrying the *Run*6 gene. C.W. Schaller (1949) identified two dominant genes *Run4* and *Run5* in barley variety Dorsett and X173-10-5-6 hybrid. In Anoidium cultivar, the resistance is controlled by the recessive gene *run7*. The *Run8* gene was isolated by D.R. Metcalfe (1966) from the winter barley line C.I. 4966 introduced to the United States from Russia. To date, 15 resistance genes to the loose smut pathogen have been identified in barley (Zang et al., 2015; Legkun et al., 2016), but according to the literature, only three of them (*Run3*, *Run6*, and *Run8*) are effective. The remaining genes are of low effectiveness and have not found practical use in breeding (Krivchenko, 1984).

Assessment of barley varieties for resistance to loose smut under natural infection conditions is not always reliable, which can lead to incorrect classification of varieties by resistance level. On the other hand, artificial inoculation is a labor-intensive and time-consuming process that spans two growing seasons as follows: in the first year, spikes are inoculated with a pathogen suspension and mature seeds are collected; in the following year, plants grown from the inoculated seeds are evaluated for resistance based on phenotypic expression. Moreover, the quality of infection largely depends on meteorological conditions during inoculation. According to L.V. Meshkova (2024), there

is a high correlation between moisture, precipitation, and infection ($r > 0.7$), and a moderate negative correlation with air temperature ($r = –0.44$). Therefore, an effective method to develop resistant varieties is to combine the classical approaches based on hybridization of resistance gene donors with recipient varieties and modern molecular marker-aided breeding methods. This approach helps reduce the time and volume of unpromising hybrid material in the breeding process.

To determine the resistance of samples, it is necessary to consider the prevalent pathogen races in the target cultivation zone of future varieties. This is the only way to ensure staying one step ahead of the pathogen in the development of resistant genotypes. Understanding the genetics of resistance in the context of pathogen virulence makes it possible to identify the effective genes, for which geneticists can select DNA markers necessary for developing hybrids with high disease resistance based on genotype. The *Run8* gene confers resistance to most races of the loose smut pathogen and is widely used in barley breeding both in Russia and abroad. This gene was mapped to the long arm of chromosome 1H (Zang et al., 2015). It codes for a protein containing two tandem protein kinase domains, which, as recently discovered, represent an important class of proteins involved in plant resistance. Polymorphism in the *Run8* protein sequences has been identified, which allows the development of intragenic diagnostic molecular markers for precise identification of *Run8* gene alleles. Before the nucleotide sequence of the *Run8* gene was discovered, molecular markers linked to this gene were used in scientific research and breeding practice (Li et al., 2001; Eckstein et al., 2002). However, the diagnostic efficiency of the latter is lower than that of intragenic markers due to recombination between the linked markers and the target gene. Therefore, the development of intragenic molecular markers based on the known nucleotide sequences of target genes becomes a relevant research problem. The goal of the present paper was to develop an intragenic molecular marker for the *Run8* gene and use it in combination with a marker linked to the *Run6* gene for genotyping barley samples, so that the lines with effective resistance genes demonstrating high loose smut resistance could be singled out from hybrid populations.

Materials and methods

Plant material. The study material consisted of 84 breeding lines selected in the F_5–F_6 generations from the Elf × Tanai cross. Elf variety is highly resistant to loose smut with the pedigree as follows: Roland × 1325 line, the latter being a dihaploid (Pervenets × Zazersky 85) F_1 × *H. Bulbosum*, which carries the *Run8* resistance gene presumably inherited from the Pervenets variety, according to the variety originators (State Register of Protected Selection Achievements). Tanai variety exhibits practical resistance to the pathogen and was developed from two breeding lines, G 20275 × G 20191, with Jet (*Run3 Run6*) and Paragon (*Run3*) varieties present in the pedigree.

The research was conducted from 2021 to 2023. In 2021, the 84 test lines were grown under field conditions in the SP1 breeding nursery. The lines selected for resistance were then reinoculated and sown in the phytopathological field of SibRIPP&B – Branch of IC&G SB RAS.

Differentiation of the loose smut pathogen *U. nuda*. To study the race composition of loose smut, an empirical set of differential varieties and a genetic set of test varieties with known resistance genes proposed by V.I. Krivchenko (1984) were used. Race identification was performed by comparing the reaction of differential varieties with a key for determining physiological races. The virulence formula of the race was determined based on the susceptibility of the genetic test set. Resistance reactions were classified according to the international scale: R standing for resistance (infection up to 10 %); S for susceptibility (infection above 10 %).

Assessment of line resistance to loose smut. Lines with confirmed resistance genes and the parental varieties Elf and Tanai were inoculated in June 2021 with the Novosibirsk population of loose smut. The susceptible indicator variety Grace was used to control inoculation quality. In 2022, the lines that showed resistance were reinoculated. Inoculation of barley spikes (at least eight per line) was carried out using the vacuum method during the early flowering stage, according to the VIR methodology (Sanin et al., 2008). The spore suspension was prepared immediately before inoculation at a concentration of 1 g of spores per 1 l of water. Smut spores were collected from infected spikes of various barley genotypes. To determine the race composition of the pathogen, differential and genetic test sets of barley varieties were inoculated simultaneously.

Infected seeds (at least 100 infected grains) were sown in the phytopathological field of SibRIPP&B – Branch of IC&G SB RAS, located in Michurinsky settlement, Novosibirsk region. Soil treatment included spring harrowing, nitrogen fertilization, and cultivation. Trials were established on fallow land. The optimal timeframe, with the soil temperature of 11.5–16.5 °C at a depth of 5–8 cm, which is favorable for both seed and pathogen germination, was chosen for sowing infected seeds.

Loose smut resistance classification of the tested samples was performed using the VIR scale (Sanin et al., 2008) by counting healthy and diseased spikes and calculating the infection percentage during the full heading to early ripening stages. When determining the race composition of the loose smut population, resistance classification was based on the international scale: R standing for resistance (infection up to 10 %); S for susceptibility (infection above 10 %).

Evaluation of lines for agronomic traits. To assess agronomically valuable traits, a structural analysis of the barley breeding lines was conducted in the autumn of 2022–2023. The traits evaluated included plant height,

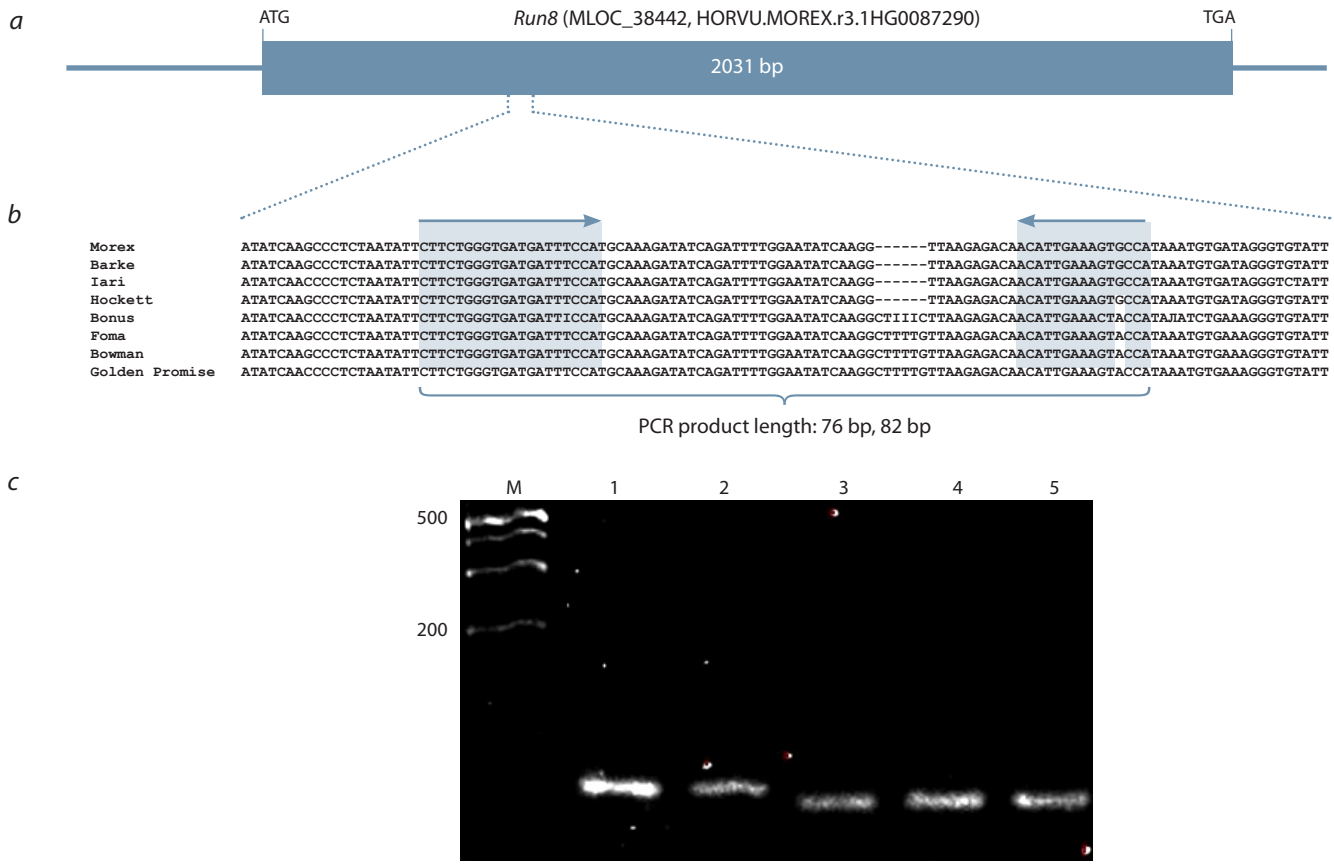


Fig. 1. Structural organization of the *Run8* gene (*a*), multiple alignment of nucleotide sequences in the gene region containing a 6 bp insertion/deletion, for which the DNA marker was developed (*b*), and results of marker testing in barley varieties with known resistance genes to loose smut: Raushan *Run8*, *Run15* (1), Elf *Run8* (2), Jet *Run3*, *Run6* (3), Keystone *Run6* (4), Grace as a susceptibility standard (5). Primer annealing sites are highlighted in gray, and the expected lengths of PCR fragments for the marker are indicated (*c*).

fertile tillers per plant, spike length, number of grains per spike, and 1,000 grain weight. The evaluation was carried out according to the Methodology of the state variety testing of agricultural crops (1985).

Development of a molecular marker for the *Run8* gene. The molecular marker *Hor7050* for the *Run8* gene was developed based on the nucleotide sequence MLOC_38442 (HORVU.MOREX.r3.1HG0087290, HORVU1Hr1G087050) (Zang et al., 2015) (Fig. 1*a*). This sequence was used as a reference in the barley pangenome database (Jayakodi et al., 2024) available on the GrainGenes web page (https://wheat.pw.usda.gov/GG3/), and, as a result, nucleotide sequences of the *Run8* gene were found in eight barley samples (Barke, Bonus, Bowman, Foma, Golden Promise, Hockett, Igri, Morex). Multiple sequence alignment was performed using the Multalin v5.4.1 software (Corpet et al., 1988), which revealed polymorphisms among different alleles of this gene. In addition to single nucleotide substitutions, a six-nucleotide deletion was identified in four of the eight varieties. To genotype this deletion, a pair of primers was selected using the IDT PrimerQuest™ (http://eu.idtdna.com/PrimerQuest/Home), with annealing sites flanking the identified deletion (Fig. 1*b*). Testing the developed marker on barley samples carrying the known loose smut resistance genes showed that electrophoretic analysis in 5 % HR agarose gel made it possible to distinguish the *Run8* gene alleles. In resistant barley samples carrying the *Run8* gene, the target amplification fragment was 82 bp, and in the susceptible ones 76 bp (Fig. 1*c*).

By comparing the genotyping data with the resistance of the lines to loose smut, the diagnostic efficiency of the marker was evaluated. It was defined as the proportion of correct test results in the total number of results (i. e., the sum of true positive and true negative results divided by the total number of results).

Genotyping of lines for the *Run6* and *Run8* genes. DNA was extracted from leaves at the tillering stage collected from five plants per line using the method described earlier (Plaschke et al., 1995). DNA was analyzed using the microsatellite marker *EBmac0541* linked to the *Run6* gene (Menzies et al., 2010) and the intragenic marker *Hor7050* developed for the *Run8* gene developed in the present paper (Table 1). Polymerase chain reaction (PCR) with the *Hor7050* marker was performed in a 20 µl reaction mixture containing 100 ng of DNA template, 1X PCR buffer (67 mM Tris-HCl, pH 8.8, 1.5 mM $MgCl_2$, 0.01 %

Table 1. DNA markers used in this paper for genotyping barley lines for the *Run6* and *Run8* genes

Gene	Chromosomal location	Marker name (type)	Primer nucleotide sequence, 5′→3′	PCR product length, bp*	Reference
Run6	3H	*EBmac0541* (linked)	F: acggatctactttagctagca	106 (S), 110+222 (R)	Rumsay et al., 2000
			R: aaacaaccccacacaatc		
Run8	1H	*Hor7050* (intragenic)	F: cttctgggtgatgatttcca	76 (S), 82 (R)	This paper
			R: tggtactttcaatgttgtctc		

* Presented here are the PCR product lengths characteristic of barley samples susceptible (S) and resistant (R) to loose smut.

Tween 20, 18 mM $(NH_4)_2SO_4$), 0.2 mM of each dNTP, 0.25 μM of each forward and reverse specific primer, and 1 unit of Taq DNA polymerase (Helicon, Moscow). PCR with the *EBmac0541* marker was performed using the BioMaster HS-Taq PCR-Color (2×) kit (Biolabmix, Novosibirsk). Amplification with the specified primers was carried out under the conditions as follows: initial denaturation for 1 min 50 s at 94 °C; denaturation for 30 s at 94 °C; primer annealing for 30 s at 50 °C (*Hor7050* marker) or 55 °C (*EBmac0541* marker), polymerization for 45 s at 72 °C, number of cycles – 45, and final elongation for 5 min at 72 °C. PCR products were separated in 5 % HR agarose gel HR Agarose PCR Grade, HydraGen, NJ, USA) in a horizontal chamber for 3–4 hours at 7 V/cm. UV imaging and gel analysis were performed using the Gel Doc™ XR+ System (Bio-Rad Laboratories, Inc., Hercules, CA, USA).

Results

A characteristic feature of the 2022–2023 growing seasons was insufficient moisture in May. In 2022, the average daily air temperature in May was +15.4 °C, which is 4.5 °C above the long-term average. In May 2023, although the average monthly temperature was close to the long-term norm, the distribution of heat across decades was uneven. In the second decade, the air temperature was 2.3 °C below the long-term average, while in the third decade it was 3.4 °C above that. The average soil temperature at the depth of 5–8 cm during sowing was 14 °C, which is favorable for the infection process of smut fungi during seed germination. The average air temperatures from May to August were as follows: 15.3, 17.3, 18.5, 16.6 °C in 2022; 11.8, 19.0, 21.6, 17.8 °C in 2023.

In 2022, the total seasonal precipitation was 153.8 mm, which is 30 % as low as the long-term average. In 2023, it was 204 mm, of which 173.8 mm fell in the third decade of July and the second and third decades of August.

Differentiation of the loose smut pathogen *U. nuda*. Based on the reaction of the empirical set of differential varieties, it was established that the loose smut population in 2022–2023 was represented by race 1, which does not infect the Keystone variety. Analysis of spore samples using the genetic test set with known resistance genes revealed the presence of biotypes virulent to varieties Trebi (*Run1* gene), OAC-21 (*Run9* and *Run10*), and Moskovsky 2 (*Run15)*. Thus, the virulence formula of race P1 is 1 – 1.9.10.15 (Supplementary Tables S1 and S2)[1].

Selection of breeding lines using the molecular marker for the *Run8* gene and assessment of their resistance to loose smut. The differences between the *Run8* gene alleles in Elf and Tanai varieties were identified using the *Hor7050* marker. A PCR product of 82 bp was obtained on the Elf DNA, and 76 bp on Tanai DNA. Due to the presence of polymorphism between the parental varieties, this marker was used for genotyping 84 F_5 breeding lines derived from the Elf × Tanai hybrid population. As a result, 18 lines were identified as having inherited the *Run8* gene from the Elf variety.

According to artificial inoculation data from 2021, among the 18 selected breeding lines carrying the *Run8* gene from Elf variety, high resistance to loose smut was confirmed in eight lines as follows: 12, 14, 25, 32, 45, 49, 78, and 79. Two lines, specifically 42 and 65, were classified as practically resistant (infection of less than 5 %). Thus, resistance to *U. nuda* was confirmed by phytopathological methods in 55.6 % of the lines, while the remaining lines showed varying degrees of susceptibility, from weak to strong (Table 2). The diagnostic efficiency of the *Hor7050* marker was 0.56.

In 2022, the selected F_6 generation lines were additionally genotyped using the *Hor7050* marker, which confirmed the presence of the *Run8* gene in resistant lines 14, 32, 42, 45, 49, 65, and 79. Unlike the Elf variety, which demonstrated stable resistance to loose smut over three years, not all selected lines carrying the *Run8* gene from this variety were resistant to the disease. Lines 4, 8, 10, 44, 57, 60, 71, and 81 showed infection levels ranging from 17.9 to 66.7 %. Since the pedigree of the parental variety Tanai includes the Jet variety known as a carrier of the *Run6* gene, the selected lines were genotyped using the microsatellite marker *EBmac541* linked to the *Run6* gene (Fig. 2).

Out of the 18 analyzed lines, nine inherited both alleles of the *EBmac541* marker from the Tanai variety (lines 4, 8, 10, 42, 44, 57, 60, 79, and 81), six from the Elf variety (lines 12, 14, 25, 32, 71, and 78), and three lines exhibited

[1] Supplementary Tables S1 and S2 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Orlova_Engl_30_3.pdf

Table 2. Resistance to loose smut in breeding lines selected from the Elf × Tanai cross and genotyping data for these lines using molecular markers for the *Run8* and *Run6* genes

Variety/line	Maximum loose smut infection, %, 2022–2023	*Hor7050* (*Run8*)		*EBmac541* (*Run 6*)
		2021	2022	
4	19.5	–	82	106
8	29.7	82	–	106
10	56.7	82	82	106
12	0	82	–	110 + 222
14	0	82	82	110 + 222
25	0	82	–	110 + 222
32	0	82	82*	110 + 222
42	3.4	82	82	106
44	62.8	82	–	106
45	0	82	82	106, 110 + 222
49	0	82	82	106, 110 + 222
57	52.8	82	–	106
60	66.7	82	–	106
65	4.1	82	82	106, 110 + 222
71	17.9	82	82	110 + 222
78	0	82	–	110 + 222
79	0	82	82	106
81	28.4	82	82	106
Tanai	1.2	76	76	106
Tanai	0	82	82	110 + 222

Note. "–" indicates that no analysis was performed; * indicates that the allele could not be definitively identified as homozygous or heterozygous.

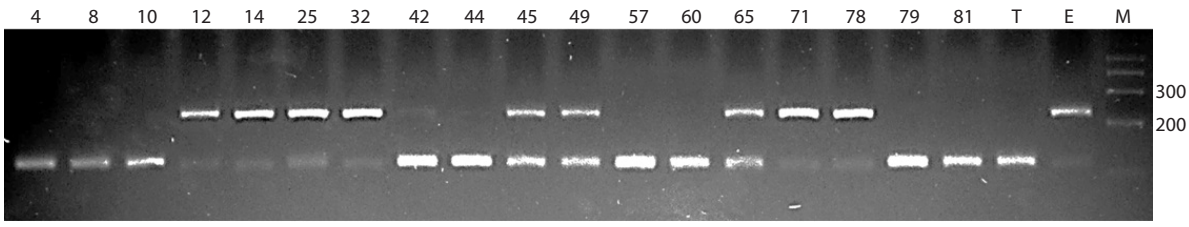


Fig. 2. Genotyping results for the selected lines obtained using the microsatellite marker *EBmac541*. Numbers above indicate line numbers, T stands for Tanai, E for Elf, and M is the 100 bp size marker (bp lengths of length marker fragments are shown on the side).

marker heterozygosity (lines 45, 49, and 65). By comparing the genotyping data with the resistance of the studied lines to loose smut, it was found that among the ten resistant lines (infection rate from 0 to 4.1 %), eight carried the *EBmac541* marker allele from the Elf variety in either homozygous or heterozygous state. Among the eight susceptible lines (infection rate from 17.9 to 66.7 %), only line 71 carried the Elf allele of the marker, while the others inherited the allele from Tanai. Based on these results, the diagnostic efficiency of the *EBmac541*, marker turned out to be 0.83.

Evaluation of selected lines for agronomic traits. To assess agronomically valuable traits, a structural analysis

Table 3. Main agronomic traits of breeding lines from the Elf × Tanai cross, F_5 generation (average for 2022–2023)

Variety/Line	Plant height, cm	Main spike length, cm	Number of grains per spike	1,000 grain weight, g	Fertile tillers per plant	Yield, g/m^2
Elf (control)	59.3	6.5	19.2	48.0	2.5	419.2
Tanai	69.5	6.8	19.4	50.05	2.8	724.4*
Line 12	62.4	5.9	16.3	51.0	3.6*	509.6
Line 14	65.4	6.8	19.4	48.9	2.8	593.9
Line 25	64.9	6.2	19.2	47.4	2.5	480.4
Line 32	70.3*	7.5*	20.9*	45.2	2.8	763.7*
Line 42	71.5*	6.5	17.3	51.8	2.8	568.5
Line 45	68.2	6.8	19.8	47.4	3.0	516.0
Line 49	60.0	6.2	17.3	44.5	2.9	587.2
Line 65	70.7*	7.8*	21.9*	45.8	3.2*	661.7*
Line 78	72.2*	6.8	18.0	50.2	2.8	521.3
Line 79	75.3*	7.6*	21.6*	47.1	2.9	712.0*
LSD_{05}	10.681	0.308	0.992	4.3997	0.646	219.8

* Indicates statistically significant differences compared to the parental variety Elf.

was conducted in the autumn of 2022–2023 for the resistant barley breeding lines carrying the *Run8* gene. Both biometric and productivity-related traits were evaluated. The parental variety Elf, carrying the *Run8* gene, was used as a control.

All tested lines outperformed the control variety Elf in plant height. Lines 32, 42, 65, 78, and 79 also surpassed the paternal form Tanai, with average heights over two years ranging from 70.0 to 75.3 cm. The remaining lines showed an intermediate performance between the parental forms. Spike length is known to positively influence productivity, as it has a high correlation with the number of grains per spike (Naumova, 2021). In the selected lines, spike length ranged from 5.9 (line 12) to 7.8 cm (line 65). Lines 32, 65, and 79 significantly exceeded both parental forms in spike length and number of grains per spike. Lines 12, 42, 49, and 78 showed low spike productivity (Table 3).

Grain size is another important yield component. In 2022, the highest 1,000 grain weight of 54.5 g was observed in line 49 compared to 51.1 g in Elf. Over two years, grain sizes comparable to Tanai were recorded in lines 12, 42, and 78 reaching 51.0, 51.8, and 50.2 g, respectively.

The number of fertile tillers per plant in lines 12 and 65 was 1.4 to 1.3 times as high as that of the control variety Elf reaching 3.6–3.2 fertile stems per plant.

Based on the two-year data, three productive lines were identified, namely 32, 65, and 79, which significantly outperformed the control variety Elf in yield. In line 65, yield varied by year from 809.7 g/m^2 in 2022 to 513.7 g/m^2 in 2023. Line 32 showed the most stable yield, with 756.0–771.3 g/m^2 across years, and on average over two years, it exceeded the best parental form, Tanai. Line 79 stood out for its earliness and productivity.

Thus, as a result of molecular genetic analysis, 18 breeding lines carrying the *Run8* loose smut resistance gene were selected from the Elf × Tanai (F5) hybrid combination. Resistance in 10 of these lines was confirmed by phytopathological testing. Lines 12, 14, 25, 32, 42, 45, 49, 65, 78, and 79 were classified as highly or practically resistant barley genotypes. Based on structural analysis data, the standout hybrid lines 32, 65, and 79 were transferred to breeding nurseries for further evaluation of agronomically important traits.

Discussion

Due to the specific life cycle of the basidiomycete *U. nuda* (Jens.) Roster, phenotyping for resistance to the disease it causes, i. e. loose smut, is a lengthy and labor-intensive process. Therefore, diagnostic molecular markers gain special relevance in breeding for resistance to the disease. However, to date, many *Run* genes that control resistance to loose smut have not yet been mapped in the barley genome (Abo-Elyousr et al., 2022). Among the genes with known chromosomal locations, only two, namely *Run6* and *Run8*, have been mapped to chromosomes 3H and 1H, respectively (Menzies et al., 2010; Zang et al., 2015). Precise mapping has traced the latter back to a specific nucleotide sequence coding for a protein kinase (Zang et al., 2015). Long-term

testing of varieties with identified resistance genes has shown that *Run6* and *Run8* were effective against loose smut races in Western Siberia. In the present paper, the source of the *Run8* gene, namely the Elf variety, was used as a donor in the development of loose smut-resistant lines. To select resistant lines, the *Hor7050* molecular marker was developed based on the nucleotide sequence of the *Run8* gene identified by W. Zang et al. (2015). When comparing the amino acid sequences of protein kinases predicted from *Run8* gene sequences, it was shown in the present paper that the group of loose smut-resistant samples was characterized by 15 unique amino acid substitutions. These substitutions were caused by single nucleotide polymorphisms (SNPs), the detection of which by routine PCR is labor-intensive and requires the development of specific primers, probes, or selection of restriction endonucleases. In this paper, a common six-nucleotide deletion was identified by comparing *Run8* gene sequences from the barley pangenome database. This deletion results in the loss of two amino acid residues in the protein kinase. When comparing *Run8* protein sequences, the deletion of two amino acid residues was found in loose smut-susceptible barley samples, whereas the presence of these two amino acids was observed in both resistant and susceptible samples (Zang et al., 2015). Thus, the marker developed specifically for this deletion only has partial diagnostic efficiency, since not all barley samples carrying the six-nucleotide insertion will be resistant to loose smut. However, it can be used to track the inheritance of the *Run8* allele from a known donor. The total of 18 barley lines carrying the *Run8* gene from the resistant donor Elf were selected using this marker. However, unlike the Elf variety, some of the selected lines showed susceptibility to race 1 of loose smut, as identified using differential varieties. Since the second parent, Tanai, has the Jet variety in its pedigree, which is known as a donor of the *Run6* gene, a hypothesis was posed that the resistant lines might carry the *Run6* gene from the Tanai variety in addition to *Run8* from Elf. To test this hypothesis, the selected lines were genotyped using the *EBmac0541* marker linked to the *Run6* gene (Menzies et al., 2010). However, the analysis showed that most of the lines demonstrating resistance to loose smut inherited the *EBmac0541* marker from Elf, rather than Tanai. Lines 4, 8, 10, 44, 57, 60, and 81, which inherited the *EBmac0541* marker from Tanai, were susceptible to loose smut, despite carrying the *Run8* gene from Elf. The only line that carried the *EBmac0541* marker from Tanai and was resistant to loose smut was line 79. Since *EBmac0541* is linked to *Run6* gene, the discrepancy between the marker allele and the loose smut resistance trait may be explained by recombination between the marker used for genotyping and the gene controlling the trait. The analysis performed makes it possible to assume that the Elf genome may also contain *Run6* gene effective against race 1 of loose smut in addition to *Run8*. Further analysis of *Run8* nucleotide sequences in known resistance donors, as well as the co-segregation observed in this paper between the *EBmac0541* marker from Elf and resistance to loose smut, will help clarify the roles of *Run6* and *Run8* in resistance to pathogen races prevalent in Western Siberia.

Phytopathological assessment and productivity evaluation of the selected lines over two years made it possible to identify three promising breeding lines resistant to loose smut and significantly outperforming the control variety Elf in yield. These selected lines have been transferred to breeding nurseries for further evaluation of agronomically valuable traits.

Conclusion

In the present paper, a diagnostic intragenic molecular marker was developed based on the nucleotide sequence of the *Run8* gene. Its application, in combination with the microsatellite marker *EBmac0541* linked to the *Run6* gene made it possible to identify an association between the *EBmac0541* allele from the Elf variety and resistance of barley hybrids to loose smut. As a result of the research, promising breeding lines were selected that are comparable in productivity to the original varieties.

References

Abo-Elyous K.A.M., Mourad A.M.I., Baenziger P.S., Shehata A.H.A., Eckstein P.E., Beattie A.D., Sallam A. Identification of putative SNP markers associated with resistance to egyptian loose smut race(s) in spring barley. *Genes*. 2022;13(6):1075. doi 10.3390/genes13061075

Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*. 1988;16(22):10881-10890. doi 10.1093/nar/16.22.10881

Druzhin A.E., Krupnov V.A. Wheat and Dusty Firebrand. Saratov, 2008 (in Russian)

Eckstein P.E., Krasichynska N., Voth D., Duncan S., Rossnagel B.G., Scoles G.J. Development of PCR-based markers for a gene (*Un8*) conferring true loose smut resistance in barley. *Can J Plant Pathol*. 2002;24(1):46-53. doi 10.1080/07060660109506970

Jayakodi M., Lu Q., Pidon H., Rabanus-Wallace M.T., Bayer M., Lux T., Guo Y., … Zhang X.-Q., Wicker T., Dockter C., Mascher M., Stein N. Structural variation in the pangenome of wild and domesticated barley. *Nature.* 2024;636(8043):654-662. doi 10.1038/s41586-024-08187-1

Krivchenko V.I. Resistance of Grain Ears to Pathogens of Head Diseases. Moscow, 1984 (in Russian)

Legkun I.B. Breeding and evaluation of winter barley varieties forgroup resistance to smut diseases. *Russ J Genet Appl Res*. 2016;6:264-269. doi 10.1134/S2079059716030060

Li C.D., Eckstein P.E., Lu M., Rossnagel B.G., Scoles G.J. Targeted development of a microsatellite marker associated with a true loose smut resistance gene in barley (*Hordeum vulgare* L.). *Mol Breed*. 2001;8:235-242. doi 10.1023/A:1013738108871

Livingston J.E. The inheritance of resistance to *Ustilago nuda. Phytopathology.* 1942;32:451-466

Menzies J.G., Steffenson B.J., Kleinhofs A. A resistance gene to *Ustilago nuda* in barley is located on chromosome 3H. *Can J Plant Pathol.* 2010;32(2):247-251. doi 10.1080/07060661003739977

Meshkova L.V., Sabaeva O.B., Nikolaev P.N. The dusty smut of barley in the Omsk region and sources of resistance to it. *Bulletin of NSAU.* 2024;4(73):70-78. doi 10.31677/2072-6724-2024-73-4-70-78 (in Russian)

Metcalfe D.R. Inheritance of loose smut resistance: III. Relationships between the “Russian” and “Jet” genes for resistance and genes in

10 barley varieties of diverse origin. *Can J Plant Sci*. 1966;46(5): 487-495. doi 10.4141/cjps66-082

Methodology of the State Variety Testing of Agricultural Crops. Moscow, 1985 (in Russian)

Naumova N.A. Peculiarities of grain productivity formation and its elements in varieties of spring barley collection of VIR in conditions of arid climate of Astrakhan region. *Agragnyy Nauchnyy Zhurnal = The Agrarian Scientific Journal.* 2021;(5):29-34. doi 10.28983/asj.y2021i5pp29-34 (in Russian)

Nielsen J., Thomas P. Loose smut. Bunt and smut diseases of wheat: Concepts and methods of disease management. Mexico: D.F. CIMMYT, 1996

Orlova E.A., Bekhtold N.P., Likhenko I.E. Impact of the pathogen barley smut under economically valuable characteristics of plants. *Dostizheniya nauki i tekhniki APK = Achievements of Science and Technology of AIC*. 2015;29(3):4-6 (in Russian)

Plaschke J., Ganal M.W., Röder M.S. Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor Appl Genet*. 1995;91(6-7):1001-1007. doi 10.1007/bf00223912

Ramsay L., Macaulay M., degli Ivanissevich S., MacLean K., Cardle L., Fuller J., Edwards K.J., … Marmiroli N., Sjakste T., Ganal M., Powell W., Waugh R. A simple sequence repeat-based linkage map of barley. *Genetics*. 2000;156(4):1997-2005. doi 10.1093/genetics/156.4.1997

Sanin S.S., Nekles N.P., Sanina A.A., Pakholkova E.V. Methodological recommendations on the creation of infectious backgrounds for immunological studies of wheat. Moscow, 2008 (in Russian)

Schaller C.W. Inheritance of resistance to loose smut, *Ustilago nuda*, in barley. *Phytopathology.* 1949;39(12):959-979

Skoropad W.P., Johnson L.P. Inheritance of resistance to *Ustilago nuda* in barley. *Can J Bot*. 1952;30(5):525-536. doi 10.1139/b52-038

Usoltsev Yu.A. Reduction of losses of the harvest of spring barley from head diseases. *Bull Kurgan State Agric Acad*. 2018;3:65-69 (in Russian)

Zang W., Peter E., Eckstein P.E., Colin M., Voth D., Himmelbach A., Beier S., Stein N., Scoles G.J., Beattie A.D. Fine mapping and identification of a candidate gene for the barley *Un8* true loose smut resistance gene. *Theor Appl Genet*. 2015;128(7):1343-1357. doi 10.1007/s00122-015-2510-4

Conflict of interest. The authors declare no conflict of interest.
Received June 30, 2025. Revised November 12, 2025. Accepted November 13, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):381-390

doi 10.18699/vjgb-26-42

Somaclonal variation in *Saccharum* spp.: unraveling its potential despite current neglect

S. Munir [1, 2], M.A.B. Jaffar [1], S. Yasmeen[1], M.T. Khan[3], I.A. Khan[1]

[1] Sugarcane Biotechnology Group, Nuclear Institute of Agriculture (NIA), Tando Jam, Pakistan
[2] Plant Breeding and Genetics Division, Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan
[3] Agricultural Biotechnology Division, National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan
ch.sana11@gmail.com

Abstract. Hybridization of different landraces or wild crop species facilitates genetic recombination and leads to the development of improved cultivars, particularly in sexually propagated crops. In contrast, genetic recombination via hybridization in asexually propagated crops like sugarcane (*Saccharum* spp. hybrid) is challenging due to self or cross incompatibility (low fertility). Such crops can be improved by somaclonal variation, which is achieved by tissue culture techniques. As a major contributor to global sugar and bioethanol production, sugarcane suffers substantial yield loss due to various biotic or abiotic stresses, which may be attributed to its poor resistance mechanism. Despite the potential of *in vitro* culture techniques, somaclonal variation remains underexplored in sugarcane breeding programs. To address the challenges posed to sugarcane under changing environmental dynamics, this review critically evaluates the role of somaclonal variation in sugarcane variety development, its underlying mechanism, practical applications, and factors affecting its occurrence. This review also discusses the limitations and challenges in the practical implementation of this technique in variety development, resulting in its neglect in modern breeding efforts. The focus on the potential of somaclonal variation, sustained by cutting-edge approaches, can unlock its limitations and fulfill the growing future demands of sugar, biofuel, and bioenergy industries.

Key words: callus; genetics; *in vitro* culture; markers; regeneration; somaclonal variation; sugarcane

For citation: Munir S., Jaffar M.A.B., Yasmeen S., Khan M.T., Khan I.A. Somaclonal variation in *Saccharum* spp.: unraveling its potential despite current neglect. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):381-390. doi 10.18699/vjgb-26-42

Funding. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Authors contribution. SM conceived the study and wrote the first draft of the manuscript. ABJ designed the figure and wrote the manuscript. SY, MTK and IAK critically review and edit the manuscript. All authors revised the final version.

Data availability statements. All data supporting the findings of this study are available within the paper. Data sharing is not applicable to this article, as no new data were created or analyzed in this study.

Сомаклональная изменчивость у видов рода *Saccharum*: раскрытие ее потенциала в современных условиях

С. Мунир [1, 2], М.А.Б. Джаффар [1], Ш. Ясмин[1], М.Т. Хан[3], И.А. Хан[1]

[1] Группа биотехнологий сахарного тростника, Ядерный институт сельского хозяйства (NIA), Тандо Джам, Пакистан
[2] Отдел селекции и генетики растений, Ядерный институт сельского хозяйства и биологии, Фейсалабад, Пакистан
[3] Отдел сельскохозяйственной биотехнологии, Национальный институт биотехнологии и генной инженерии, Фейсалабад, Пакистан
ch.sana11@gmail.com

Аннотация. Гибридизация различных местных сортов или диких видов сельскохозяйственных культур способствует генетической рекомбинации и приводит к созданию улучшенных сортов, в особенности у культур, размножающихся половым путем. В отличие от этого, у культур с бесполым размножением, таких как сахарный тростник (гибрид *Saccharum* spp.), генетическая рекомбинация посредством гибридизации затруднена из-за само- или перекрестной несовместимости (низкой фертильности). Такие сельскохозяйственные культуры могут быть улучшены за счет сомаклональной изменчивости с помощью методов культуры ткани. Сахарный тростник, являющийся одним из основных источников мирового производства сахара и биоэтанола, из-за недостаточной устойчивости к биотическим и абиотическим стрессовым факторам несет значительные потери урожая. Несмотря на эффективность методов культивирования *in vitro*, сомаклональная изменчивость остается недостаточно изученной в программах селекции сахарного тростника. Для решения проблем, появляющихся у сахарного тростника в условиях меняющейся динамики окружающей среды, в настоящем обзоре критически оценивается роль сомаклональной изменчивости в создании сортов сахарного тростника, рассмотрены ее основные механизмы, практическое применение и факторы, влияющие

на ее возникновение. Кроме того, обсуждаются ограничения и проблемы практического применения этого метода при создании сортов, не позволяющие адекватно использовать его в современных селекционных работах. Использование потенциала сомаклональной изменчивости в сочетании с передовыми технологиями позволит преодолеть данные ограничения и удовлетворить растущие потребности в производстве сахара и биотоплива и в биоэнергетической промышленности.

Ключевые слова: каллус; генетика; культивирование *in vitro*; маркеры; регенерация; сомаклональная изменчивость; сахарный тростник

Introduction

The plant tissue culture technique is traditionally used for development of uniform genetic plants through micropropagation, which is the main benefit of clonal cultivars used for commercial cultivation (Duta-Cornescu et al., 2023). However, variation in plants generated from any cell or tissue culture under *in vitro* conditions has been identified and termed as "somaclonal variation" (Larkin, Scowcroft, 1981). This variation provides a potential substitute for crop improvement, especially in asexually propagated crops, such as sugarcane, or crops with narrow genetic bases, self or cross-incompatibility, and irregular inbreeding depression, which are considered significant limitations in using conventional methods for crop improvement. Such genetic improvement assists in reducing the effect of abiotic and biotic stresses in crops like sugarcane via development of resilient cultivars.

Sugarcane is a prominent genetic feedstock used in production of sugar (Azu et al., 2022), biofuel (Marques et al., 2024), and bioenergy (Ahmad, Ming, 2024). Sugar crops rank second in total production among staple crops (Fig. 1*a*). Among sugar crops, sugarcane yields up to 85 % of global sugar production with the rest contributed by sugarbeet (FAO, 2024). Although a large number of different crops are cultivated worldwide, only four account for half of the global production; sugarcane ranks first, being followed by maize, rice, and wheat, as shown in Figure 1*b*. Beyond sugar industry, the significance of sugarcane extends from biomass resources to biofuel production (Ahmad, Ming, 2024).

Conventional breeding by selecting superior parents for mating is hampered in sugarcane by limitations like asynchronous flowering, low seed viability, low fertility, and lack of flowering caused by environmental conditions (Khan M.T. et al., 2018). In fact, sugarcane flowering is specific to certain regions of the world. Continuous efforts have been made to reduce the generational interval in other crops (e. g: doubled haploid, shuttle breeding, etc.), but sugarcane breeding has not benefited from such techniques.

These limitations have led to a narrow genetic base; as the modern genetic feedstock of sugarcane exhibits poor resistance to several diseases (Budeguer et al., 2021). To overcome the common breeding constraints, *in vitro* culture is of keen interest as a way to generate genetic variability for selecting desired genotypes. Variants with a good yield potential developed through tissue culture in sugarcane can be utilized for facing regional environmental challenges and biotic/abiotic stresses. Sugarcane variants developed from *in vitro* culture have been reported since 1971 (Heinz, Mee, 1971).

This paper focuses on how somaclonal variation can complement or replace conventional breeding to generate genetic variations among clones in sugarcane. It covers the mechanisms involved in development of variants along with the factors leading to a satisfactory rate of variation. Despite the successful practical application in developing variants in sugarcane, the progress over the last decade has been neglected. Therefore, this review draws attention towards the significance of somaclonal variation by highlighting its limitations, which may constrain its application on a commercial scale.

In vitro culture techniques

The most promising tissue culture techniques widely adopted in various crops include anther culture, cell suspension culture, callus culture, embryo culture, micropropagation, and proto-

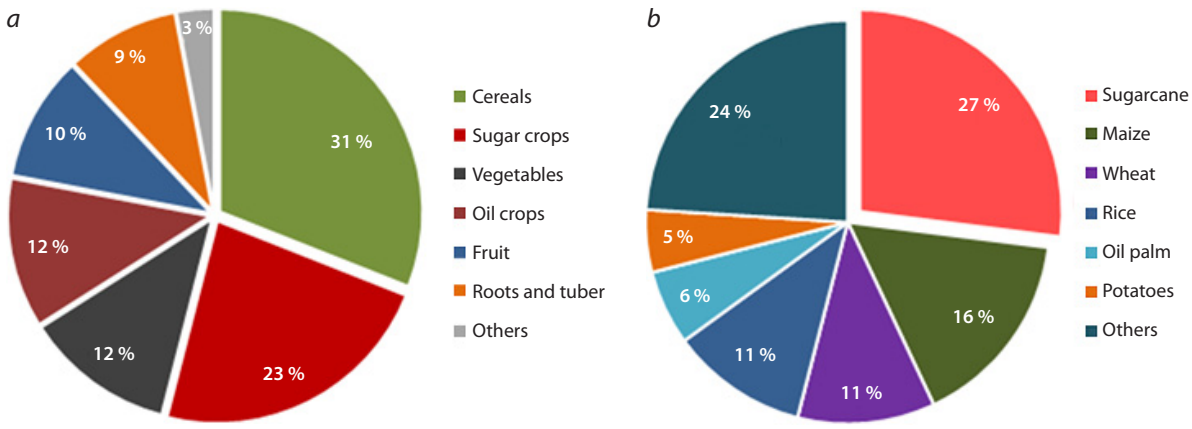


Fig. 1. Contribution of sugarcane to global production in 2023 by (*a*) commodity group and (*b*) individual commodity.
Source: (FAOSTAT, 2024).

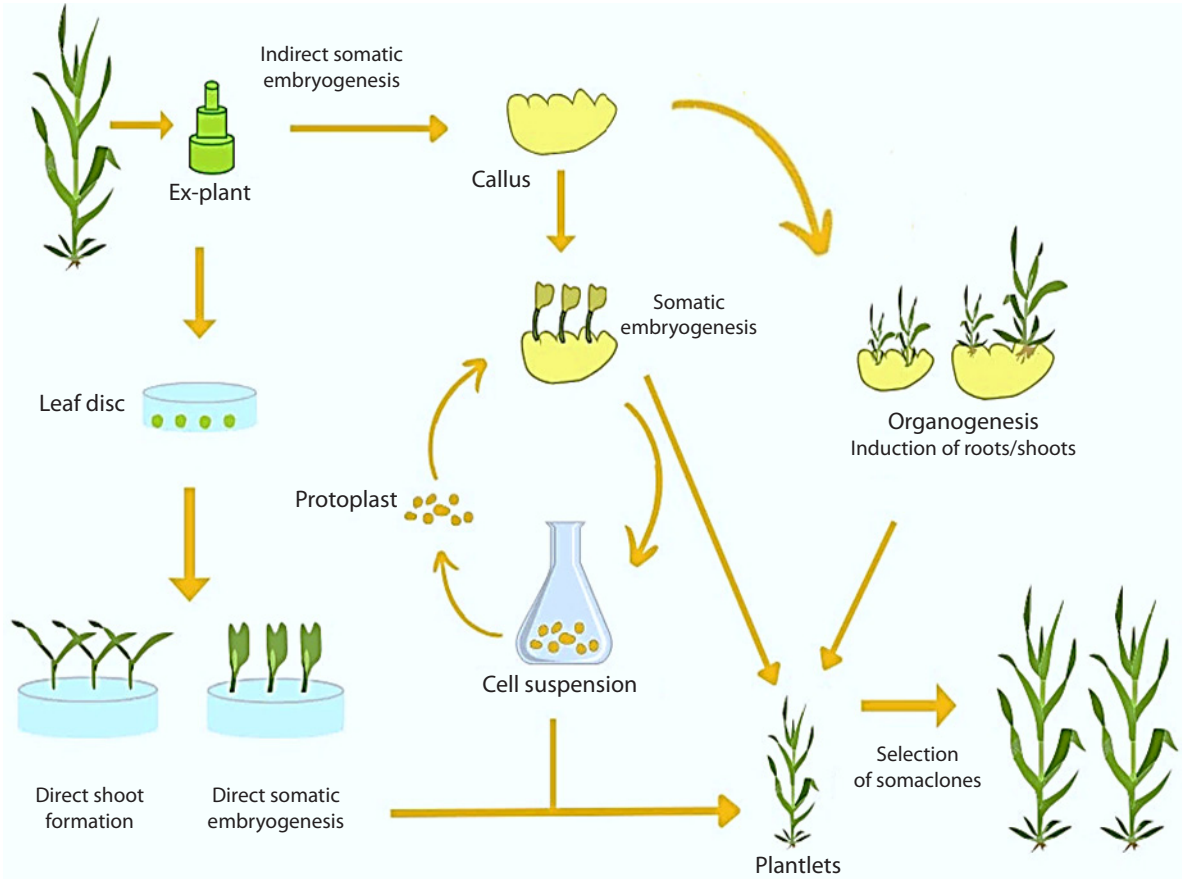


Fig. 2. Regeneration pathways in sugarcane.

plast culture. Different researchers have established different protocols for successful induction of somaclonal variation in sugarcane, but detailed discussion of their protocols is beyond the scope of this paper. However, the principles of extensively used techniques for *in vitro* culture and their regeneration in sugarcane are discussed briefly below.

Organogenesis and somatic embryogenesis are the main pathways for *in vitro* plant regeneration. They require totipotency and competency for plant regeneration under *in vitro* conditions as shown in Figure 2. Newly regenerated plants need to be acclimated to external growth conditions by placing them under low relative humidity and high illumination, which results in cuticle production (Warren, 1991).

In organogenesis, shoots are induced to differentiate from cells or cell clusters. The process involves the progression of leaves, shoots or roots from the explant and their transfer to a diverse medium for root formation (Bidabadi, Jain, 2020). It can be further classified into direct and indirect organogenesis. In the former, the root or shoot develops directly without going into a callus phase. Thorat et al. (2018) has found that sugarcane plants developed from indirect organogenesis exhibit more genetic variation than those developed from direct organogenesis.

In somatic embryogenesis, a cell or group of somatic cells gives rise to a somatic embryo, which morphologically resembles zygote embryos. The somatic embryo has a bipolar structure giving rise to both shoot and root meristems (Khan I.A., Khatri, 2006). Somatic embryogenesis can also follow the direct or indirect (with an intermediate callus phase) pathways.

The first phase in somatic embryogenesis is the induction of embryogenic cells, which requires time to dedifferentiate. Competence and embryo formation depend upon auxins, mainly in the form of 2,4-dichlorophenoxyacetic acid (2,4-D). It represses morphogenesis and disrupts cell-to-cell interaction, leading to the fragmentation of a single cell or group of cells – the mechanism of the proliferation of embryogenic suspension cultures (De Klerk et al., 1997). It is a commonly used pathway for sugarcane micropropagation (Naz et al., 2008; Almeida et al., 2022), with different established protocols. However, these protocols are efficient for not all genotypes due to their differential behavior (Di Pauli et al., 2021).

Nickell and Maretzki (1969) were first to report cell suspension culture in sugarcane. For this purpose, friable callus is dispersed in a liquid medium and then agitated on a rotary shaker to obtain a culture of isolated cells. This process is genotype-dependent, and those with high phenolic compound levels do not perform well. Generally, suspension cultures have been utilized to study metabolic and physiological processes in sugarcane (Goldner et al., 1991; Thorat et al., 2017; Bottcher et al., 2021). However, Thorat et al. (2017) reported the presence of variation in sugarcane somaclones developed through cell suspension culture, although at a lower frequency, as confirmed by molecular markers.

Mechanisms of somaclonal variation

Somaclonal variation is a multifaceted phenomenon comprising genetic or epigenetic mechanisms (Leva, Rinaldi, 2017). The stability of mutations of qualitative features in subsequent generations relies on their molecular basis (Hung et al., 2018; Azizi et al., 2020). If these mutations are unstable or not inherited by the next generations, they may be regarded as epigenetic changes (Hung et al., 2018). Moreover, modifications in tails of histones and DNA methylation are also viewed as a part of epigenetic change (Azizi et al., 2020), which are usually caused by residual effects of plant growth regulators in tissue culture media (Jackson, Lyndon, 1990). Chromosomal variations are usually caused by explant source or culture media conditions (Phillips et al., 1994). Such variations also have been observed in sugarcane somaclones (Sobhakumari, 2012).

Somaclonal variation has been observed in several species and extensively utilized due to its commercial potential, irrespective of the type of mechanisms involved (Leva, Rinaldi, 2017). Economically, somaclonal variation should be valuable and heritable. Genetic changes can be induced by a variety of mechanisms, including changes in chromosome structure and number, chromosomal breakage or aberrations (deletions/inversions), and cryptic changes. Variations in chromosome number, structure, and aberrations are observed in regenerated plants (Mujib et al., 2007). Such changes may lead to loss of genes or their function (Leva et al., 2012). Point mutations responsible for cryptic changes may affect mitochondrial and chloroplast genomes and are typically predicted to take place. Moreover, cryptic changes such as gene silencing or activation of a transposable element play a vital role in somaclonal variation (Barret et al., 2006).

Factors affecting somaclonal variation

The frequency and type of somaclonal variations in sugarcane can be affected by different factors, such as the genotype composition, explant source, culture medium, time period, and number of culture cycles as discussed below.

Genotype

Different plant species or genotypes vary in their susceptibility to somaclonal variation. The most significant factor among various factors may be plant genotype (Shen et al., 2007; Tican et al., 2008; Nwauzoma, Jaja, 2013). Moreover, different genotypes show different responses under stress, pointing out that somaclonal variation is also affected by genotypic composition.

Shen et al. (2007) assessed plant regenerants and showed that plant genotype is a key factor influencing somaclonal variation during *in vitro* culture. The rate of somaclonal variation varies from 2.6 to 40.4 % depending upon the plant genotypes.

Thus, somaclonal variation is a genotype-dependent phenomenon specifically linked to the regeneration potential of different varieties (Manchanda et al., 2018).

Explant source

The origin of the explant exerts an important influence on the rate of somaclonal variation (Ahuja, 1998). Also, it has been reported that the explant source may greatly influence genetic fidelity (Krikorian et al., 1993). Regenerates from explants with pre-existing meristems (such as axillary buds or shoot tips), exhibited less somaclonal variation than those developed from highly differentiated tissue (such as leaves, stems, or roots), which usually produce significant somaclonal variants (Duncan, 1997; Sharma et al., 2007).

However, successful cases of *in vitro* regeneration using leaf pieces, leaf sheath, shoot apical meristem, or pith have been reported in sugarcane (Ali et al., 2007; Abdullah et al., 2013). A comparative study investigated better sources of explant in sugarcane. Leaves, shoot apical meristem, or pith were used as explant sources. Leaves performed better than all other sources, followed by shoot apical meristem and pith during callus induction (Ali et al., 2007). Moreover, Abdullah et al. (2013) reported the leaf as the explant source that performed best for indirect somatic embryogenesis technique in sugarcane as compared to pith. The genetic fidelity of somaclones developed from leaves as explants showed polymorphism up to 81 %, demonstrating that direct regeneration from immature leaf slices could be helpful in exploitation of genetic variation and improvement of the existing genotype of sugarcane (Khan I.A. et al., 2009).

Culture medium composition

From the viewpoint of *in vitro* culture conditions, the composition of culture media is the most influential element for plant regeneration in sugarcane (Abdullah et al., 2013). For the practical application of *in vitro* culture, appropriate concentrations and combinations of cytokinins and auxins are necessary (Letham, Gollnow, 1985). The role of cytokinins in the development of shoot apical meristem of sugarcane has been reported in (Ali et al., 2007).

Plant growth regulators in culture media are potent agents of somaclonal variations (Leva et al., 2012). Suboptimal or supraoptimal concentrations of growth regulators, mainly synthetic compounds, have been associated with variation. Somaclonal variation is also favored by rapid, disorganized growth (Karp, 1994). Adding auxin to the culture media of unorganized callus increases the variation by augmenting the rate of DNA methylation (LoSchiavo et al., 1989). Shahid et al. (2012) reported superiority of 2,4-D as an auxin over IAA during callogenesis in sugarcane. Highly significant differences were observed among somaclonal variants developed in sugarcane using different levels of 2,4-D in media, thus pointing out that culture media greatly influence the occurrence of somaclonal variation. In comparison to explant source or plant genotype, remarkable differences were reported for 2,4-D in sugarcane (Abdullah et al., 2013).

2,4-D and kinetin are two common growth regulators that can lead to changes in chromosome number and induce chromosomal abnormalities (Daub, 1986). Growth regulators increase the cell division rate in the culture during the somatic differentiation stage, leading to endopolyploidy, polyteny, and chromosomal changes (D'Amato et al., 1977; Larkin, Scowcroft, 1981). Culture medium conditions are often responsible for chromosomal variations (Phillips et al., 1994). Such variations have also been observed in sugarcane somaclones (Sobhakumari, 2012). Khan I.A. et al. (2008) also

reported 2,4-D as the best source of somaclonal variation when comparing different auxins (Dicamba, Picloram, and IAA) in sugarcane. The maximum number of chlorophyll mutants was developed with 2,4-D as compared to other auxins, representing the highest rate of somaclonal variation in sugarcane. Plants obtained with 2,4-D auxin showed higher phenotypic variability in sugarcane mainly due to the true change in genetic makeup while showing its enhancing effect on cane yield and plant height.

Culture age and number of subcultures

Somaclonal variation increases with the duration of the culture (Farahani et al., 2011; Sun et al., 2013). As callus age increases, chances of diverse plant production increase during successive subcultures (Zayova et al., 2010). Khan S. et al. (2011) showed that somaclonal variation increased after the eighth subculture.

However, not only the number of subcultures but also the duration of each subculture is significant in creating somaclonal variation particularly in callus culture and cell suspension (Bairu et al., 2006; Sun et al., 2013). Somaclonal variation is more promising in plants regenerated after longer culturing, as rapid multiplication of tissue may influence the genetic stability (Israeli et al., 1995; Etienne, Bertrand, 2003; Petolino et al., 2003). About 30 % of genetic variations marked by morphological changes appeared in sugarcane clones after the fourth subculture (Nogueira et al., 2022).

Detection of somaclonal variation

Detection of somaclonal variation is the foremost step in selecting a cultivar with desirable traits or discarding material with unwanted characteristics. Various biochemical, cytological, morphological, and molecular approaches have been employed to detect the extent and type of somaclonal variation in sugarcane.

Somaclonal variants can be simply assessed by morphological traits such as abnormal pigmentation, plant height, or leaf morphology (Israeli et al., 1991).

The existence of somaclonal variation among *in vitro*-cultured sugarcane plants has been shown by different researchers using morphological markers (Doule, 2006; Dalvi et al., 2012; Khan I.A. et al., 2015; Yasmeen et al., 2017; Abo-Elwafa et al., 2021). In sugarcane variety CoJ 64, morphological variations in traits such as leaf length, stalk diameter, etc. were observed during three years (Sobhakumari, 2012). Likewise, Rajeswari et al. (2009) reported 51 commercially promising sugarcane somaclones out of 700 and evaluated their genetic variability using morphological traits. Morphological markers have been widely used in assessment of genetic variations among sugarcane somaclones (Khan I.A. et al., 2015; Yasmeen et al., 2017; Abo-Elwafa et al., 2021). Nonetheless, the detection of somaclonal variants using morphological tools is environmentally sensitive and time-consuming (Bairu et al., 2011).

Change in chromosome number or structure is assessed by cytological analysis (Sobhakumari, 2012; Abreu et al., 2014). Flow cytometry and complex microscopic techniques are usually used to detect somaclonal variation in tissue-cultured plants. Flow cytometry is widely accepted (Doležel et al., 2004), though it is time-consuming and chemicals used to prepare cytosolic suspensions may interfere with DNA content (Krishna et al., 2016).

In characterization of the complex genome of sugarcane, genetic difference has been verified by flow cytometry (Oliveira et al., 2015; Metcalfe et al., 2019). Thus, Oliveira et al. (2015) detected variation in DNA content in sugarcane varieties and proposed a reliable analysis technique for flow cytometry. Sixteen varieties were classified into four groups relative to their DNA content. Similarly, Nogueira et al. (2022) observed changes in DNA content in regenerated sugarcane plants. Three out of ten genotypes showed change in DNA content. Among these three, two showed reduction of 0.45 pg DNA, while one genotype showed an increase up to 61 % in DNA content during *in vitro* culture. This increase in genome size might be due to polyploidy or transposable elements (Bennetzen et al., 2005), whereas the decrease might be caused by chromosomal breakage or deletions (Petrov, 2001).

Phenotypic variations among organisms are the outcome of biochemical variations that can be due to polymorphism in the genetic makeup of organisms. Isozyme analysis was first performed in sugarcane in 1969 (Heinz, 1969). Enzymes, such as peroxidase, superoxide dismutase, and malate dehydrogenase were used to study variation in sugarcane (Weising et al., 2005). Sugarcane somaclones showed improved activity of antioxidants like catalase, superoxide dismutase, peroxidase, ascorbate peroxidase, and ascorbate under drought stress (Naheed et al., 2020). Variation in sugarcane based on biochemical parameters was also reported by Yasmeen et al. (2017) and Dalvi et al. (2012). Nevertheless, the implementation of isozyme analysis is limited due to some limitations, such as their being tissue-specific, few in numbers, and being affected by environmental conditions.

The complete picture of variations in complex plant genomes can be assessed with the help of molecular markers. They are more reliable tools, as compared to biochemical markers, as they are more numerous, environment-independent, reproducible, and highly specific (Manchanda et al., 2018). Currently, different molecular markers, such as amplified fragment length polymorphism (AFLP), randomly amplified polymorphic DNA (RAPD), simple sequence repeats (SSR), and inter-simple sequence repeats (ISSR), are utilized in different plants to study somaclonal variation (Leva et al., 2012). Somaclonal variation was widely detected in sugarcane by using molecular markers (Thumjamras et al., 2011; Seema et al., 2014; Mahmud et al., 2015; Tawar et al., 2016). In particular, SSR markers are frequently utilized to study variation among somaclones in sugarcane (Nair et al., 2002). Smiullah et al. (2012) rated SSR markers the best tool to study genetic diversity and found up to 51 % polymorphisms using SSR primers in sugarcane somaclones.

Achievements through somaclonal variation in sugarcane

The success of any conventional breeding program depends on genetic variability. However, a program lasts for about 10–15 years and comprises genotype selection and variety testing, passing through different crop development stages. Therefore, to boost the variation, especially in asexually propa-

gated crops, plant tissue culture presents a novel technique, which assists breeders in creating variability in crops in a short time (Karp, 1992; Mathur, 2013). Tissue culture techniques significantly contributed to crop improvement in sugarcane. Useful somaclones have been developed around the globe against different biotic and abiotic stresses (Oloriz et al., 2011; Dalvi et al., 2012; Kumar et al., 2012; Nikam et al., 2015).

The most significant somaclones developed mainly during the last two decades, along with their improved features, are shown in the Table. Ono was one of the pioneer tissue-cultured

Promising clones developed in sugarcane

Plant material	Findings	Resistance if any	Reference
LSC and B36464	Reported efficient callus-mediated regeneration system with a survival rate of 85–91 % of regenerated plants	–	Azu et al., 2022
GT-54 9	Eleven somaclones were assessed for yield and quality parameters	–	Abo-Elwafa et al., 2021
CPF-248	Somaclones (IPSV1 and IPSV2) exhibited improved photosynthesis and antioxidant response	Drought	Naheed et al., 2020
Co 05011	Induction of regeneration of shoot culture at a higher rate and multiplication ratio via pretreatment of explant with thidiazuron	–	Kumari et al., 2017
CoC 671	Somaclones with improved agronomical traits, tolerance to biotic and abiotic stress in sugarcane	Drought and red rot	Tawar et al., 2016
Isd 37, Isd 38, Isd 39, and Isd 40	Disease-tolerant somaclones were screened and further assessed for their genetic variability using RAPD and SSR markers	Red rot	Mahmud et al., 2015
NIA98, BL4, and AEC82-1026	RAPD was utilized to confirm the somaclonal variation in sugarcane clones derived from callus cultures	–	Seema et al., 2014
HSF-242	SSR markers were applied for the identification of somaclonal variation	Sugarcane mosaic virus	Abdullah et al., 2013
CoC 671	Clones with superior yield parameters	Smut	Dalvi et al., 2012
BF-162	Assessment of red rot disease tolerant clones developed from susceptible genotype BF-162 and genetic variation using SSR and RAPD markers	Red rot	Shahid et al., 2012
CoJ 64	Somaclones were evaluated for different morphological parameters	–	Sobhakumari, 2012
S97US297	Determination of somaclonal variation using RAPD and SSR markers	Red rot	Shahid et al., 2011
CP48-103	Salt-tolerant clones were developed via *in vitro* culture	Salinity	Shomeili et al., 2011
K84-200	Application of SSR and RAPD markers to detect genetic variability in somaclones in sugarcane	Salinity	Thumjamras et al., 2011
NIA-98, BL4, and NIA-2004	Use of SSR markers to confirm variations among somaclones	–	Khan I.A. et al., 2009
S. officinarum × *S. spontaneum*, *S. officinarum* × *Erianthus arundinaceus*	Thirty-nine hybrids were used as donors for *in vitro* culture. Fifty-one sub-clones showed good commercial potential	–	Rajeswari et al., 2009
CoJ 88 and CoJ 64	Somaclones regenerated from calli screened for red rot disease showed better tolerance to the disease than parents in the field	Red rot	Sengar et al., 2009
CoC 671	The selection of 147 plantlets was made at different salt concentrations	Salinity	Patade et al., 2008
CoJ 88	Three resistant and four moderately resistant clones of red rot disease were developed	Red rot	Singh et al., 2008
CoC671	Somaclones showed good performance for quality and yield parameters	–	Doule, 2006
CP65-357	Salt tolerant lines were developed via *in vitro* culture	Salinity	Gandonou et al., 2006
COC671	RAPD markers were used to detect genetic variability among somaclones	Salinity, drought	Yadav et al., 2006
CP-43/33	Somaclones with desirable traits were developed	Salinity	Khan S.J. et al., 2004
Q77N1232, Co6519	No inferior morphological parameter was observed in any developed somaclone	Drought	Wagih et al., 2004
CP 70-321, LCP 85-384, and HoCP 85-845	Plantlets regenerated from the callus of leaf roll have lower stalk weight and diameter with a higher stalk population than bud-propagated cane	–	Hoy et al., 2003

sugarcane varieties developed from the susceptible variety Pindar that was found resistant to Fiji disease (Krishnamurthi, Tlaskal, 1974). Two somaclonal variants of sugarcane variety CoC 671 were released as cultivars: Phule Savitri, which has high sucrose content, early maturity, and moderate tolerance to smut and red rot (Jalaja et al., 2006), and VSI 434, which showed an almost double increase in sugar and cane yield (Tawar et al., 2016). After VSI 434, no significant commercially released variety or clones have been produced in the last decade, as shown in the Table.

Challenges encountered in somaclonal variation

Regardless of various somaclones developed in sugarcane, genetic improvement remains limited with few varieties released to date. Some distressing concerns hindering its practical application on commercial scale are listed below:

- Quantitative inheritance in sugarcane is a major limiting factor for the generation of favorable mutations via tissue culture, explicitly regarding yield and sucrose content.
- Somaclonal variation is an expensive technique, as it demands a closed facility.
- The rate of somaclonal variation depends on the genotype of explant used.
- Somaclones exhibit uncontrolled and unpredictable variation.
- Most of the traits acquired by this variation are uncontrolled, unstable, and epigenetic in nature. Therefore it is essential to study the genetic diversity of tissue-cultured plants after their transplantation into field.
- Multilocation trials are demanded to test the stability of desired traits over generations.
- To assess the genetic stability of somaclones, extensive field trials are required.
- Issues such as poor acclimatization, contamination, and tissue dying also impede the practical application of variant development.

To overcome these limitations, somaclones can serve as an initial material for plant breeding rather than as finished cultivars. They act as donors of specific valuable traits, requiring further improvement through conventional breeding methods. Integrating conventional breeding with tissue culture technique, along with other advanced molecular tools such as marker assisted selection, genetic mapping, or gene editing tool (e. g. CRISPR-Cas9) or mutation can enhance sugarcane resilience to abiotic and biotic stresses. Moreover, gene editing approaches enable the modification of targeted genes linked to stress tolerance, ultimately improving the resistance mechanism in sugarcane. Implementing these advance approaches can help breeders overcome the aforesaid concerns.

Conclusions and future prospects

Genetic variability exists both in wild-type and commercial varieties, serving as the foundation for improving cultivars. However, in sugarcane, aspects such as the complex genome, narrow genetic pool, long breeding cycle, and low fertility are the key limiting factors for the development of superior cultivars via recombination. In this context, the tissue culture technique can overcome the hardship of creating genetic variation in sugarcane. Somaclonal variation induced by tissue culture has become the most prevalent approach in sugarcane for developing varieties and superior lines or clones with tolerance to biotic or abiotic stresses, particularly in regions where climatic conditions limit fertilization.

Despite of its potential, somaclonal variation has been underexplored in recent years due to several challenges. These challenges can be met by integrating the tissue culture technique with other biotechnological or latest approaches. For instance, identification of variants at an early stage by means of molecular markers can be productive, as early selection or elimination of variants reduces the time expenditure, especially in long duration crops like sugarcane. Combinatorial use of induced mutagenesis with *in vitro* culture can further enhance variation frequency in crops with large complex genomes and lead to development of sugarcane for commercial use under changing environmental conditions. Moreover, harnessing genetic edition tools like CRISPR-Cas9 can enhance crucial traits related to biofuel and bioenergy production and fulfill the growing future demands in sugar, biofuel, and bioenergy industries.

References

Abdullah, Smiullah, Khan F.A., Iftikhar R., Raza M.M., Aslam R., Hammad G., Ijaz A., Zafar W., Ijaz U. Detection of somaclonal variation in micropropagated plants of sugarcane and SCMV screening through ELISA. *J Agr Sci.* 2013;5:199-208. doi 10.5539/jas.v5n4p199

Abo-Elwafa A., Bakheit B.R., El-Taib A.B., Noby N.Y. Evaluation of some new somaclones of sugarcane for yield and quality. *SVU-IJAS.* 2021;3:129-139. doi 10.21608/svuijas.2021.59073.1073

Abreu I.S., Carvalho C.R., Clarindo W.R. Massal induction of *Carica papaya* L. 'Golden' somatic embryos and somaclone screening by flow cytometry and cytogenetic analysis. *Cytologia.* 2014;79:475-484. doi 10.1508/cytologia.79.475

Ahmad K., Ming R. Harnessing genetic tools for sustainable bioenergy: a review of sugarcane biotechnology in biofuel production. *Agriculture*. 2024;14:1312. doi 10.3390/agriculture14081312

Ahuja M.R. Somaclonal genetics of forest trees. In: Jain S.M., Brar D.S., Ahloowalia B.S. (Eds) Somaclonal Variation and Induced Mutations in Crop Improvement. Dordrecht: Kluwer Academic, 1998; 105-121

Ali A., Naz S., Alam S., Iqbal J. *In vitro* induced mutation for screening of red rot (*Colletotrichum falcatum*) resistance in sugarcane (*Saccharum officinarum*). *Pak J Bot.* 2007;39(6):1979-1994

Almeida F.A., Santa-Catarina C., Silveira V. Somatic embryogenesis in sugarcane (*Saccharum* spp.). In: Ramírez-Mosqueda M.A. (Ed.) Somatic Embryogenesis. Methods in Molecular Biology. New York: Humana, 2022;83-95. doi 10.1007/978-1-0716-2485-2_7

Azizi P., Hanafi M.M., Sahebi M., Harikrishna J.A., Taheri S., Yassoralipour A., Nasehi A. Epigenetic changes and their relationship to somaclonal variation: a need to monitor the micropropagation of plantation crops. *Funct Plant Biol.* 2020;47(6):508-523. doi 10.1071/FP19077

Azu E., Elegba W., Asare A.T., Asare K., Akama C., Asare P., Annor C., Azure S., Danso K. Efficient callus-mediated system for commercial production of sugarcane (*Saccharum spp.*) planting material in Ghana. *Afr J Biotechnol*. 2022;21:208-217. doi 10.5897/AJB2021.17440

Bairu M.W., Fennell C.W., van Staden J. The effect of plant growth regulators on somaclonal variation in Cavendish banana (*Musa* AAA cv. '*Zelig*'). *Sci Hortic.* 2006;108:347-351. doi 10.1016/j.scienta.2006.01.039

Bairu M.W., Aremu A.O., Staden J.V. Somaclonal variation in plants: causes and detection methods. *Plant Growth Regul.* 2011;63:147-173. doi 10.1007/s10725-010-9554-x

Barret P., Brinkman M., Beckert M. A sequence related to rice *Pong* transposable element displays transcriptional activation by *in vitro* culture and reveals somaclonal variations in maize. *Genome.* 2006; 49:1399-1407. doi 10.1139/g06-109

Bennetzen J.L., Ma J., Devos K.M. Mechanisms of recent genome size variation in flowering plants. *Ann Bot.* 2005;95(1):127-132. doi 10.1093/aob/mci008

Bidabadi S.S., Jain S.M. Cellular, molecular, and physiological aspects of *in vitro* plant regeneration. *Plants.* 2020;9(6):702. doi 10.3390/plants9060702

Bottcher A., Domingues-Junior A.P., Souza L.P., Tohge T., Araújo W.L., Fernie A.R., Mazzafera P. Sugarcane cell suspension reveals major metabolic changes under different nitrogen starvation regimes. *Bragantia.* 2021;80:e2921. doi 10.1590/1678-4499.2021-0009

Budeguer F., Enrique R., Perera M.F., Racedo J., Castagnaro A.P., Noguera A.S., Welin B. Genetic transformation of sugarcane, current status and future prospects. *Front Plant Sci.* 2021;12:768609. doi 10.3389/fpls.2021.768609

Dalvi S.G., Vasekar V.C., Yadav A., Tawar P.N., Dixit G.B., Prasad D.T., Deshmukh R.B. Screening of promising sugarcane somaclones for agronomic traits, and smut resistance using PCR amplification of inter transcribed region (ITS) of *Sporisorium scitamineum*. *Sugar Tech.* 2012;14:68-75. doi 10.1007/s12355-011-0132-y

D'Amato F., Buiatti M., Raghavan V., Johri B.M., Bhojwani S.S., Rangaswamy N.S., de Nettancourt D., Devreux M. Cytology, cytogenetics and plant breeding. In: Reinert J., Bajaj Y.P.S. (Eds) Applied and Fundamental Aspects of Plant Cell, Tissue, and Organ Culture. Springer, 1977;343-464. doi 10.1007/978-3-662-02279-5_3

Daub M.E. Tissue culture and the selection of resistance to pathogens. *Annu Rev Phytopathol.* 1986;24:159-186. doi 10.1146/annurev.py.24.090186.001111

De Klerk G.J., Arnholdt-Schmitt B., Lieberei R., Neumann K.H. Regeneration of roots, shoots and embryos: physiological, biochemical and molecular aspects. *Biol Plant.* 1997;39:53-66. doi 10.1023/A:1000304922507

Di Pauli V., Fontana P.D., Lewi D.M., Felipe A., Erazzú L.E. Optimized somatic embryogenesis and plant regeneration in elite Argentinian sugarcane (*Saccharum* spp.) cultivars. *J Genet Eng Biotechnol.* 2021;19(1):171. doi 10.1186/s43141-021-00270-8

Doležel J., Valárik M., Vrána J., Lysák M.A., Hřibová E., Bartoš J., Gasmanová N., Doleželová M., Šafář J., Šimková H. Molecular cytogenetics and cytometry of bananas (*Musa* spp). In: Jain S.M., Swennen R. (Eds) Banana Improvement: Cellular, Molecular Biology, and Induced Mutations. Enfield: Science Publishers Inc, 2004;229-244

Doule R.B. Cane yield and quality characters of some promising somaclonal variants of sugarcane. *Sugar Tech.* 2006;8:191-193. doi 10.1007/BF02943660

Duncan R.R. Tissue culture-induced variation and crop improvement. *Adv Agron.* 1997;58:201-240. doi 10.1016/S0065-2113(08)60256-4

Duta-Cornescu G., Constantin N., Pojoga D.M., Nicuta D., Simon-Gruita A. Somaclonal in micropropagation of variation-advantage or disadvantage the medicinal plants. *Int J Mol Sci.* 2023;24(1):838. doi 10.3390/ijms24010838

Etienne H., Bertrand B. Somaclonal variation in *Coffea arabica*: effects of genotype and embryogenic cell suspension age on frequency and phenotype of variants. *Tree Physiol.* 2003;23:419-426. doi 10.1093/treephys/23.6.419

Farahani F., Yari R., Masoud S. Somaclonal variation in Dezful cultivar of olive (*Olea europaea* subsp. *europaea*). *Gene Conserve.* 2011; 10:216-221

Food and Agriculture Organization. Agricultural production statistics 2010–2023. FAOSTAT database. 2024. https://www.fao.org/statistics/highlights-archive/highlights-detail/agricultural-production-statistics-2010-2023/en

Gandonou C.B., Errabii T., Abrini J., Idaomar M., Senhaji N.S. Selection of callus cultures of sugarcane (*Saccharum* sp.) tolerant to NaCl and their response to salt stress. *Plant Cell Tissue Organ Cult.* 2006;87:9-16. doi 10.1007/s11240-006-9113-3

Goldner W., Thom M., Maretzki A. Sucrose metabolism in sugarcane cell suspension cultures. *Plant Sci.* 1991;73(2):143-147. doi 10.1016/0168-9452(91)90021-Y

Heinz D.J. Isozyme prints for variety identification. *ISSCT Breeders' Newsletter.* 1969;24:8

Heinz D.J., Mee G.W.P. Morphologic, cytogenetic, and enzymatic variation in *Saccharum* species hybrid clones derived from callus tissue. *Am J Bot.* 1971;58:257-262. doi 10.1002/j.1537-2197.1971.tb09971.x

Hoy J.W., Bischoff K.P., Milligan S.B., Gravois K.A. Effect of tissue culture explant source on sugarcane yield components. *Euphytica.* 2003;129:237-240. doi 10.1023/A:1021928823445

Hung Y.-H., Liu F., Zhang X.-Q., Xiao W., Hsieh T.-F. Sexual and non-sexual reproduction: inheritance and stability of epigenetic variations and consequences for breeding application. In: Gallusci P., Bucher E., Mirouze M. (Eds) Plant Epigenetics Coming of Age for Breeding Applications. Advances in Botanical Research. Academic Press, 2018;88:117-163. doi 10.1016/bs.abr.2018.09.002

Israeli Y., Reuveni O., Lahav E. Qualitative aspects of somaclonal variations in banana propagated by *in vitro* techniques. *Sci Hort.* 1991;48(1-2):71-88. doi 10.1016/0304-4238(91)90154-Q

Israeli Y., Lahav E., Reuveni O. *In vitro* culture of bananas. In: Gowen S. (Ed.) Bananas and Plantains. World Crop Series. Springer, 1995;147-178. doi 10.1007/978-94-011-0737-2_6

Jackson J.A., Lyndon R.F. Habituation: cultural curiosity or developmental determinant*? Physiol Planta.* 1990;79:579-583. doi 10.1111/j.1399-3054.1990.tb02120.x

Jalaja N.C., Sreenivasan T.V., Pawar S.M., Bhoi P.G., Garker R.M. Co 94012 – a new sugarcane variety through somaclonal variation. *Sugar Tech.* 2006;8:132-136. doi 10.1007/BF02943647

Karp A. The role of growth regulators in somaclonal variation. *Br Soc Plant Growth Regul Ann Bull.* 1992;2:1-9

Karp A. Origins, causes and uses of variation in plant tissue cultures. In: Vasil I.K., Thorpe T.A. (Eds) Plant Cell and Tissue Culture. Springer, 1994;139-151. doi 10.1007/978-94-017-2681-8_6

Khan I.A., Khatri A. Plant regeneration via organogenesis or somatic embryogenesis in sugarcane: histological studies. *Pak J Bot.* 2006; 38(3):631-636

Khan I.A., Dahot M.U., Seema N., Bibi S., Khatri A. Genetic variability in plantlets derived from callus culture in sugarcane. *Pak J Bot.* 2008;40(2):547-564

Khan I.A., Dahot M.U., Seema N., Yasmeen S., Bibi S., Raza G., Khatri A., Naqvi M.H. Direct regeneration of sugarcane plantlets: a tool to unravel genetic heterogeneity. *Pak J Bot.* 2009;41(2): 797-814

Khan I.A., Seema N., Raza S., Yasmeen S. Comparative performance of sugarcane somaclones and exotic germplasm under agro-climatic conditions of Tando Jam. *Pak J Bot.* 2015;47(3):1161-1166

Khan M.T., Khan I.A., Yasmeen S., Seema N., Nizamani G.S. Field evaluation of diverse sugarcane germplasm in agroclimatic conditions of Tandojam, Sindh. *Pak J Bot.* 2018;50:1441-1450

Khan S.J., Khan M.A., Ahmed H.K., Khan R.D., Zafar Y. Somaclonal variation in sugarcane through tissue culture and subsequent screening for salt tolerance. *Asian J Plant Sci.* 2004;3:330-334. doi 10.3923/ajps.2004.330.334

Khan S., Saeed B., Kauser N. Establishment of genetic fidelity of *in vitro* raised banana plantlets. *Pak J Bot.* 2011;43:233-242

Krikorian A.D., Irizarry H., Cronauer-Mitra S.S., Rivera E. Clonal fidelity and variation in plantain (*Musa* AAB) regenerated from vegetative stem and floral axis tips *in vitro*. *Ann Bot.* 1993;71:519-535. doi 10.1006/anbo.1993.1068

Krishna H., Alizadeh M., Singh D., Singh U., Chauhan N., Eftekhari M., Sadh R.K. Somaclonal variations and their applications in horticul-

tural crops improvement. *3 Biotech.* 2016;6:54. doi 10.1007/s13205-016-0389-7

Krishnamurthi M., Tlaskal J. Fiji disease resistant *Saccharum officinarum* var. Pindar sub-clones from tissue cultures. In: Proceedings XV Congress, International Society of Sugar Cane Technologists. 1974;15(1):130-137

Kumar P., Agarwal A., Tiwari A., Lal M., Jabri M. Possibilities of development of red rot resistance in sugarcane through somaclonal variation. *Sugar Tech.* 2012;14:192-194. doi 10.1007/s12355-012-0138-0

Kumari K., Lal M., Saxena S. Enhanced micropropagation and tiller formation in sugarcane through pretreatment of explants with thidiazuron (TDZ). *3 Biotech.* 2017;7:282. doi 10.1007/s13205-017-0910-7

Larkin P.J., Scowcroft W.R. Somaclonal variation – a novel source of variability from cell culture for plant improvement. *Theor Appl Genet*. 1981;60:197-214. doi 10.1007/BF02342540

Letham D., Gollnow B. Regulators of cell division in plant tissues. XXX. Cytokinin metabolism in relation to radish cotyledon expansion and senescence. *J Plant Growth Regul*. 1985;4:129-145. doi 10.1007/BF02266951

Leva A., Rinald L.M.R. Somaclonal variation. In: Thomas B., Murray B.G., Murphy D.J. (Eds) Encyclopedia of Applied Plant Sciences. Cambridge: Academic Press, 2017;468-473. doi 10.1016/B978-0-12-394807-6.00150-7

Leva A., Petruccelli R., Rinaldi L. Somaclonal variation in tissue culture: a case study with olive. In: Leva A., Rinaldi L.M.R. (Eds) Recent Advances in Plant *in vitro* Culture. London: Intech Open, 2012. doi 10.5772/50367

LoSchiavo F., Pitto L., Giuliano G., Torti G., Nuti-Ronchi V., Marazziti D., Vergara R., Orselli S., Terzi M. DNA methylation of embryogenic carrot cell cultures and its variations as caused by mutation, differentiation, hormones and hypomethylating drugs. *Theor Appl Genet.* 1989;77:325-331. doi 10.1007/BF00305823

Mahmud K., Nasiruddin K.M., Hossain M.A., Hassan L. Screening sugarcane somaclones and their parent varieties against red rot (*Colletotrichum falcatum*) and assessment of variability by RAPD and SSR markers. *SAARC J Agric.* 2015;13(2):173-182. doi 10.3329/sja.v13i2.26578

Manchanda P., Kaur A., Gosal S.S. Somaclonal variation for sugarcane improvement. In: Gosal S.S., Wani S.H. (Eds) Biotechnologies of Crop Improvement. Vol. 1. Springer, 2018;299-326. doi 10.1007/978-3-319-78283-6_9

Marques J.C., Gasi F., Lourenço S.R. Biofuel in the automotive sector: viability of sugarcane ethanol. *Sustainability.* 2024;16:2674. doi 10.3390/su16072674

Mathur S. Conservation of biodiversity through tissue culture. *Res Rev J Microbiol Biotechnol.* 2013;2(3):1-6

Metcalfe C.J., Li J., Giorgi D., Piperidis N., Aitken K.S. Flow cytometric characterisation of the complex polyploid genome of *Saccharum officinarum* and modern sugarcane cultivars. *Sci Rep.* 2019;9:19362. doi 10.1038/s41598-019-55652-3

Mujib A., Banerjee S., Dev Ghosh P. Callus induction, somatic embryogenesis and chromosomal instability in tissue culture raised hippeastrum (*Hippeastrum hybridum* cv. United Nations). *Propag Ornam Plants.* 2007;7(4):169-174

Naheed R., Arfan M., Farhat F., Ijaz S., Khalid H. Acclimatization of drought tolerance with somaclonal variants of sugarcane (*Saccharum officinarum* L.). *Adv Life Sci.* 2020;8(1):57-62

Nair N.V., Selvi A., Sreenivasan T.V., Pushpalatha K.N. Molecular diversity in Indian sugarcane cultivars as revealed by randomly amplified DNA polymorphisms. *Euphytica.* 2002;127:219-225. doi 10.1023/A:1020234428681

Naz S., Ali A., Siddique A. Somatic embryogenesis and plantlet formation in different varieties of sugarcane (*Sacchrum officinarum* L.) HSF-243 and HSF-245. *Sarhad J Agric.* 2008;24(4):593-598

Nickell L.G., Maretzki A. Growth of suspension cultures of sugarcane cells in chemically defined media. *Physiol Plant.* 1969;22:117-125. doi 10.1111/j.1399-3054.1969.tb07847.x

Nikam A.A., Devarumath R.M., Ahuja A., Babu H., Shitole M.G., Suprasanna P. Radiation-induced *in vitro* mutagenesis system for salt tolerance and other agronomic characters in sugarcane (*Saccharum officinarum* L.). *Crop J.* 2015;3:46-56. doi 10.1016/j.cj.2014.09.002

Nogueira G.F., Luis Z.G., Salles L.A., Pasqual M., Scherwinski-Pereira J.E. High-efficiency organogenesis and evaluation of the regenerated plants by flow cytometry of a broad range of *Saccharum* spp. hybrids. *Biologia.* 2022;77:3265-3278. doi 10.1007/s11756-022-01176-7

Nwauzoma A.B., Jaja E.T. A review of somaclonal variation in plantain (*Musa spp*): mechanisms and applications. *J Appl Biosci.* 2013;67:5252-5260. doi 10.4314/jab.v67i0.95046

Oliveira A.C., Pasqual M., Bruzi A.T., Pio L.A., Mendonça P.M., Soares J.D. Flow cytometry reliability analysis and variations in sugarcane DNA content. *Genet Mol Res.* 2015;14(2):7172-7183. doi 10.4238/2015.June.29.11

Oloriz M.I., Gil V., Rojas L., Veitia N., Hofte M., Jimenez E. Selection and characterisation of sugarcane mutants with improved resistance to brown rust obtained by induced mutation. *Crop Pasture Sci.* 2011;62(12):1037-1044

Patade V.Y., Suprasanna P., Bapat V.A. Gamma irradiation of embryogenic callus cultures and *in vitro* selection for salt tolerance in sugarcane (*Saccharum officinarum* L.). *Agric Sci China.* 2008;7(9):1147-1152. doi 10.1016/S1671-2927(08)60158-3

Petolino J.F., Roberts J.L., Jayakumar P. Plant cell culture: a critical tool for agricultural biotechnology. In: Vinci V.A., Parekh S.R. (Eds) Handbook of Industrial Cell Culture. Humana Press, 2003;243-258. doi 10.1007/978-1-59259-346-0_10

Petrov D.A. Evolution of genome size: new approaches to an old problem. *Trends Genet.* 2001;17:23-28. doi 10.1016/s0168-9525(00)02157-0

Phillips R.L., Kappler S.M., Olhof R. Genetic instability of plant tissue cultures: breakdown of normal controls. *Proc Natl Acad Sci USA.* 1994;91(12):5222-5226. doi 10.1073/pnas.91.12.5222

Rajeswari S., Thirugnanakumar S., Anandan A., Krishnamurthi M. Somaclonal variation in sugarcane through tissue culture and evaluation for quantitative and quality traits. *Euphytica.* 2009;168:71-80. doi 10.1007/s10681-009-9889-4

Seema N., Khan I.A., Raza S., Yasmeen S., Bibi S., Nizamani G.S. Assessment of genetic variability in somaclonal variation in sugarcane. *Pak J Bot.* 2014;46:2107-2111

Sengar A., Thind K., Kumar B., Pallavi M., Gosal S.S. *In vitro* selection at cellular level for red rot resistance in sugarcane (*Saccharum* sp.). *Plant Growth Regul.* 2009;58:201-209. doi 10.1007/s10725-009-9368-x

Shahid M.T., Khan F.A., Saeed A., Fareed I. Variability of red rot-resistant somaclones of sugarcane genotype S97US297 assessed by RAPD and SSR. *Genet Mol Res.* 2011;10:1831-1849. doi 10.4238/vol10-3gmr1122

Shahid M.T.H., Khan F.A., Saeed A. Development of somaclones in sugarcane genotype BF-162 and assessment of variability by random amplified polymorphic DNA (RAPD) and simple sequence repeats (SSR) markers in selected red rot resistant somaclones. *Afr J Biotechnol.* 2012;11(15):3502-3513. doi 10.5897/AJB11.2729

Sharma S., Bryan G., Winfield M., Millam S. Stability of potato (*Solanum tuberosum* L.) plants regenerated via somatic embryos, axillary bud proliferated shoots, microtubers and true potato seeds: a comparative phenotypic, cytogenetic and molecular assessment. *Planta.* 2007;226:1449-1458. doi 10.1007/s00425-007-0583-2

Shen X., Chen J., Kane M., Henny R. Assessment of somaclonal variation in *Dieffenbachia* plants regenerated through indirect shoot organogenesis. *Plant Cell Tissue Organ Cult.* 2007;91:21-27. doi 10.1007/s11240-007-9273-9

Shomeili M., Nabipour M., Meskarbashee M., Memari H.R. Evaluation of sugarcane (*Saccharum officinarum* L.) somaclonals tolerance to salinity via *in vitro* and *in vivo*. *HAYATI J Biosci.* 2011;18(2):91-96. doi 10.4308/hjb.18.2.91

Singh G., Sandhu S., Meeta M., Singh K., Gill R., Gosal S.S. *In vitro* induction and characterization of somaclonal variation for red rot and other agronomic traits in sugarcane. *Euphytica.* 2008;160:35-47. doi 10.1007/s10681-007-9531-2

Smiullah, Khan F.A., Abdullah, Afzal A., Javed M.A., Iqbal Z., Iftikhar R., Wattoo J.I. *In vitro* regeneration, detection of somaclonal variation and screening for mosaic virus in sugarcane (*Saccharum* spp.) somaclones. *Afr J Biotechnol.* 2012;11:10841-10850. doi 10.5897/AJB11.4073

Sobhakumari V.P. Assessment of somaclonal variation in sugarcane. *Afr J Biotechnol.* 2012;11:15303-15309. doi 10.5897/AJB12.1627

Sun S., Zhong J., Li S., Wang X. Tissue culture-induced somaclonal variation of decreased pollen viability in torenia (*Torenia fournieri* Lind.). *Bot Stud.* 2013;54:36. doi 10.1186/1999-3110-54-36

Tawar P.N., Sushir K.V., Meti N.T. Somaclonal variation an aid for sugarcane improvement. *Int J Curr Res Biosci Plant Biol.* 2016;3:47-55.

Thorat A.S., Sonone N.A., Choudhari V.V., Devarumath R.M., Babu K.H. Plant regeneration from cell suspension culture in *Saccharum officinarum* L. and ascertaining of genetic fidelity through RAPD and ISSR markers. *3 Biotech*. 2017;7:16. doi 10.1007/s13205-016-0579-3

Thorat A.S., Sonone N.A., Choudhari V.V., Devarumath R.M., Babu K.H. Plant regeneration from direct and indirect organogenesis and assessment of genetic fidelity in *Saccharum officinarum* using DNA-based markers. *Biosci Biotechnol Res Commun.* 2018;11:60-69. doi 10.21786/bbrc/11.1/9

Thumjamras S., Iamtham S., Lersrutaiyotin R., Prammanee S. Identification of sugarcane somaclones derived from callus culture by SSR and RAPD marker analysis. *Thai J Agric Sci.* 2011;44(5):71-76

Tican A., Câmpeanu G., Chiru N., Ivanovici D. Using of unconventional methods for obtaining somaclonal variations, having as goal making of new potato varieties with resistance at diseases and pests. *Rom Biotechnol Lett.* 2008;13(4):3791-3798

Wagih M.E., Ala A., Musa Y. Regeneration and evaluation of sugarcane somaclonal variants for drought tolerance. *Sugar Tech.* 2004;6:35-40. doi 10.1007/BF02942615

Warren G. The regeneration of plants from cultured cells and tissues. In: Stafford A., Warren G. (Eds) Plant Cell and Tissue Culture. Milton Keynes: Open University Press, 1991;82-100

Weising K., Nybom H., Wolff K., Kahl G. DNA Fingerprinting in Plants: Principles, Methods, and Applications. New York: CRC Press, 2005

Yadav P.V., Suprasanna P., Gopalrao K.U., Anant B.V. Molecular profiling using RAPD technique of salt and drought tolerant regenerants of sugarcane. *Sugar Tech.* 2006;8:63-68. doi 10.1007/BF02943744

Yasmeen S., Rajput M.A., Khan I.A., Hasseny S.S. Induced mutations and somaclonal variations in three sugarcane (*Saccharum officinarum* L.) varieties. *Pak J Bot.* 2017;49:955-964

Zayova E., Vassilevska I.R., Kraptchev B., Stoeva D. Somaclonal variations through indirect organogenesis in eggplant (*Solanum melongena* L.). *Biol Divers Conserv.* 2010;3:1-5

Conflict of interest. The authors declare no conflict of interest.
Received May 8, 2025. Revised September 21, 2025. Accepted February 16, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):391-402

doi 10.18699/vjgb-26-43

Factors of wheat frost hardiness – ice recrystallization inhibitor proteins

N.E. Korotaeva [1, 2], A.V. Fedyaeva [2], K.K. Musinov [2, 3], A.S. Surnachev [2, 3], G.B. Borovskii [1]

[1] Siberian Institute of Plant Physiology and Biochemistry of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Siberian Research Institute of Plant Production and Breeding – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Krasnoobsk, Novosibirsk region, Russia

knev73@yandex.ru

Abstract. For winter wheat, winter hardiness is one of the complex traits that determine the successful cultivation of this crop, and the responsible genes are recognized as highly significant for breeding work. The accumulation of proteins that prevent ice recrystallization (ice recrystallization inhibition proteins, IRIP) correlates with the survival of winter wheat, which indicates the importance of taking this trait into account when obtaining more frost-resistant varieties. The importance of IRIPs is determined by their ability to integrate into growing ice crystals, which limits the formation of large ice conglomerates in the tissues of winter plants. Wheat IRIPs, which accumulate mainly in the apoplast of leaves and in the crowns during cold acclimation, are characterized by a typical duality of structural organization that determines both the manifestation of IRI activity and anti-pathogenic properties. The wheat IRIP molecule contains at the C-terminus a conserved NxVx(x)G fragment that repeats several times, forming a β-helix responsible for binding to the ice surface; at the N-terminus, there is an LRR sequence typical of pathogen-activated kinases, as well as a guiding signal peptide. The wheat genome contains up to eleven *IRI* genes. The *TaIRI* gene promoter contains typical basic *cis*-activating elements and some elements that respond to abiotic stress and hormones. Isoforms of proteins responsible for protecting against pathogens (pathogenesis related proteins, PRP), which accumulate in winter wheat during cold acclimation, also have IRI activity. The expression of the IRIP and PRP genes positively correlates with the cold resistance of winter wheat plants. According to modern data, the regulation of the IRIP genes and cold-activated PRP genes is ABA-independent, but depends on the presence of jasmonic acid and on some proteomic transcription factors. The review provides examples of the practical use of isolated winter wheat IRIPs. The issue of the factors regulating the activity of the IRIP genes and cold-activated PRPs is the least developed to date. The association of these proteins with the winter hardiness of wheat indicates the prospects for their further study.

Key words: ice recrystallization inhibition proteins; *Triticum aestivum* L.; winter hardiness; regulation of gene activity

For citation: Korotaeva N.E., Fedyaeva A.V., Musinov K.K., Surnachev A.S., Borovskii G.B. Factors of wheat frost hardiness – ice recrystallization inhibitor proteins. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):391-402. doi 10.18699/vjgb-26-43

Funding. This research was funded by a grant of Russian Science Foundation No. 25-24-00117 (https://rscf.ru/project/25-24-00117/).

Факторы морозоустойчивости пшеницы – белки-ингибиторы рекристаллизации льда

Н.Е. Коротаева [1, 2], А.В. Федяева [2], К.К. Мусинов [2, 3], А.С. Сурначёв [2, 3], Г.Б. Боровский [1]

[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] Сибирский научно-исследовательский институт растениеводства и селекции – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, р. п. Краснообск, Новосибирская область, Россия

knev73@yandex.ru

Аннотация. Для озимой пшеницы зимостойкость – один из комплексных признаков, определяющих успешное возделывание этой культуры, а отвечающие за нее гены признаны высоко значимыми для селекционных работ. Накопление белков, препятствующих рекристаллизации льда (ice recrystallization inhibition proteins, IRIP), коррелирует с выживаемостью озимой пшеницы, что указывает на важность учета этого признака при получении более морозоустойчивых сортов. Значение IRIP для выживаемости определяется их способностью встраиваться в растущие кристаллы льда, что ограничивает формирование крупных ледяных конгломератов в тканях озимых растений. IRIP пшеницы, накапливающиеся преимущественно в апопласте листьев и узлов кущения в период холодовой акклимации, характеризуются типичной двойственной структурной

организацией, определяющей как проявление IRI-активности, так и антипатогенные свойства. Молекула IRIP пшеницы содержит на С-конце консервативный несколько раз повторяющийся фрагмент NxVx(x)G, формирующий ответственную за связывание с поверхностью льда β-спираль; на N-конце расположены типичная для активируемых патогенами киназ LRR-последовательность, а также направляющий сигнальный пептид. Геном пшеницы содержит до 11 *IRI*-генов. Промотор генов *TaIRI* содержит типичные основные *цис*-активирующие элементы и некоторые элементы, реагирующие на абиотический стресс и гормоны. Изоформы ответственных за защиту от патогенов белков (pathogenesis related proteins, PRP), накапливающихся у озимой пшеницы в период холодовой акклимации, также обладают IRI-активностью. Экспрессия генов IRIP и PRP положительно коррелирует с холодоустойчивостью растений озимой пшеницы. Регуляция генов IRIP и генов PRP, активируемых холодом, согласно современным данным, АБК-независимая, но зависит от присутствия жасмоновой кислоты и от некоторых протеомных факторов транскрипции. В обзоре приведены примеры практического использования изолированных IRIP озимой пшеницы. Вопрос о факторах регуляции активности генов IRIP и активируемых холодом PRP является наименее разработанным на сегодняшний день. Связь этих белков с зимостойкостью пшеницы указывает на перспективность их дальнейшего изучения.

Ключевые слова: белки-ингибиторы рекристаллизации льда; *Triticum aestivum* L.; зимостойкость; регуляция активности генов

Introduction

Winter wheat is an economically valuable cereal crop that is of special interest for plant breeders. This crop is cultivated worldwide in regions with significantly different combinations of climatic conditions, which requires a highly diverse selection of high-yielding, regionally adapted varieties. The availability of a large set of studied traits and their associated genes significantly expands the potential for regional adaptation of wheat and for increasing its yield under specific regional conditions.

Active vegetative growth ensuring high yield of winter wheat depends on the overwintering success rate of winter cereal crops. Traits ensuring high survival during the winter-spring period are of special importance. Winter hardiness is a complex trait responsible for stability against several environmental stresses, including soil freezing (Ambroise et al., 2020).

Soil freezing, which leads to ice formation in the crowns of winter wheat, is one of the most significant damaging factors contributing to survival and yield reduction. The accumulation of antifreeze proteins (AFPs) protects the crowns and leaf parts of the plants from freezing and acts as an indicator of increased frost hardiness. The results of studies on AFPs in cereals were presented earlier in the review by J.G. Duman and M.G. Wisniewski (2014). The reviews available to us mostly focus either on cold tolerance or AFPs in plants in general. The information obtained so far on these proteins in wheat remains fragmented.

The goal of this paper is to provide a comprehensive overview of the currently available literature data on AFPs in soft wheat *Triticum aestivum* L. Here, a general characterization of these proteins will be presented, their molecular organization at the proteomic and genomic levels will be examined, and current knowledge on the regulation of their gene activity and practical applications of AFPs in *T. aestivum* will also be presented.

General characterization of antifreeze proteins

AFPs were first described in Antarctic fish more than fifty years ago (DeVries, Wohlschlag, 1969). Since then, the body of knowledge about these proteins has significantly expanded, and methods for their practical application have been proposed and tested (Voets, 2017; Liu et al., 2021). Several terms are used in research papers to refer to these proteins, including ice-structuring proteins (ISP) (Kontogiorgos et al., 2007; Wang X. et al., 2024), ice-binding proteins (IBP) (Janech et al., 2006), thermal hysteresis proteins (THP) (Ewart et al., 1999), or ice recrystallization inhibitor proteins (IRIP) (Knight et al., 1995). Some plant AFPs are glycosylated proteins, commonly referred to as antifreeze glycoprotein proteins (AFGP) (Griffith et al., 1992). The term “IRI proteins” (IRIP) is currently accepted for wheat AFPs with annotated genes, as it largely reflects their functional activity (Wisniewski et al., 2020).

AFPs in plants were first described in bittersweet nightshade (Urrutia et al., 1992) and winter rye (Griffith et al., 1992). It was later found that plant AFPs had a specific distinctive feature compared to those of other living organisms. Plant AFPs are characterized by a weak ability to lower the freezing point of water below its melting point, a property known as thermal hysteresis. Only a few plant species exhibit thermal hysteresis up to +2 °C (Jarząbek et al., 2009; Bayer-Giraldi et al., 2011). In contrast, this property is strongly expressed in animal AFPs (Kristiansen et al., 1999, 2011). In wheat leaves, thermal hysteresis measured in February is about 0.2 °C (Duman, Olsen 1993).

Ice recrystallization occurs during one of the water freezing stages, when latent heat is slowly released as a result of ice formation (Puchkov, 2017). The ability of plant IRI proteins (IRIPs) to inhibit ice recrystallization lies in the binding of a protein molecule to the surface of a growing ice crystal and its incorporation into the crystal, which shapes the crystal as a hexagonal disk (at

low IRIP concentrations) or a bipyramidal prism (at high IRIP concentrations) (Griffith et al., 2005). Importantly, this prevents the formation of large ice conglomerates (Yang et al., 2024), which are likely to damage the living tissues.

The level of IRI activity depends on the concentration of IRIPs in the freezing solution (Regand, Goff, 2006). The ability of plant IRIPs to inhibit ice recrystallization was first described in rye (Griffith et al., 1992) and later in other cold-tolerant species (Antikainen, Griffith, 1997). This property turned out to only be characteristic of frost-hardy plants (Hon et al., 1995; Antikainen, Griffith, 1997).

As an overwintering cold-tolerant species, wheat contains *IRI* genes. The accumulation of IRIPs in wheat is especially relevant in late autumn and early winter, preceded by cold acclimation at low positive temperatures. At temperatures ranging from –6 to –8 °C, cold-acclimated wheat plants exhibit intercellular ice formation in the vascular transition zone of crowns, where large quantities of water migrate from the meristem zone. This water redistribution contributes to the better protection of the meristem zone from ice damage and ensures winter survival (Willick et al., 2019). IRIPs accumulate primarily in the vascular transition zone of crowns compared to the stem meristem zone (Willick et al., 2018).

It is worth noting that the minimum temperature at which meristematic cells survive in the overwintering wheat is close to the LT50 (Lethal Temperature 50 – the temperature at which 50 % of the test plant population dies), whereas IRIP accumulation correlates with the overall winter survival (Chun et al., 1998). In spring, *IRI* genes affect growth, heading, and the onset of flowering in wheat by stimulating the activity of genes responsible for vernalization (Cao et al., 2021).

Snow cover creates favorable conditions for the survival of pathogenic fungi (Gaudet, Laroche, 1997). As a result of tissue damage caused by growing ice conglomerates, there is a high risk of pathogenic infection in winter wheat plants after the loss of snow cover at the end of winter. In response to pathogen attack, plants synthesize PR proteins (PRPs), which are released into the apoplast upon infection to inhibit fungal enzymes or enzymatically degrade their cell walls (Jain, Khurana, 2018).

The LRR amino acid sequence, which is common among PRPs, is also contained in wheat IRIP molecules (Tremblay et al., 2005), where it is responsible for the PR activity (Juurakko et al., 2022). Some wheat PRPs, such as β-1,3-glucanases, chitinases, thaumatin-like proteins, and polygalacturonase-inhibition proteins, exhibit antifreeze activity (Hon et al., 1995; Kontogiorgos et al., 2007). In response to cold, wheat accumulates specific PRP isoforms acting as IRIPs, modifying ice growth during freezing, and ensuring pathogen resistance even before infection (Griffith, Yaish, 2004).

Wheat IRIPs have been found in plant organs such as roots, stems, leaves, and crowns (Duman, Olsen, 1993; Willick et al., 2018, 2019; Jin et al., 2022; Vaitkevičiūte et al., 2024). The initial activation of *TaIRI* genes after cold exposure occurs in the leaves and then propagates to roots and stems (Jin et al., 2022), which is likely due to leaves being above-ground organs, and as such, the first to be affected by freezing. For example, the *TaIRI1* transcript was detected in leaves, crowns, and roots, while the *TaIRI2* transcript was only found in leaves (Tremblay et al., 2005).

A significant accumulation of IRIPs in the extracellular space is observed in winter cereals during cold acclimation (Antikainen, Griffith, 1997). PRPs, the genes of which are regulated by cold acclimation/deacclimation, are likely to act as IRIPs and may be localized in the extracellular space, plasma membrane, nucleus, or cytoplasm (Vaitkevičiūte et al., 2024).

Amino acid sequence of *T. aestivum* IRIPs

It was previously shown that the mature *T. aestivum* IRIPs TaIRI-1 and TaIRI-2 with molecular masses of 26.8 and 40.7 kDa respectively, share 40.4 % sequence identity (Tremblay et al., 2005). A conserved fragment located at the C-terminus and consisting of NxVxxG or NxVxG repeat motifs, where “x” represents a non-conserved residue, is recognized as the main conserved fragment of plant IRIPs, including those of wheat (Tremblay et al., 2005; Middleton et al., 2012; Jin et al., 2018). This conserved amino acid sequence, repeated several times at the C-terminal region, forms a β-helix structure. The two sides of this helix make it possible for IRIPs to bind to the surfaces of two forming ice crystals (Middleton et al., 2009).

Presumably, the number of repeats of this conserved sequence evolved through duplication events during the acquisition of cold tolerance in plants, and an increased number of repeats is likely associated with enhanced resistance to ice formation (Sandve et al., 2008). The number of repeats of the conserved C-terminal motif may vary. For example, TaIRI-1 and TaIRI-2, the two IRIPs of *T. aestivum*, exhibit five and six repeats, respectively (Tremblay et al., 2005). In a later study, the six identified *T. aestivum* IRIPs were found to contain 9 to 13 repeats (Jin et al., 2018).

The N-terminal region of IRIPs contains a leucine-rich segment (Jin et al., 2018). The LRR amino acid sequence, which is common among both eukaryotic and prokaryotic organisms, is known to determine the properties of protein-protein and ligand-binding interactions. In plants, it is incorporated into protein molecules synthesized in response to stress caused by pathogenic

organisms (McHale et al., 2006). The LRR domain of *T. aestivum* proteins TaIRI-1 and TaIRI-2 turned out to be homologous to receptor-like kinases (Tremblay et al., 2005) involved in signaling pathways triggered by the emergence of pathogen-associated elicitor molecules (He, Wu, 2016).

Another region of TaIRI molecules turned out to be homologous to the receptor kinase of the phytosulfokine peptide phytohormone (Tremblay et al., 2005) involved in differentiation of the plant immune response depending on the pathogen type (Sauter, 2015). *T. aestivum* proteins coded for by genes *Tr001_B19* and *Tr001_M19* showed high similarity to cold-regulated IRIP sequences and were annotated as protein kinases with LRR sequences (Monroy et al., 2007). Although IRIPs with kinase domains do not exhibit protein kinase activity (Tremblay et al., 2005), these findings indicate a similarity between IRIPs and PRPs. The N-terminal region of *T. aestivum* IRIPs also contains a signal peptide directing proteins to the extracellular space or to the membrane (Tremblay et al., 2005).

Eight wheat IRIPs were grouped based on the similarity of their amino acid sequences. According to phylogenetic analysis, the amino acids located at the C- or N-terminal regions determine the classification of some wheat IRIPs into group 2 (TaIRI-6, TaIRI-7, and TaIRI-8). Group 1 includes the proteins TaIRI-1, TaIRI-3, TaIRI-4, and TaIRI-5. TaIRI-2 protein having a distinct LRR domain was attributed to the third group (Jin et al., 2018).

T. aestivum IRIP gene organization

The IRI gene family in cold-tolerant grasses of the Pooideae subfamily began to form and acquire specific features approximately 75 Mya, following the divergence of this subfamily into a separate taxon (Sandve et al., 2008). According to the prevailing evolutionary hypothesis with regard to *LRR-IRI*-like genes, the most recent common ancestor of *IRI* genes in Pooideae was the *OsLRR-PSR*-like gene identified in rice. This gene contained at its 5′ end an LRR sequence characteristic of the phytosulfokine receptor kinase. Notably, this sequence is also present in *TaIRI* genes (Houde et al., 2006). It is assumed that the IRI-coding fragment appeared in *TaIRI* genes as a result of an insertional event (Tremblay et al., 2005). The major expansion of the IRI gene group in Pooideae occurred around 36 Mya due to duplications of a specific IRI region within the genes (Sandve et al., 2008). In various studies, *T. aestivum* was found to possess two (Tremblay et al., 2005), six (Jin et al., 2018), or eleven *IRI* genes (Sandve et al., 2008) containing both LRR and IRI coding fragments. The sequence identity between *TaIRI1* and *TaIRI2* is 74.7 and 54.1 % at the 5′ and 3′ ends, respectively (Tremblay, 2005).

With few exceptions, the grouping of *IRI* genes in monocotyledons based on genomic sequence similarity generally corresponds to genus affiliation and phylogenetic relationships between the taxons. It is noteworthy that the organization of *IRI* genes in *Triticum* is closely related to that of *Aegilops* genus plants, which are the progenitors of allohexaploid wheat. Differences in characteristic regions of *TaIRI* (*T. aestivum*) gene sequences were found when compared against *TdiIRI1* (*T. dicoccoides*), *TmIRI1* (*T. monococcum*) and *TuIRI1* (*T. urartu*), but not against *TdIRI1* (*T. durum*) (Jin et al., 2018).

Eight *TaIRI* gene sequences are divided into four groups, with one gene each in groups III and IV (*TaIRI6* and *TaIRI2*, respectively), four genes in group I (*TaIRI1*, *TaIRI3*, *TaIRI4*, and *TaIRI5*), and two genes in group II (*TaIRI7* and *TaIRI8*) (Sandve et al., 2008; Jin et al., 2018).

The promoters of *TaIRI* genes contain typical core *cis*-acting elements as well as elements responsive to abiotic stress and hormones. These responsive elements include three elements in charge of responses to low temperature and drought stresses, one element specifically in charge of stress response at low temperatures, two elements involved in methyl jasmonate signaling response, four light-responsive elements, one abscisic acid-responsive element (ABRE), and several additional regulatory elements (ARE, CAT-box, O2-site and TC-rich repeats) (Jin et al., 2022).

Mapping of *T. aestivum* IRIP and cold-activated PRP genes

To date, a substantial amount of information has been obtained regarding the chromosomal localization of traits associated with wheat frost hardiness. Relevant traits for frost hardiness are located in QTL (Quantitative Trait Loci) regions on chromosomes 1D, 2A, 2B, 6D, and 7B (Båga et al., 2007); 5B and 5D (Chun et al., 1998); 5A, 2D, 2A, and 4B (Case et al., 2014); 5A and 4B (Kruse et al., 2017); as well as in chromosomes from group 5, 2B, and 4B (Sutka, 1994, 2001). The *Fr1* and *Fr2* genes controlling frost hardiness have been mapped on chromosomes 5A and 5D, respectively (Sutka et al., 1997). Substitutions of chromosomes from the Cheyenne variety for chromosomes 4A, 6A, 3B, 5D, and 3D of the frost-sensitive Chinese Spring variety resulted in increased protein accumulation in the apoplast, while the substitutions of chromosomes 1A, 5A, 4B, 5B, 6B, 1D, and 5D resulted in increased antifreeze activity.

The key regulatory genes responsible for increasing both the antifreeze activity and protein accumulation in the leaf apoplast were observed on chromosomes 5B and 5D (Chun et al., 1998). Two *TaIRI* genes were mapped on the same chromosomes obtained from the leaves of the cold-acclimated Zhoumai variety of *T. aestivum* (Zheng

et al., 2020). The *Tr015_M14* and *Tr017_M15* genes coding for IRIPs in the Chinese Spring variety, were located in the 4A locus (Monroy et al., 2007). Notably, the analysis was conducted using plant parts that included both the crown and leaves, despite the fact that the highest IRIP content in wheat plants exposed to freezing is found in the vascular transition zone of crowns, while the meristematic zone and leaves exhibit their own specific responses to freezing (Willick et al., 2018).

Since IRIPs contain a fragment related to PRPs, it is reasonable to search for loci that simultaneously control responses to cold and pathogen attack. QTL regions in *T. aestivum* associated with resistance to both stress factors have been identified on chromosome 5A and are closely linked to the *Fr2* locus. Separate QTLs associated with freezing hardiness were found on chromosomes 5A and 4B, while a QTL associated with snow mold resistance was identified on chromosome 6B (Kruse et al., 2017). The accumulation of β-1,3-glucanases, chitinases, and thaumatin-like proteins under freezing conditions was reduced in plants with chromosome substitutions 2A, 3A, 6B, and 7A. Substitution of chromosome 7A led to a significant prolongation of the freezing period followed by PRP gene activation (Chun et al., 1998).

It can be seen that the composition of QTL regions responsible for frost hardiness and antifreeze activity in wheat varies among varieties. In addition, the specific plant part used for mapping plays a significant role. Therefore, when genome organization of wheat plants is analyzed in studies of proteins with IRI activity, it is essential to take these factors into account.

Effect of low temperatures on the activity of *T. aestivum* IRIP genes

To date, substantial evidence has been accumulated showing that the expression of *T. aestivum* AFPs is induced by cold exposure (see the Table) and is associated with cold and frost hardiness of the variety.

The expression of three *TaIRI* genes (*TaIRI16*, *TaIRI17*, and *TaIRI18*) in stem, leaf, and root tissues of young *T. aestivum* plants was induced by cold (+4 °C, from 2 to 72 hours), and *TaIRI16* showed the most significant induction among them (Jin et al., 2018, 2022). The *TaIRI1* transcript in soft wheat begins to accumulate after low temperature exposure (+4 °C) and reaches its maximum level after 36 days of acclimation. Further exposure to low temperature does not lead to a higher transcript level. The *TaIRI2* transcript also accumulates immediately after the onset of cold hardening, but it reaches its peak earlier in the acclimation period. After deacclimation, transcript levels return to those of non-acclimated plants (Tremblay et al., 2005).

Cold hardening (1–6 days) of seedlings from the Jagger and Alabaskaya varieties showed increased transcript levels of the *Tr001_B19* (GB ID CK197231) and *Tr001_M19* (GB ID CK201227) genes exhibiting high sequence similarity with cold-regulated IRIPs (Monroy et al., 2007). The activity of *TaIRI* genes related to genes coding for *Lolium perenne* IRIPs more than doubled after moving cold-hardened *T. aestivum* plants to subzero temperatures (Herman et al., 2006). The activity of the AFP I precursor gene (represent sequence ID CK197682) was more than eight times as high within one day of freezing young wheat plants at –5 °C and reached even higher levels by the third day of freezing (Kang et al., 2013).

Levels of *TaIRI-1* and *TaIRI-2* transcripts increase significantly in response to both gradual and abrupt temperature drops. Their induction occurs later than that of PRP-encoding genes. The highest accumulation levels are observed in leaves, especially when it comes to the *TaIRI-2* transcript (Winfield et al., 2010).

Much of the evidence for the role of IRIPs in cold and frost hardiness has been obtained using wheats varying in their cold tolerance. Enhanced expression of *TaIRI-1* and *TaIRI-2* in response to low temperatures was observed in winter wheat cultivars compared to spring ones (Winfield et al., 2010). The more cold-tolerant cultivar Mironovskaya 808 has demonstrated higher *TaIRI* gene activity under all tested cooling regimes (Jin et al., 2022). After the first six hours of cold acclimation, transcript levels of *Tr015_M14* and *Tr017_M15* coding for AFPs were lower in the winter cultivar CDC Clair (LT50 –17 °C) than in the spring cultivar Quantum (LT50 –8 °C). However, the gene activity in the winter cultivar exceeded that in the spring cultivar after the first day of hardening, with almost three times the difference in *Tr017_M15*. Expression of the *Tr015_J17* gene, annotated as IRIP-encoding, was two times as high in winter wheat seedlings after one day of acclimation at +4 °C compared to the spring cultivar, and this difference increased with longer acclimation up to day 14. Transcript levels of *Tr001_B19* in the winter cultivar CDC Clair exceeded those in the spring cultivar Quantum at all stages of acclimation (from 6 hours to 14 days) (Monroy et al., 2007).

Additional evidence for the role of IRIPs in *T. aestivum* cold tolerance was obtained through gene transfer experiments. Expression of the *TaIRI4* and *TaIRI6* genes from the frost-resistant cultivar Mironovskaya 808 in transgenic tobacco under freezing conditions led to reduced membrane permeability (with the transformation of *TaIRI4*) and decreased lipid peroxidation activity (with transformations in both *TaIRI4* and *TaIRI6*) (Jin et al., 2018).

These studies provide convincing evidence that the expression of *T. aestivum* IRIP genes is activated by both subzero and low positive temperatures and contributes to the development of frost hardiness in this species.

IRIP- and PR-encoding genes, the activity of which increases in response to cold exposure

Gene name	Treatment conditions	Stage of development or plant organ	Reference
	IRIP		
TaIRI4, *TaIRI16*, *TaIRI17*, *TaIRI1*	+4 °C, 2–72 h	leaves, stems, roots	Jin et al., 2018, 2022
TaIRI-1	+4 °C, 36 d	leaves, roots crowns	Tremblay et al., 2005
TaIRI-2	+4 °C, 6 d	leaves	
Tr001_B19 (GB ID K197231); *Tr001_M19* (GB ID K201227); *Tr015_M14* (GB ID Y742123); *Tr017_M15* (GB ID Y742648)	+4 °C, 1–14 d	seedlings	Monroy et al., 2007
Tr015_J17	+4 °C, 1–6 d	seedlings	
Contig3670_at; *Contig3668_at*; *Contig7221_s_at*	+3 °C, 3 weeks → –3 °C, 6 h–3 d	–	Herman et al., 2006
Ta.12663.1.S1_at (CK197682)	–5 °C, 1–3 d	leaves	Kang et al., 2013
TaIRI-1; *TaIRI-2*	"shock" (temperature decline from +16 to +4 °C during 2 d); "acclimation" (temperature decline from +16 to +4°C during 5–9 weeks)	leaves	Winfield et al., 2010
TaIRI-1	+3 °C, 3 weeks	leaves	Livingston et al., 2021
	PRP		
Pathogenesis-related protein 1-21 (TraesCS7B03G0275300)	+2 °C, 7 weeks → +10 °C, 24 h–1 weeks → + 2 °C, 24 h–2 weeks	crowns	Vaitkevičiūte et al., 2024
	PRP – chitinases		
*CL386Contig5** *CL386Contig1** *CL1911Contig1** *CL40Contig11** *CL40Contig9** *CL754Contig1**	+4 °C, 14 d	–	Houde et al., 2006
JC5845	+3 °C, 3 weeks → –3 °C, 6 h–3 d	–	Herman et al., 2006
Chitinase (ID AB029936.1)	–5 °C, 1–3 d	leaves	Kang et al., 2013
Chitinase (class II)	+2 °C, 1–28 d	crowns	Gaudet et al., 2000
TaCHT-1	+3 °C, 3 weeks	leaves	Livingston et al., 2021
Chitinase EC 3.2.1.14 (TraesCS1B03G0732400) (TraesCS3A03G0663800) (TraesCS4D03G0756500)	+2 °C, 7 weeks → +10 °C, 24 h–1 week → + 2 °C, 24 h–2 weeks	crowns leaves	Vaitkevičiūte et al., 2024
	PRP – thaumatin-like proteins		
TraesCS5B03G1182300 TraesCS2A03G0227600 TraesCS4A03G0142000 TraesCS4A03G0750400 TraesCS4A03G1252200 TraesCS4D03G0549700 TraesCS5A03G0810000 TraesCS5A03G0043700 TraesCS7B03G1303000 TraesCS7B03G1122500 TraesCS5B03G0034800 TraesCS7A03G1360800 TraesCS6B03G1288000	+2 °C, 7 weeks → +10 °C, 24 h–1 week → + 2 °C, 24 h–2 weeks	crowns leaves	Vaitkevičiūte et al., 2024
	PRP – glucanases		
TaGLB2b	+16 → +4 °C during 5–9 weeks	leaves	Winfield et al., 2010
β-1,3-glucanase	+2 °C, 1–28 d	crowns	Gaudet et al., 2000

* The Arabidopsis TAIR database was used to find homologues (Hannah et al., 2005).
GB ID – GenBank Identification Description.

Regulation of *T. aestivum* PRP gene activity by low temperatures

A number of PRP-encoding wheat genes are activated by low temperatures (see the Table). It is assumed that such genes code for PRP isoforms with IRI activity. M. Houde et al. (2006) obtained high-quality ESTs (Expressed Sequence Tags) from wheat cDNA libraries associated with cold stress. A set of 75,488 unique sequences (31,580 contigs and 43,908 singletons/singlets) enriched with stress-regulated genes was obtained. Among them were ESTs containing contigs coding for chitinases, β-1,3-glucanase, and thaumatin-like proteins. Fifteen of these sequences were annotated as precursors or actual chitinases, six of which being cold-inducible. The transcript level of the *TaGLB2b* gene coding for glucanase significantly increases only in response to gradual temperature decrease, while no increase is observed under sudden cold shock. Its induction occurs earlier than that of *TaIRI* genes, i. e. between the third and the fifth weeks of cooling (Winfield et al., 2010).

The activity of the *JC5845* gene coding for chitinase more than doubles after moving cold-hardened *T. aestivum* plants to subzero temperatures (Herman et al., 2006). The gene coding for chitinase 3 (represent sequence ID AB029936.1) was among those, the activity of which was more than eight times as high in response to freezing young wheat plants at –5 °C. Its expression level rose significantly depending on the duration of exposure, with nearly a 20 times difference between the measurements taken on days 1 and 3 of freezing (Kang et al., 2013).

Differences in PRP gene expression levels in response to cooling were observed between wheats varying in PR and frost hardiness. When comparing the genomes of spring and winter wheat cultivars under cold acclimation (+4 °C, from 6 hours to 14 days), no significant differences were found in the activity of genes coding for thaumatin or chitinase. However, the increase in transcript levels of the *Tr017_C17* gene coding for chitinase II precursor in response to hardening was higher in the winter cultivar (Monroy et al., 2007).

A number of chitinase and β-1,3-glucanase genes are regulated by autumn cold acclimation: their expression begins in late autumn, peaks in mid-winter, then decreases, and reaches maximum levels again in spring. Under cold acclimation, β-1,3-glucanase gene transcripts are weakly expressed during short-term exposure, but the expression increases significantly with prolonged exposure and remains elevated after returning to control conditions. The chitinase-encoding gene begins active expression immediately after the onset of cooling, and its expression decreases after returning to control conditions. Differences in expression levels of these genes between cultivars resistant and less resistant to snow mold began to manifest themselves during autumn cold acclimation, with higher expression observed in the more resistant cultivar (Gaudet et al., 2000).

D.P. Livingston et al. (2021) found that cold acclimation of young wheat plants enhances the expression of genes coding for chitinase-1 (*TaCHT-1*) and IRIPs. However, the relationship between the freezing temperature of leaves from pre-acclimated plants and the expression levels of these genes turned out to be negative, i. e. the expression was lower in leaves that froze at lower temperatures.

Since plant AFPs lower the freezing point by only a fraction of a degree (Tremblay et al., 2005), the absence of a direct positive correlation between *TaIRI-1* and *TaCHT-1* expression and leaf freezing temperature is understandable. Expression levels of *TaIRI-1* and *TaCHT-1* show positive correlation with the abundance and complexity of bacterial and fungal communities, likely due to the structural and functional similarities between AFPs and PRPs.

These findings indicate the existence of differential expression of PRP genes regulating their activity depending on the wheat cultivar's cold tolerance, the duration and direction of exposure, and organ-specific features.

G. Vaitkevičiūte et al. (2024) studied the effects of cold deacclimation and reacclimation on genome activity in frost-hardy and frost-sensitive *T. aestivum* cultivars. It was shown that a significant number of genes associated with pathogen response reacted to both deacclimation and repeated cold exposure, with gene activity affected by the duration of exposure. For example, both short- and long-term deacclimation of crowns led to reduced activity of the thaumatin-like protein gene, while longer exposure additionally reduced the activity of chitinase-encoding genes.

During reacclimation, differential expression of genes coding for PRPs with similar action was observed. Among the genes coding for chitinase in crowns, some exhibited decreased activity during reacclimation, while others showed increased activity. Reacclimation in crowns also resulted in differential activity in thaumatin-like protein genes depending not only on the direction of change but also on the duration of exposure.

A significantly more diverse set of genes coding for thaumatin-like protein and chitinase in *T. aestivum* leaves compared to crowns responded with differential expression to deacclimation and reacclimation depending on the duration of exposure. The activity of PR genes in the crowns of the resistant cultivar Lakaja DS was less affected by temperature fluctuations compared to the more sensitive cultivar KWS Ferrum. In contrast, a greater number of genes were regulated by temperature fluctuations in the leaves of the more resistant cultivar compared to the more sensitive one, many of them being thaumatin-like protein genes (Vaitkevičiūte et al.,

2024). These findings reveal a link between wheat cold tolerance and the transcription of PRP genes presumably also acting as IRIPs.

Given the above, we can conclude that wheat genes coding for PRPs are activated by cold and reacclimation and contribute to enhanced cold and frost hardiness. At the same time, genes belonging to the same PRP subgroup (chitinases, thaumatin-like proteins) are regulated differentially, and not all of them are activated by cooling.

Regulatory pathways of *T. aestivum* IRIP and cold-induced PRP gene expression

It has been shown that wheat *IRI* genes begin to express in response to decreasing temperatures. The activity of plant genes responsible for cold tolerance is regulated through both ABA-dependent and ABA-independent pathways. Among the latter, the ICE–CBF–COR signaling pathway is considered the primary one. This includes cold-regulated genes (COR – COld Related), which activated by C-repeat Binding Factor (CBF) with the activity of CBF controlled by a dedicated inducer (ICE – Inducer of CBF Expression) (Chinnusamy et al., 2003). The role of this pathway in activating cold adaptation genes in *T. aestivum* is widely recognized. However, the regulation of *IRI* genes by the ICE–CBF–COR transcription factor group has not been definitively proven, despite the fact that ICE proteins in *T. aestivum* bind to MYC elements (Badawi et al., 2008) present in *TaIRI* promoters (Jin et al., 2022). The literature search revealed one research suggesting a possible link between the ICE–CBF–COR pathway and *IRI* regulation in *T. aestivum* (Tchagang et al., 2017).

Another known regulatory pathway for cold adaptation is ABA-dependent (Xue-Xuan et al., 2010). Although ABA plays an important role in the development of cold and frost hardiness in *T. aestivum* (Zheng et al., 2020), some factors responsible for these traits are activated independently of ABA (Wang W. et al., 2017). The ABA involvement in the activation of *TaIRI* and PRP genes has not been confirmed, although *TaIRI* gene promoters contain elements associated with ABA-responsiveness (Jin et al., 2022). No activation of the *TaIRI1* and *TaIRI2* genes was observed after spraying 6-day-old *T. aestivum* seedlings with ABA (Tremblay et al., 2005).

In *T. monococcum*, ABA levels in leaves and crowns increase during the initial phase of non-lethal cold exposure, but the adaptation begins on the third day of exposure. Notably, the decrease in ABA levels coincides with an increase in salicylic acid content. Thus, it is possible that ABA's influence on *TaIRI* and cold-induced PRP gene expression during cold exposure is limited based on exposure duration, and ABA may not be involved in the adaptive response during later phases requiring IRIP accumulation, when the regulatory role in adaptive changes may shift to other factors (Vanková et al., 2014).

Wheat IRIPs are structurally similar to PRPs, and its PRPs (β-1,3-glucanases, thaumatin-like proteins, and chitinases) also exhibit AFP properties. It is reasonable to assume that PR-associated factors, particularly the WRKY transcription factor family, may act as gene transcription factors for plant AFPs (Pandey, Somssich, 2009). Several findings indirectly support this possibility. For example, WRKY gene expression is associated with increased wheat resistance during the early stages of cold acclimation (Talanova et al., 2008; Winfield et al., 2010); some WRKY genes are activated by cold only in winter wheat (Winfield et al., 2010); and WRKY gene activity increases at subzero temperatures and depends on exposure duration (Herman et al., 2006). In addition, some wheat genes coding for glucanases, chitinases, and thaumatin-like proteins are co-regulated with genes coding for WRKY transcription factors (Winfield et al., 2010). The *TaWRKY1* gene is activated by both pathogen attack and cold (Wang W. et al., 2017), and some WRKY proteins exhibit IRIP-like properties on their own (Huang, Duman, 2002).

It was shown that phytohormones involved in pathogen response, such as ethylene, jasmonic acid, and salicylic acid, also participate in the AFP gene regulation. These phytohormones are also involved in regulating IRI gene activity in wheat (Tremblay et al., 2005), although their role in cold and frost hardiness is not always definitively confirmed (Kurepin et al., 2013). In addition, it has been shown that ethylene is not associated with wheat cold tolerance (Macháčcková et al., 1989).

Expression of the *TaIRI17*, *TaIRI18*, and especially *TaIRI16* genes in stem, root, and leaf tissues of young *T. aestivum* plants increased after exogenous treatment of wheat shoots, roots, and leaves with methyl jasmonate combined with cold exposure (+4 °C, 24 h) (Jin et al., 2022). Exogenous methyl jasmonate treatment also stimulated the activity of *TaIRI* gene promoters (Jin et al., 2022). However, jasmonic acid had different effects on *TaIRI1* and *TaIRI2* expression in 6-day-old seedlings of a frost-hardy *T. aestivum* cultivar, it induced *TaIRI1* expression but not *TaIRI2* (Tremblay et al., 2005). Analysis of motif activity of *TaIRI* gene promoters in tobacco and their expression in *A. thaliana* after the respective transformations showed that MYB, Mybb, MYC, CGTCA, and TGACG motifs from the *TaIRIp1* promoter were highly active and inducible by cold and methyl jasmonate (Jin et al., 2022).

The role of salicylic acid in AFP gene regulation still remains to be confirmed. No activation of *TaIRI1* and *TaIRI2* was observed after spraying 6-day-old seedlings of the frost-hardy *T. aestivum* cultivar with salicylic acid (Tremblay et al., 2005). Treatment of young wheat plants of cultivar Yangmai 16 with salicylic acid solution followed by 6 days of field freezing (at the average temperatures of +5.3 °C and minimum of –7 °C) did not

increase *IRI2* (AY968589) expression compared to control temperature conditions (average +12 °C) (Wang W. et al., 2021). At the same time, freezing without salicylic acid solution treatment led to a twofold increase in gene expression compared to control. According to the authors, previous experiments showed that pretreatment with salicylic acid could increase *TaIRI2* expression during 48-hour cold exposure (Wang W. et al., 2020). Thus, it may be suggested that salicylic acid regulation of *TaIRI2* expression may depend on exposure duration, though this requires further study.

Ethylene is involved in regulating antifreeze activity in winter rye in response to cold and drought, with β-1,3-glucanases, chitinases, and thaumatin-like proteins accumulating in the leaf apoplast (Yu et al., 2001). In *T. aestivum*, spraying 6-day-old seedlings of a frost-hardy cultivar with ethylene solution led to differential regulation of the *TaIRI1* and *TaIRI2* genes: specifically, it induced *TaIRI1* expression but not that of *TaIRI2* (Tremblay et al., 2005). However, ethylene turned out to have a negative association with wheat cold tolerance (Macháčcková et al., 1989).

Overall, data on transcription factor involvement in AFP gene expression regulation remain limited, and information on phytohormonal regulation is often contradictory. It appears that the duration, intensity, and phase of cold exposure play a significant role in phytohormonal regulation. It also remains unclear why the studied phytohormones do not enhance wheat cold tolerance despite their ability to activate *TaIRI* genes. This aspect of antifreeze activity regulation requires further investigation.

Practical application of *T. aestivum* IRIPs and cold-induced PRPs

The use of plant AFPs is limited by their reduced thermal hysteresis capacity compared to animal AFPs and is primarily justified by their ice recrystallization inhibition (IRI) activity. One of the most promising directions for the practical application of AFPs is the development of more frost-hardy plants and cultivars. Although gene transfer studies involving wheat AFPs have yielded encouraging results (Jin et al., 2022), practical use of the available knowledge on wheat IRIPs remains limited.

In present, the main area of practical application for plant AFPs is cryopreservation in medicine and food industry. Naturally derived AFPs are less toxic than dimethyl sulfoxide (DMSO), which is commonly used today, making them promising cryoprotectants (Chow-Shi-Yée et al., 2020). A crude extract from wheat seedlings turned out to protect rat hepatocytes during cryopreservation; the effect could be successfully enhanced by further purification of the extract or by using recombinant TaIRI-2 protein, which enabled longer storage of hepatocytes and recovery of a greater number of viable cells restoring their functions upon thawing (Grondin et al., 2009).

When recombinant TaIRI-2 was used for cryopreserving insulin-secreting INS832/13 cells, it was found that TaIRI-2 could penetrate these cells as well as hepatocytes. This likely contributed to improved cell survival after thawing and reduced cell death (Chow-Shi-Yée et al., 2016). However, it is important to consider that plant-based cryoprotectants may have immunogenic properties, potentially triggering unwanted immune or allergic reactions, which is a significant limitation for medical cryopreservation uses (Kostyaev et al., 2016).

In the food industry, natural AFPs are widely used in ice cream production, with several patents granted, including one for the use of AFPs from winter rye (Boonsupthip, Lee, 2003). AFPs extracted from cold-acclimated wheat significantly improved ice cream texture, and this effect was not diminished by shock freezing (Regand, Goff, 2006). Adding AFPs extracted from flour of frost-hardy winter wheats to bread dough reduces water flow and migration during freeze-thaw cycles, helping to preserve the dough's internal structure and gluten network, increase viscosity and elasticity, and improve textural properties (Wang X. et al., 2024).

Direct use of raw plant extracts in the food industry is preferable to using purified extracts due to their low yield and the complexity of the purification process. However, using concentrated extracts can be problematic due to the presence of enzymes that may negatively affect the taste or texture of the final product. Undesired enzymatic activity in the extract can be eliminated by heat treatment, so thermal stability becomes necessary for natural AFPs used in food production. A promising thermostable thaumatin-like AFP was identified in the leaf apoplast extract of cold-acclimated *T. aestivum* (Kontogiorgos et al., 2007).

Conclusion

The relationship between the expression of AFP genes (with the subsequent accumulation of their respective products) and increased frost hardiness in winter wheat has been convincingly demonstrated in a number of model experiments. It has been shown that the accumulation of apoplastic proteins and AFPs correlates with the survival of winter wheat plants under field conditions (Chun et al., 1998). However, it remains unclear how significant the contribution of antifreeze activity is to plant survival compared to other known factors (e. g., accumulation of free proline and oligosaccharides, depth of occurrence of the crown). Therefore, including AFP accumulation as a target trait in breeding programs still requires more solid justification. Wheat AFPs exhibit dual antifreeze and antipathogenic activity. On the one

hand, this duality complicates the assessment of the specific contribution of AFPs to wheat survival; on the other hand, it makes the presence of this trait particularly valuable, as it reflects the presence of two types of resistance and yield-enhancing factors.

In the vast majority of model experiments, the effect of low positive temperatures on the expression of IRIP and cold-induced PRP genes was assessed, while only a few studies examined the impact of subzero temperatures (Herman et al., 2006; Kang et al., 2013; Willick et al., 2018). As a result, researchers can currently draw conclusions only about the association of these genes with cold tolerance. However, the actual need for these proteins arises under freezing conditions. Simulating field-like subzero temperatures in experimental setups could not only deepen our understanding of IRIP and PRP gene expression but also bring us closer to understanding how their functions are realized in real-world conditions.

Understanding the regulatory mechanisms controlling AFP gene activity remains a relevant problem, as the overall picture of their regulation is still unclear, especially regarding the protein factors involved in gene expression. The hormonal regulation patterns of wheat AFPs are also not fully understood. It is highly likely that the dual nature of plant AFPs has led to the development of a regulatory system in which signaling pathways for both biotic and abiotic stress responses are involved. This hypothesis requires particularly thorough investigation.

The promising results of practical applications of wheat AFPs, along with evidence linking cold tolerance, gene expression, and protein accumulation in winter wheat, support the need for continued research into this group of low-temperature adaptation proteins.

References

Ambroise V., Legay S., Guerriero G., Hausman J.F., Cuypers A., Sergeant K. The roots of plant frost hardiness and tolerance. *Plant Cell Physiol*. 2020;61(1):3-20. doi 10.1093/pcp/pcz196

Antikainen M., Griffith M. Antifreeze protein accumulation in freezing-tolerant cereals. *Physiol Plant.* 1997;99(3):423-432. doi 10.1111/j.1399-3054.1997.tb00556.x

Badawi M., Reddy Y.V., Agharbaoui Z., Tominaga Y., Danyluk J., Sarhan F., Houde M. Structure and functional analysis of wheat ICE (inducer of CBF expression) genes. *Plant Cell Physiol.* 2008;49(8):1237-1249. doi 10.1093/pcp/pcn100

Båga M., Chodaparambil S.V., Limin A.E., Pecar M., Fowler D.B., Chibbar R.N. Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. *Funct Integr Genomics.* 2007;7(1):53-68. doi 10.1007/s10142-006-0030-7

Bayer-Giraldi M., Weikusat I., Besir H., Dieckmann G. Characterization of an antifreeze protein from the polar diatom *Fragilariopsis cylindrus* and its relevance in sea ice. *Cryobiology*. 2011;63(3):210-219. doi 10.1016/j.cryobiol.2011.08.006

Boonsupthip W., Lee T.-C. Application of antifreeze protein for food preservation: effect of type III antifreeze protein for preservation of gel-forming of frozen and chilled actomyosin. *J Food Sci*. 2003;68(5):1804-1809. doi 10.1111/j.1365-2621.2003.tb12333.x

Cao Y., Hu G., Zhuang M., Yin J., Wang X. Molecular cloning and functional characterization of *TaIRI9* gene in wheat (*Triticum aestivum* L.). *Gene*. 2021;791:145694. doi 10.1016/j.gene.2021.145694

Case A.J., Skinner D.Z., Garland-Campbell K.A., Carter A.H. Freezing tolerance-associated quantitative trait loci in the brundage × coda wheat recombinant inbred line population. *Crop Sci.* 2014;54(3):982-992. doi 10.2135/cropsci2013.08.0526

Chinnusamy V., Ohta M., Kanrar S., Lee B.H., Hong X., Agarwal M., Zhu J.K. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes Dev*. 2003;17(8):1043-1054. doi 10.1101/gad.1077503

Chow-Shi-Yée M., Briard J.G., Grondin M., Averill-Bates D.A., Ben R.N., Ouellet F. Inhibition of ice recrystallization and cryoprotective activity of wheat proteins in liver and pancreatic cells. *Protein Sci.* 2016;25(5):974-986. doi 10.1002/pro.2903

Chow-Shi-Yée M., Grondin M., Ouellet F., Averill-Bates D.A. Control of stress-induced apoptosis by freezing tolerance-associated wheat proteins during cryopreservation of rat hepatocytes. *Cell Stress Chaperones*. 2020;25(6):869-886. doi 10.1007/s12192-020-01115-y

Chun J., Yu X., Griffith M. Genetic studies of antifreeze proteins and their correlation with winter survival in wheat. *Euphytica.* 1998;102:219-226. doi 10.1023/A:1018333730936

DeVries A.L., Wohlschlag D.E. Freezing resistance in some Antarctic fishes. *Science.* 1969;163(3871):1073-1075. doi 10.1126/science.163.3871.1073

Duman J.G., Olsen T.M. Thermal hysteresis protein activity in bacteria, fungi, and phylogenetically diverse plants. *Cryobiology*. 1993;30(3):322-328. doi 10.1006/cryo.1993.1031

Duman J.G., Wisniewski M.J. The use of antifreeze proteins for frost protection in sensitive crop plants. *Env Exp Bot*. 2014;106:60-69. doi 10.1016/j.envexpbot.2014.01.001

Ewart K.V., Lin Q., Hew C.L. Structure, function and evolution of antifreeze proteins. *Cell Mol Life Sci*. 1999;55(2):271-283. doi 10.1007/s000180050289

Gaudet D.A., Laroche A. Winter survival of cereals parasitized by Snow Mold. In: Li P.H., Chen T.H.H. (Eds) Plant Cold Hardiness. Boston: Springer, 1997;331-342. doi 10.1007/978-1-4899-0277-1_31

Gaudet D.A., Laroche A., Frick M., Davoren J., Puchalski B., Ergon Å. Expression of plant defence-related (PR-protein) transcripts during hardening and dehardening of winter wheat. *Physiol Mol Plant Pathol.* 2000;57(1):15-24. doi 10.1006/pmpp.2000.0275

Griffith M., Yaish M.W. Antifreeze proteins in overwintering plants: a tale of two activities. *Trends Plant Sci.* 2004;9(8):399-405. doi 10.1016/j.tplants.2004.06.007

Griffith M., Ala P., Yang D.S., Hon W.C., Moffatt B.A. Antifreeze protein produced endogenously in winter rye leaves. *Plant Physiol*. 1992;100(2):593-596. doi 10.1104/pp.100.2.593

Griffith M., Lumb C., Wiseman S.B., Wisniewski M., Johnson R.W., Marangoni A.G. Antifreeze proteins modify the freezing process in planta. *Plant Physiol*. 2005;138(1):330-340. doi 10.1104/pp.104.058628

Grondin M., Hamel F., Averill-Bates D.A., Sarhan F. Wheat proteins improve cryopreservation of rat hepatocytes. *Biotechnol Bioeng*. 2009;103(3):582-591. doi 10.1002/bit.22270

Hannah M.A., Heyer A.G., Hincha D.K. A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. *PLoS Genet*. 2005;1(2):e26. doi 10.1371/journal.pgen.0010026

He K., Wu Y. Receptor-like kinases and regulation of plant innate immunity. *Enzymes*. 2016;40:105-142. doi 10.1016/bs.enz.2016.09.003

Herman E.M., Rotter K., Premakumar R., Elwinger G., Bae H., Ehler-King L., Chen S., Livingston D.P. 3rd. Additional freeze hardiness in wheat acquired by exposure to –3 degrees C is associated with extensive physiological, morphological, and molecular changes. *J Exp Bot.* 2006;57(14):3601-3618. doi 10.1093/jxb/erl111

Hon W.C., Griffith M., Mlynarz A., Kwok Y.C., Yang DSC. Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiol.* 1995;109(3):879-889. doi 10.1104/pp.109.3.879

Houde M., Belcaid M., Ouellet F., Danyluk J., Monroy A.F., Dryanova A., Gulick P., Bergeron A., Laroche A., Links M.G., MacCarthy L., Crosby W.L., Sarhan F. Wheat EST resources for functional genomics of abiotic stress. *BMC Genomics*. 2006;7:149. doi 10.1186/1471-2164-7-149

Huang T., Duman J.G. Cloning and characterization of a thermal hysteresis (antifreeze) protein with DNA-binding activity from winter bittersweet nightshade, *Solanum dulcamara*. *Plant Mol Biol*. 2002; 48(4):339-350. doi 10.1023/A:1014062714786

Jain D., Khurana J.P. Role of pathogenesis-related (PR) proteins in plant defense mechanism. In: Singh A., Singh I. (Eds) Molecular Aspects of Plant-Pathogen Interaction. Singapore: Springer, 2018;265-281. doi 10.1007/978-981-10-7371-7_12

Janech M.G., Krell A., Mock T., Raymond J.A. Ice-binding proteins from sea ice diatoms (Bacillariophyceae). *J Phycol*. 2006;42(2):410-416. doi 10.1111/j.1529-8817.2006.00208.x

Jarząbek M., Pukacki P.M., Nuc K. Cold-regulated proteins with potent antifreeze and cryoprotective activities in spruces (*Picea* sp.). *Cryobiology*. 2009;58(3):268-274. doi 10.1016/j.cryobiol.2009.01.007

Jin Y.N., Bai L.P., Guan S.X., Zhong M., Ma H., Wang S., Guo Z.F. Identifcation of an ice recrystallisation inhibition gene family in winter-hardy wheat and its evolutionary relationship to other members of the *Triticeae*. *J Agron Crop Sci.* 2018;204(4):400-413. doi 10.1111/jac.12270

Jin Y., Ding X., Li J., Guo Z. Isolation and characterization of wheat ice recrystallisation inhibition gene promoter involved in low temperature and methyl jasmonate responses. *Physiol Mol Biol Plants.* 2022;28(11-12):1969-1979. doi 10.1007/s12298-022-01257-6

Juurakko C.L., diCenzo G.C., Walker V.K. *Brachypodium* antifreeze protein gene products inhibit ice recrystallisation, attenuate ice nucleation, and reduce immune response. *Plants.* 2022;11(11):1475. doi 10.3390/plants11111475

Kang G., Li G., Yang W., Han Q., Ma H., Wang Y., Ren J., Zhu Y., Guo T. Transcriptional profile of the spring freeze response in the leaves of bread wheat (*Triticum aestivum* L.). *Acta Physiol Plant.* 2013;35(02):575-587. doi 10.1007/s11738-012-1099-3

Knight C.A., Wen D., Laursen R.A. Nonequilibrium antifreeze peptides and the recrystallization of ice. *Cryobiology*. 1995;32(1):23-34. doi 10.1006/cryo.1995.1002

Kontogiorgos V., Regand A., Yada R., Goff H.D. Isolation and characterization of ice structuring proteins from cold-acclimated winter wheat grass extract for recrystallization inhibition in frozen foods. *J Food Biochem*. 2007;31(2):139-160. doi 10.1111/j.1745-4514.2007.00112.x

Kostyaev A.A., Martusevich A.K., Andreev A.A. Toxicity of cryoprotectants and cryoconservants on their basis for blood components and bone marrow (review article). *Nauchnoe Obozrenie. Meditsinskie Nauki = Scientific Review. Medical Sciences.* 2016;6:54-74 (in Russian)

Kristiansen E., Pedersen S.A., Ramløv H., Zachariassen K.E. Antifreeze activity in the cerambycid beetle Rhagium inquisitor. *J Comp Physiol B.* 1999;169:55-60. doi 10.1007/s003600050193

Kristiansen E., Ramløv H., Højrup P., Pedersen S.A., Hagen L., Zachariassen K.E. Structural characteristics of a novel antifreeze protein from the longhorn beetle *Rhagium inquisitor*. *Insect Biochem Mol Biol.* 2011;41(2):109-117. doi 10.1016/j.ibmb.2010.11.002

Kruse E.B., Carle S.W., Wen N., Skinner D.Z., Murray T.D., Garland-Campbell K.A., Carter A.H. Genomic regions associated with tolerance to freezing stress and snow mold in winter wheat. *G3* (*Bethesda*). 2017;7(3):775-780. doi 10.1534/g3.116.037622

Kurepin L.V., Dahal K.P., Savitch L.V., Singh J., Bode R., Ivanov A.G., Hurry V., Hüner N.P. Role of CBFs as integrators of chloroplast redox, phytochrome and plant hormone signaling during cold acclimation. *Int J Mol Sci*. 2013;14(6):12729-12763. doi 10.3390/ijms140612729

Liu X., Pan Y., Liu F., He Y., Zhu Q., Liu Z., Zhan X., Tan S. A review of the material characteristics, antifreeze mechanisms, and applications of cryoprotectants (CPAs). *J Nanomater*. 2021;2021(1):9990709. doi 10.1155/2021/9990709

Livingston D.P. 3rd, Bertrand A., Wisniewski M., Tisdale R., Tuong T., Gusta L.V., Artlip T. Factors contributing to ice nucleation and sequential freezing of leaves in wheat. *Planta*. 2021;253(6):124. doi 10.1007/s00425-021-03637-w

Macháčcková I., Hanisova A., Krekule J. Levels of ethylene, ACC, MACC, ABA and proline as indicators of cold hardening and frost resistance in winter wheat. *Physiol Plant*. 1989;76(4):603-607. doi 10.1111/j.1399-3054.1989.tb05486.x

McHale L., Tan X., Koehl P., Michelmore R.W. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 2006;7(4):212. doi 10.1186/gb-2006-7-4-212

Middleton A.J., Brown A.M., Davies P.L., Walker V.K. Identification of the ice-binding face of a plant antifreeze protein. *FEBS Lett*. 2009; 583(4):815-819. doi 10.1016/j.febslet.2009.01.035

Middleton A.J., Marshal C.B., Faucher F., Bar-Dolev M., Braslavsky I., Campbell R.L., Walker V.K., Davies P.L. Antifreeze protein from freeze-tolerant grass has a beta-roll fold with an irregularly structured ice-binding site. *J Mol Biol*. 2012;416(5):713-724. doi 10.1016/j.jmb.2012.01.032

Monroy A.F., Dryanova A., Malette B., Oren D.H., Ridha Farajalla M., Liu W., Danyluk J., Ubayasena L.W., Kane K., Scoles G.J., Sarhan F., Gulick P.J. Regulatory gene candidates and gene expression analysis of cold acclimation in winter and spring wheat. *Plant Mol Biol.* 2007;64(4):409-423. doi 10.1007/s11103-007-9161-z

Pandey S.P., Somssich I.E. The role of WRKY transcription factors in plant immunity. *Plant Physiol*. 2009;150(4):1648-1655. doi 10.1104/pp.109.138990

Puchkov E.O. Biogenic control of ice formation. *Priroda.* 2017;2:27-37 (in Russian)

Regand A., Goff H.D. Ice recrystallization inhibition in ice cream as affected by ice structuring proteins from winter wheat grass. *J Dairy Sci*. 2006;89(1):49-57. doi 10.3168/jds.S0022-0302(06)72068-9

Sandve S.R., Rudi H., Asp T., Rognli O.A. Tracking the evolution of a cold stress associated gene family in cold tolerant grasses. *BMC Evol Biol*. 2008;8:245. doi 10.1186/1471-2148-8-245

Sauter M. Phytosulfokine peptide signaling. *J Exp Bot*. 2015;66(17): 5161-5169. doi 10.1093/jxb/erv071

Sutka J. Genetic control of frost tolerance in wheat (*Triticum aestivum* L.). *Euphytica*. 1994;77:277-282. doi 10.1007/BF02262642

Sutka J. Genes for frost resistance in wheat. *Euphytica.* 2001;119:169-177. doi 10.1023/A:1017520720183

Sutka J., Galiba G., Snape J.W. Inheritance of frost resistance in wheat (*Triticum aestivum* L.). *Acta Agron Hung.* 1997;45:257-263

Talanova V.V., Titov A.F., Topchieva L.V., Malysheva I.E., Venzhik Y.V., Frolova S.A. Expression of genes encoding the WRKY transcription factor and heat shock proteins in wheat plants during cold hardening. *Dokl Biol Sci*. 2008;423:440-442. doi 10.1134/s0012496608060215

Tchagang A.B., Fauteux F., Tulpan D., Pan Y. Bioinformatics identification of new targets for improving low temperature stress tolerance in spring and winter wheat. *BMC Bioinformatics*. 2017;18(1):174. doi 10.1186/s12859-017-1596-x

Tremblay K., Ouellet F., Fournier J., Danyluk J., Sarhan F. Molecular characterization and origin of novel bipartite cold-regulated ice recrystallization inhibition proteins from cereals. *Plant Cell Physiol.* 2005;46(6):884-891. doi 10.1093/pcp/pci093

Urrutia M.E., Duman J.G., Knight C.A. Plant thermal hysteresis proteins. *Biochem Biophys Acta.* 1992;1121(1-2):199-206. doi 10.1016/0167-4838(92)90355-h

Vaitkevičiūte G., Aleliūnas A., Brazauskas G., Armonienė R. Deacclimation and reacclimation processes in winter wheat: novel perspectives from time-series transcriptome analysis. *Front Plant Sci.* 2024; 15:1395830. doi 10.3389/fpls.2024.1395830

Vanková R., Kosová K., Dobrev P., Vítámvás P., Trávníčková A., Cvikrová M., Pešek B., Gaudinová A., Přerostová S., Musilová J.,

Galiba G., Prasil I.T. Dynamics of cold acclimation and complex phytohormone responses in *Triticum monococcum* lines G3116 and DV92 differing in vernalization and frost tolerance level. *Env Exp Bot*. 2014;101:12-25. doi 10.1016/j.envexpbot.2014.01.002

Voets I.K. From ice-binding proteins to bio-inspired antifreeze materials. *Soft Matter.* 2017;13(28):4808-4823. doi 10.1039/c6sm02867e

Wang W., Hao Q., Wang W., Li Q., Wang W. The genetic characteristics in cytology and plant physiology of two wheat (*Triticum aestivum*) near isogenic lines with different freezing tolerances. *Plant Cell Rep*. 2017;36(11):1801-1814. doi 10.1007/s00299-017-2195-z

Wang W., Wang X., Zhang J., Huang M., Cai J., Zhou Q., Dai T., Jiang D. Salicylic acid and cold priming induce late-spring freezing tolerance by maintaining cellular redox homeostasis and protecting photosynthetic apparatus in wheat. *Plant Growth Regul.* 2020;90:109-121. doi 10.1007/s10725-019-00553-8

Wang W., Wang X., Huang M., Cai J., Zhou Q., Dai T., Jiang D. Alleviation of field low-temperature stress in winter wheat by exogenous application of salicylic acid. *J Plant Growth Regul*. 2021;40:811-823. doi 10.1007/s00344-020-10144-x

Wang X., Geng H., Wu D., Wang L., Zhang N., Wang W., Yu D. Isolation of ice structuring proteins from winter wheat in frigid region (Dongnongdongmai1) and the effect on freeze-thaw stability of dough. *Food Res Int.* 2024;197(Pt. 2):115295. doi 10.1016/j.foodres.2024.115295

Willick I.R., Takahashi D., Fowler D.B., Uemura M., Tanino K.K. Tissue-specific changes in apoplastic proteins and cell wall structure during cold acclimation of winter wheat crowns. *J Exp Bot*. 2018;69(5):1221-1234. doi 10.1093/jxb/erx450

Willick I.R., Gusta L.V., Fowler D.B., Tanino K.K. Ice segregation in the crown of winter cereals: evidence for extraorgan and extratissue freezing. *Plant Cell Environ*. 2019;42(2):701-716. doi 10.1111/pce.13454

Winfield M.O., Lu C., Wilson I.D., Coghill J.A., Edwards K.J. Plant responses to cold: transcriptome analysis of wheat. *Plant Biotechnol J*. 2010;8(7):749-771. doi 10.1111/j.1467-7652.2010.00536.x

Wisniewski M., Willick I.R., Duman J.G., Livingston D.P. 3rd, Newton S.S. Plant antifreeze proteins. In: Ramløv H., Friis D. (Eds) Antifreeze Proteins. Vol. 1. Cham: Springer, 2020;189-226. doi 10.1007/978-3-030-41929-5_7

Xue-Xuan X., Hong-Bo S., Yuan-Yuan M., Gang X., Jun-Na S., Dong-Gang G., Cheng-Jiang R. Biotechnological implications from abscisic acid (ABA) roles in cold stress and leaf senescence as an important signal for improving plant sustainable survival under abiotic-stressed conditions. *Crit Rev Biotechnol*. 2010;30(3):222-230. doi 10.3109/07388551.2010.487186

Yang T., Zhang Y., Guo L., Li D., Liu A., Bilal M., Xie C., Yang R., Gu Z., Jiang D., Wang P. Antifreeze polysaccharides from wheat bran: the structural characterization and antifreeze mechanism. *Biomacromolecules*. 2024;25(7):3877-3892. doi 10.1021/acs.biomac.3c00958

Yu X.M., Griffith M., Wiseman S.B. Ethylene induces antifreeze activity in winter rye leaves. *Plant Physiol*. 2001;126(3):1232-1240. doi 10.1104/pp.126.3.1232

Zheng X., Shi M., Wang J., Yang N., Wang K., Xi J., Wu C., Xi T., Zheng J., Zhang J. Isoform sequencing provides insight into freezing response of common wheat (*Triticum aestivum* L.). *Front Genet*. 2020;11:462. doi 10.3389/fgene.2020.00462

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

Received May 13, 2025. Revised September 15, 2025. Accepted September 24, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):403-411

doi 10.18699/vjgb-26-44

Adaptive responses of pear cultivars to low-temperature stress in the spring period

A.E. Mishko, A.V. Klyukina

North Caucasian Federal Scientific Center of Horticulture, Viticulture, Winemaking, Krasnodar, Russia
mishko-alisa@mail.ru

Abstract. The pear is one of the most famous pome crops. It occupies about 7 % of the total area of perennial fruit crops in Russia. Orchard plantings are predominantly composed of foreign European cultivars. Spring frosts, which are typical for the southern regions of the country, lead to significant crop losses. This study determined the response characteristics of pear flower buds to low-temperature stress. The Crimean cultivar Dzhankoyskaya Pozdnyaya, two cultivars – Leven and Flamenco – of Krasnodar selection and the interspecific hybrid Kieffer were investigated. Flower buds at different developmental stages were exposed to a climatic chamber for 12 hours at temperatures –1.5…–2 °C. After stress exposure, the activity of certain antioxidant enzymes was determined, along with the content of phenolic compounds, malondialdehyde, and the gene expression level of its enzymes and proteins involved in cold adaptation. It was revealed that the autumn-ripening cultivar Kieffer, under conditions of the Krasnodar region, begins to bloom earlier than other studied cultivars, making it more susceptible to recurrent frosts. This is evidenced by high values of malondialdehyde and the activity level of superoxide dismutase. The Russian cultivars, Leven (winter cultivar) and Flamenco (summer cultivar), showed the highest activity of peroxidase and gene expression of *PcDREB2*, *PcCAP160*, *PcCOR413*, *PcPOX1*, with a reduced level of malondialdehyde. These cultivars typically emerged from dormancy later compared to Kieffer. The Crimean winter-ripening cultivar was closer to the interspecific hybrid in terms of the studied parameters but showed lower enzyme activity and gene expression levels. The obtained results suggest that under pear cultivation conditions in the southern region of the country, where spring frosts are possible, cultivars with flowering starting in the second-to-third decade of April and high indicators of antioxidant enzyme activity (primarily peroxidase) and gene expression levels of *PcDREB2*, *PcCAP160*, and *PcCOR413* demonstrate greater resistance.

Key words: pear cultivars; low-temperature stress; resistance; antioxidant system defense; gene expression

For citation: Mishko A.E., Klyukina A.V. Adaptive responses of pear cultivars to low-temperature stress in the spring period. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):403-411. doi 10.18699/vjgb-26-44

Funding. The study was carried out with the financial support of the state budget project FGRE-2022-0001 (No. 225020408372-9) of the Ministry of Science and Higher Education of the Russian Federation.

Пути адаптации разных сортов груши к низкотемпературному стрессу в весенний период

А.Е. Мишко, А.В. Клюкина

Северо-Кавказский федеральный научный центр садоводства, виноградарства, виноделия, Краснодар, Россия
mishko-alisa@mail.ru

Аннотация. Хорошо известная семечковая культура – груша – на территории России занимает около 7 % общих площадей насаждений многолетних плодовых культур. Основную долю грушевых садов составляют зарубежные европейские сорта. Возвратные заморозки, которые характерны для южных регионов страны в весенний период, приводят к значительным потерям урожая. В настоящей работе изучены особенности ответных реакций цветочных почек сортов груши на низкотемпературный стресс. Исследовали крымский сорт Джанкойская поздняя, два краснодарских сорта Левен и Фламенко, а также межвидовой гибрид Киффер. Цветочные почки на разных стадиях развития промораживали в климатической камере в течение 12 ч при температуре –1.5…–2.0 °C. После стресса определяли активность некоторых ферментов антиоксидантной системы, содержание фенольных соединений, малонового диальдегида и уровень экспрессии генов тех же ферментов и белков, участвующих в холодовой адаптации. Выявлено, что осенний сорт Киффер в условиях Краснодарского края быстрее других изученных сортов начинает цвести, вследствие чего наиболее подвержен воздействию возвратных заморозков, о чем свидетельствуют высокие значения малонового диальдегида и уровень активности супероксиддисмутазы. Отечественные сорта Левен (зимний сорт) и Фламенко (летний) обладали самыми высокими показателями активности пероксидазы и экспрессии генов *PcDREB2*, *PcCAP160*, *PcCOR413*, *PcPOX1* на фоне сниженного уровня малонового диальдегида. Как правило, данные сорта позднее выходили из состояния покоя по сравнению

с сортом Киффер. Крымский сорт зимнего срока созревания был ближе по исследуемым параметрам к межвидовому гибриду, но отличался меньшими показателями ферментативной активности и экспрессией рассматриваемых генов. Полученные результаты позволяют заключить, что в условиях произрастания груши на территории южного региона страны, где возможны возвратные заморозки в весенний период, большей физиологической устойчивостью обладают сорта с началом цветения во второй-третьей декадах апреля, характеризующиеся высокими показателями антиоксидантной ферментативной активности, в первую очередь пероксидазы, и уровнем экспрессии генов *PcDREB2*, *PcCAP160* и *PcCOR413*.

Ключевые слова: груша; холодовой стресс; устойчивость; антиоксидантная система защиты; экспрессия генов

Introduction

Pome crops, and pears in particular, are a valuable resource for the horticulture of the Krasnodar region. The share of pears in the total plantings of perennial fruits is extremely small and amounts to about 7 % of the total, though the economic interest in the crop keeps growing. The reasons for its limited distribution, even under the most favorable conditions of southern Russia, are obvious: compared to apple trees, pears are more demanding of growing conditions, mainly water supply, especially at the beginning of the growing season and throughout fruit ripening (Asayesh et al., 2023). The frost resistance of the crop is high: according to various researchers, in a state of organic dormancy, pear flower buds can withstand temperatures up to –25 °C (Sotnik et al., 2017; Bandurko, 2024). Not only sub-zero temperatures and frost duration are critical, but also the phase of the development of generative organs (Xiao et al., 2022).

One of the essential features of adaptation to growing conditions is the resistance of fruit crop cultivars to low-temperature stress during the growing season. In the Krasnodar region, late-spring frosts, which have become more common recently, cause significant economic losses. Based on the long-term weather monitoring, the main stress factors for pears during the winter–spring period have been revealed: the absence of low temperatures in December–January that are necessary for plant hardening, and frosts at the beginning of the growing season (March–April). Each of these stress factors might lead to a reduction in yield or a complete loss of harvest in pear cultivars (Evers et al., 2021; Lee et al., 2023; Zhao et al., 2023). It has been shown that any crop – and any particular cultivar – has its own threshold of critical minimum temperatures that may trigger the death of flower buds (Klyukina et al., 2024). Moreover, different cultivars are characterized by different timeframes of the deep dormancy period: its duration under certain temperature conditions will determine the speed of dormancy release (Gabay, Flaishman, 2024).

The resistance of cultivars to low-temperature stress is composed of the complex interaction of physiological, biochemical, and molecular genetic processes that are reflected in a number of indicators. Any significant stress experienced by a plant provokes secondary oxidative stress caused by an increase in reactive oxygen species (ROS) in cells (Suzuki et al., 2012). One of the key protective reactions to such a negative impact is the increased activity of antioxidant enzymes that inhibit the accumulation of free radicals, which might lead to the destruction of lipids, proteins, and nucleic acids, to the point of cell death. Superoxide dismutase neutralizes superoxide anion, while peroxidases neutralize hydrogen peroxide (Dumanovic et al., 2021). Polyphenol oxidase catalyzes the oxidation reactions of various phenolic compounds, synthesizing or degrading the metabolites necessary for maintaining plant homeostasis (Zhang S., 2023).

Along with enzymes, numerous phenolic compounds also perform a protective function (Dumanovic et al., 2021). Flavonoids are one of the main phenolic antioxidants. They are capable of reducing ROSs to stable and less harmful forms (Treml, Mejkal, 2016). At the genetic level, low-temperature stress induces the expression of functional and regulatory genes, the products of which either directly protect cells from damage (osmolite biosynthesis enzymes, detoxification enzymes, etc.) or participate in signal transduction and the regulation of gene expression (Yang, Huang, 2018; Zhang Y. et al., 2023).

Due to the increased frequency of late frosts in the Krasnodar region, the aim of this research is to study the response of pear cultivars to low-temperature stress during the spring season.

Material and methods

The survey was conducted in the Kuban horticultural zone of the Krasnodar region (45°16′N, 38°93′E) in March–April 2023–2025 on the genetic collection of the North Caucasian Federal Scientific Center of Horticulture, Viticulture, Winemaking (NCFSCHVW). The objects of the study were the pear (*Pyrus communis* L.) cultivars of domestic origin Leven and Flamenco (NCFSCHVW breeding), the Crimean cultivar Dzhankoyskaya Pozdnyaya and the American cultivar Kieffer (an interspecific hybrid of *P. communis* × *P. pyrifolia* Nakai). The cultivars were grafted onto the BA-29 rootstock. The planting year was 2007, and the planting pattern was 5 × 2 m. Ten shoots with flower buds were selected from 5–7 trees of each cultivar.

The response of generative organs to low spring temperatures was analyzed by the critical minimum temperatures for each phase of flower bud development in the BPC500D/CVSI-Spector climate chamber (Fujian Jiupo Biotechnology Co, China) using the standard methodology (Program and Methodology of Varietalization…, 1999; New Methods…, 2023). Phases of flower bud development were determined according to the generally accepted techniques (Kolomiyts, 1952; Shitt, 1958). The temperature data for the studied period were obtained from a local weather station (synoptic index 34927). Natural late frosts were simulated under laboratory conditions, with a gradual decrease in temperature to a critical level of –1.5 to –2.0 °C over 24 hours. Control samples were not exposed to low-temperature stress. Upon completing the experiment, the plant material was frozen in liquid nitrogen for further use in a series of biochemical and molecular genetic analyses.

The total content of phenolic compounds and flavonoids was determined in pear flower buds (Ainsworth, Gillespie, 2007; Hikmawanti et al., 2021). The extraction of soluble proteins was conducted using the method outlined by Z. Wei and colleagues (2018). The concentration of soluble proteins was assessed by coloring with Coomassie solution (Bradford, 1976). The activity of superoxide dismutase (SOD), peroxidase (POX), and polyphenol oxidase (PPO) enzymes was measured according to standard colorimetric methods (Boyarkin, 1951; Queiroz et al., 2011; Efimova et al., 2018). The content of malondialdehyde (MDA) was identified by the reaction with thiobarbituric acid (Bonyanpour, Jamali, 2020).

To analyze gene expression levels, RNA was extracted from pear flower buds using a modified CTAB method (Sundyreva et al., 2018). cDNA synthesis was carried out using MMLV reverse transcriptase (Eurogen, Russia) according to the manufacturer's instructions. Real-time PCR was conducted by means of the qPCRmix-HS SYBR kit (Eurogen, Russia). The primers were selected from literature sources (Table 1). A conservative fragment of the actin gene sequence was used as a reference gene. The relative expression of the studied genes was calculated via the $2^{-\Delta\Delta Ct}$ method (Livak, Schmittgen, 2001).

Measurements were taken in 2–4-fold analytical repetition. The statistical significance of differences was determined based on the results of the Tukey test of one-way ANOVA at the 0.05 significance level. Statistical analysis was carried out using the STATISTICA 12 software. The results are presented as mean values and their standard errors.

Table 1. Sequences for primers used in quantitative real-time PCR

Gene	Primer sequence 5′→3′		Reference
PPO1	F	CCTACTCACAAAGCCCAAGC	Busatto et al., 2021
	R	CCTCCAAGACCAAGAAGCAC	
SOD1	F	GGGAGATGGCCCAACTACTG	Azarabadi et al., 2017
	R	CCAGTTGACATGCAACCGTT	
POX1	F	AAGGCATGCATGTGGTCAGT	Zhai et al., 2018
	R	CGACATATCCACCATGCCCA	
DREB2	F	GCAAAGAAACAGACCTTGTGC	Nham et al., 2017
	R	GCATATAAGTCGTCCATCAACC	
COR413	F	GGTCGAACAGCACTGAAGGA	
	R	CTCAAATGGGTTGCCTCCCT	
CAP160	F	GCCACTACTGCTATTGCCGA	
	R	ATACCCTGTTGCTCAGGTGC	
Actin	F	CTGCTGGCATTCATGAGACT	Şahin et al., 2022
	R	TCTGGTGGAGCTACAACCTT	

Note. F – forward primer sequence, R – reverse primer sequence.

Results

The temperature conditions during the study period were variable (Table 2), as can be seen in the phases of flower bud development (Fig. 1). The plant material was selected in the third decade of March or the first decade of April, when the pear starts blossoming.

The conditions for bursting of generative buds were favorable in 2023. After spring frosts in March (down to –6.3 °C), the temperature grew up to +5.4 °C in April, and this resulted in the early start of blossoming for the Kieffer cultivar, while Flamenco, Leven, and Dzhankoyskaya Pozdnyaya remained in the phase of flower bud separation.

The development of generative organs in 2024 was different from that in 2023. Temperatures in March dropped as low as –4.3 °C, while in April the minimum temperature was only +2.0 °C. Most of the cultivars (Leven, Dzhankoyskaya Pozdnyaya, and Flamenco) remained in the phase of bud burst, whereas Kieffer was in the phase of flower bud separation.

Over the three years of the study, the lowest temperature was recorded in March (down to –13.9 °C) only in 2025. The minimum air temperature in the first ten days of April was +0.2 °C. Under these conditions, it was found that Dzhankoyskaya Pozdnyaya and Flamenco were in the phase of bud burst, while Kieffer and Leven reached the phase of flower bud separation.

The total phenolic content in pear flower buds varied in different years of the study (Fig. 2*a*). The maximum values for most cultivars were recorded in 2023, averaging from 17.2 to 18.6 mg/g FW. The minimum values were registered in the spring of 2025: 5.1–11.7 mg/g FW. Upon short-term low-temperature stress, a significant increase in this parameter was observed only in 2025. Flamenco, Dzhankoyskaya Pozdnyaya, and Leven had almost a twofold increase in phenols. Through-

Table 2. Temperature regime in the winter-spring period of 2023–2025

Month	Minimum air temperature, °C								
	2023			2024			2025		
	I	II	III	I	II	III	I	II	III
January	**–13.4**	–9.7	–8.1	–11.3	**–14.2**	–12.0	**–6.2**	–5.5	–3.1
February	**–14.5**	–11.5	–8.4	–2.9	–1.6	**–4.5**	–8.3	–8.4	**–18.6**
March	**–6.3**	2.3	1.5	**–4.3**	–3.5	–0.8	**–13.9**	–0.6	0.2
April	**0.2**	5.4	5.2	**2.0**	10.1	2.0	**0.2**	0.3	3.1

Note. I–III – decades of the months; the absolute minimum temperatures for the month are shown in bold.

Fig. 1. Phases of pear flower bud development.
a – bud burst; *b* – flower bud separation; *c* – start of blossoming.

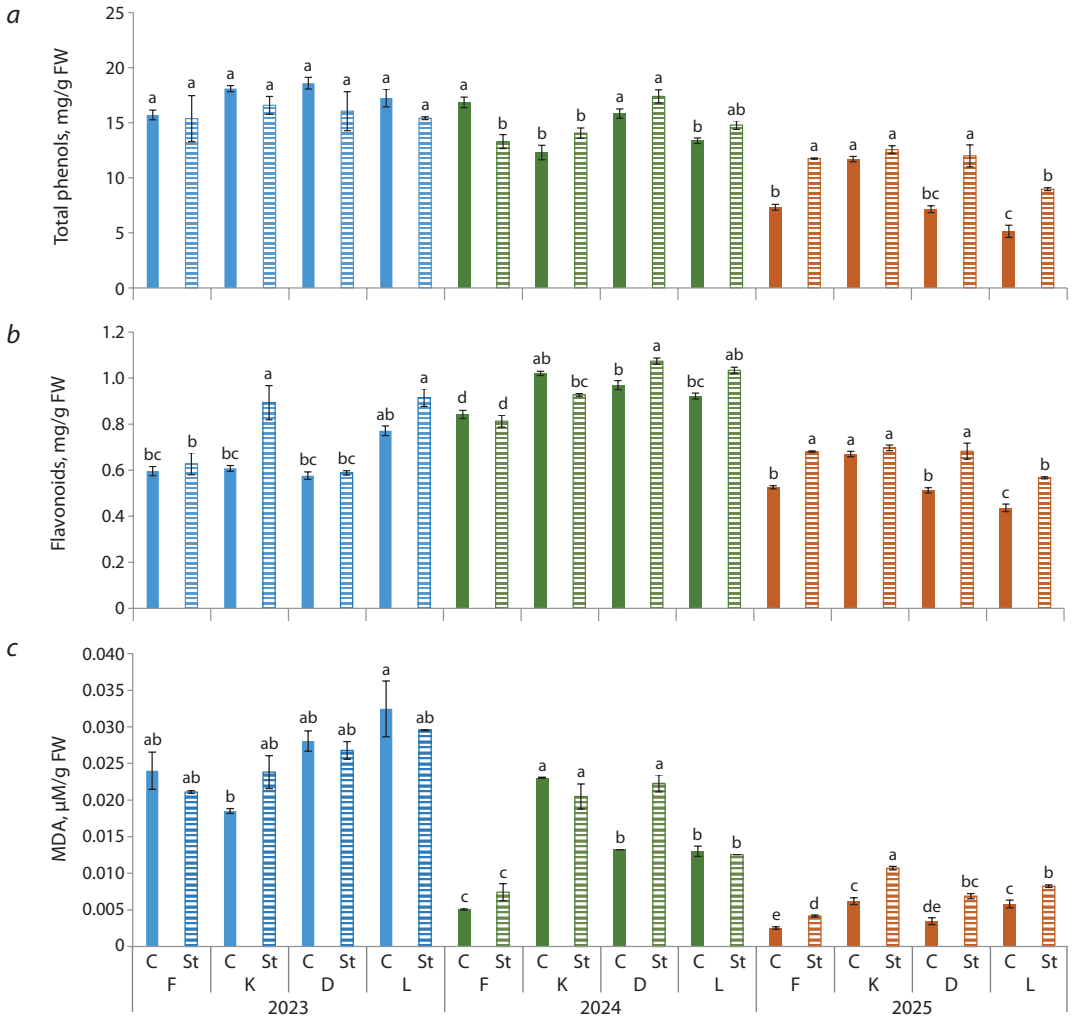


Fig. 2. Content of total phenols, flavonoids, malondialdehyde in pear flower buds under low-temperature stress conditions.

Here and in Figures 2–5: F – Flamenco, K – Kieffer, D – Dzhankoyskaya Pozdnyaya, L – Leven; C – control, St – stress; 2023–2024 – years; significant differences of the data are shown by different lowercase letters based on the results of the Tukey test at $p \leq 0.05$.

out the study period, Leven on average was characterized by the lowest values of this parameter, i. e. ~12 mg/g FW.

With respect to the flavonoid content, the highest values were recorded in 2024, ranging from 0.8 to 1.1 mg/g FW (Fig. 2*b*). In 2025, the accumulation of flavonoids in pear flower buds reached its minimum (0.4–0.6 mg/g FW). Due to the stress, the flavonoid content increased by more than 30 % in 2025, except for Kieffer, which surged by 1.5 times in 2023. Thus, during the three-year experiment, the maximum levels of flavonoids after stress were found in different cultivars: in

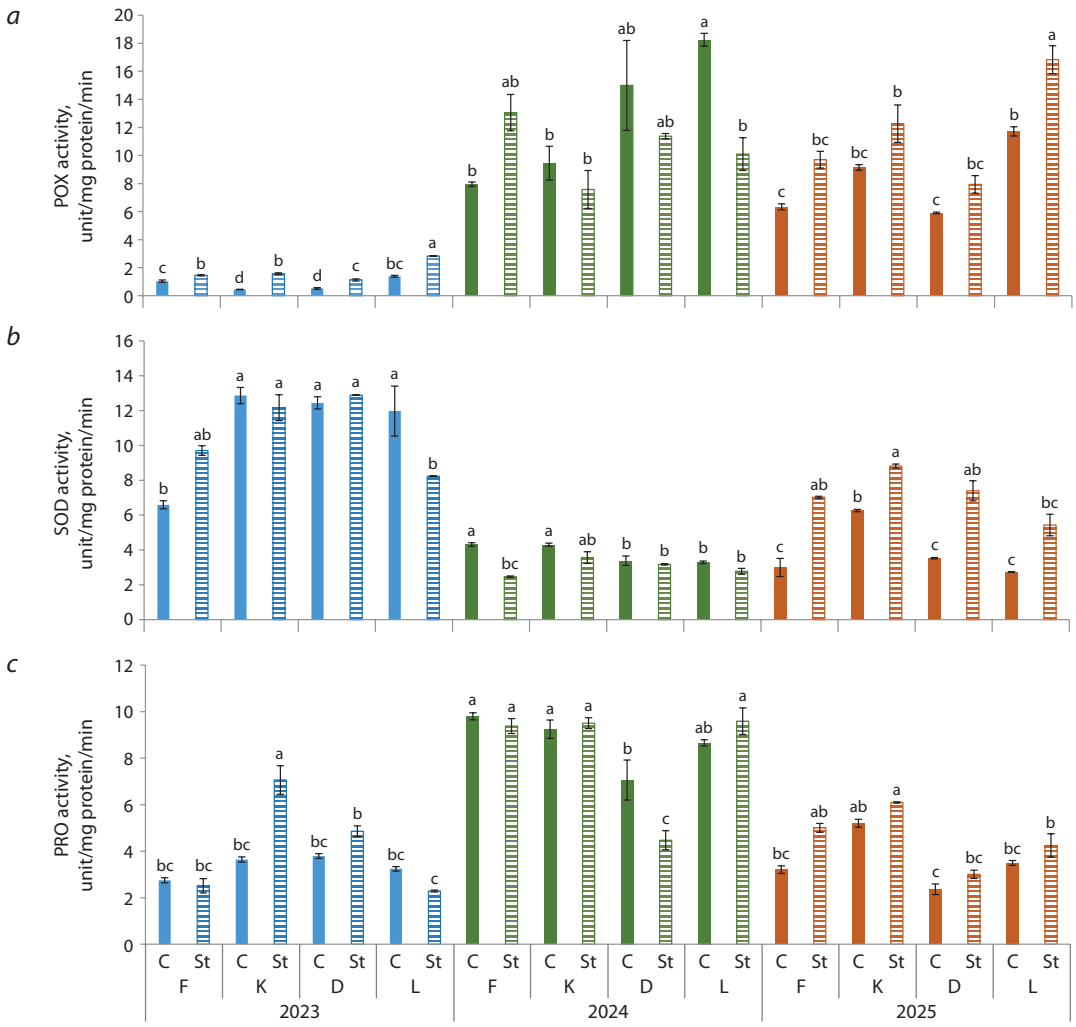


Fig. 3. Activities of peroxidase, superoxide dismutase, polyphenol oxidase in pear flower buds under low-temperature stress conditions in 2023–2025.

2023, in Kieffer and Leven; in 2024, in Dzhankoyskaya Pozdnyaya and Leven; and in 2025, in Flamenco, Dzhankoyskaya Pozdnyaya, and Kieffer.

Significant differences in the malondialdehyde accumulation were found depending on the year of the study. The highest values were specific to 2023, averaging 0.03 μM/g FW, and the lowest values were observed in 2025, averaging 0.004 μM/g FW (Fig. 2*c*). Low-temperature stress caused an increase in MDA by 1.5–2 times for all pear cultivars in 2025; for Flamenco and Dzhankoyskaya Pozdnyaya, by 45 and 65 %, respectively, in 2024; and by 30 %, for Kieffer in 2023. The analysis of the mean values of this parameter for the 2023–2025 period showed the maximum stress values in Leven, Dzhankoyskaya Pozdnyaya, and Kieffer.

Changes in the activity of the studied enzymes also depended on the year of the study. Peroxidase activity reached its maximum in the springs of 2024–2025 (from 5.9 to 18.3 unit/mg protein/min), and its minimum in 2023, ranging from 0.5 to 1.4 unit/mg protein/min (Fig. 3*a*). Most times stress led to an increase in POX activity, in some cases more than twofold. The highest enzyme activity levels were observed in Leven after exposure to low air temperatures under simulated conditions, averaging 11.6 units/mg protein/min over the entire study period.

The dynamics of superoxide dismutase activity were opposite to the changes of peroxidase activity, with peak growth in 2023 and its minimum in 2024. Throughout the study period, the highest SOD values were recorded in Kieffer, averaging 6.8 unit/mg protein (Fig. 3*b*). After stress, the enzyme activity in pear flower buds increased by 1.4–2.4 times only in 2025. In the course of three years, the highest values after short-term negative effects were found in Kieffer and Dzhankoyskaya Pozdnyaya, averaging 8.0 unit/mg protein.

The activity of polyphenol oxidase also varied from year to year: the maximum was in 2024, and the minimum was in 2023 and 2025 (Fig. 3*c*). Cultivar differences were not great; only for Dzhankoyskaya Pozdnyaya PPO values were the lowest in 2024 and 2025, 7.1 and 3.0 unit/mg protein, respectively. Low-temperature stress did not lead to a significant change in enzyme activity, except for two cases: a 2-fold PPO increase in Kieffer in 2023 and a 60 % PPO decrease in Dzhankoyskaya Pozdnyaya in 2024.

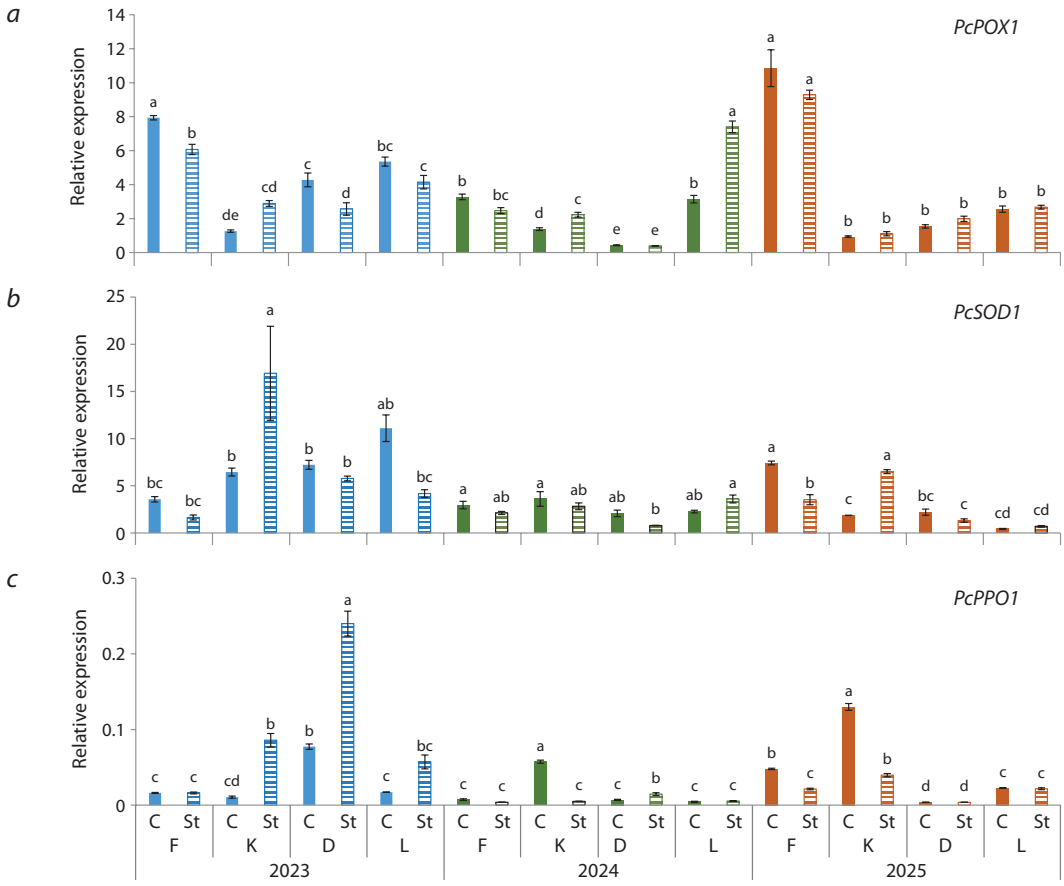


Fig. 4. Relative expression of the *PcPOX1*, *PcSOD1*, *PcPPO1* genes in pear flower buds under low-temperature stress conditions in 2023–2025.

According to the level of relative gene expression of the same antioxidant enzymes, the highest mean values of *PcPOX1* and *PcSOD1* were found in 2023 (Fig. 4*a, b*). Meanwhile, the maximum relative expression of the peroxidase gene was recorded for Kieffer in 2025. Regarding the *PcPPO1* gene, a high level of expression was observed in 2025, with a maximum in Kieffer (Fig. 4*c*). Subsequent to the stress, the expression level of antioxidant enzyme genes did not change significantly, except for certain cases: a spike in *PcSOD1* was observed in Kieffer in 2023 and 2025; *PcPOX1* also increased in Leven in 2024; a decrease was recorded in *PcPOX1* for Flamenco and Dzhankoyskaya Pozdnyaya in 2023, then in *PcSOD1* for Flamenco in 2025, and in *PcPPO1* for Kieffer in 2024–2025 (in 2023, on the contrary, an increase in this parameter was noted).

Genetic analysis was also conducted for a group of genes responsible for resistance to low-temperature stress: the transcription factor *PcDREB2* (Fig. 5*a*) and the proteins of cold acclimation *PcCAP160*, *PcCOR413* (Fig. 5*b, c*) were studied. No cultivar differences were found under the control conditions throughout the study, except for *PcCOR413* in Flamenco in 2023, which exceeded the values of the others by 78 %.

The exposure of pear generative buds to low temperatures caused a significant increase in the level of relative expression for *PcCOR413* – by 2–5 times in 2023 alone for all the studied cultivars. In Flamenco, an increase in stress values of this gene was detected in 2024 (by two times), since in the first year of the study, as noted earlier, the reference indicators were quite high. As for the level of relative expression for *PcCAP160*, its values under stressful conditions surged in 2023, especially in Kieffer, as well as in Flamenco and Leven in 2024 by 8–14 times, and in the latter cultivar, by three times in 2025. The changes in indicators for *PcDREB2* were even more dynamic. In 2023, the reference values were exceeded by 30–46 times for Flamenco and Dzhankoyskaya Pozdnyaya, by 109 times for Leven, and by 122 times for Kieffer with minimal values. In the following year, the growth of *PcDREB2* indicators was not so significant – by 2–17 times, though the expression level in Flamenco increased by 68 times. The maximum values for *PcDREB2* were recorded after the stress in Flamenco and Leven in 2025. An increase in expression, but less significant, was also registered in Kieffer and Dzhankoyskaya Pozdnyaya.

Discussion

According to the data obtained, it was revealed that physiological, biochemical, and molecular genetic characteristics of the generative bud are defined by its developmental stage. The "bud burst" stage is characterized by a high activity of POX and PPO enzymes, as well as accumulation of flavonoids. Under colder conditions in the spring of 2025, at the

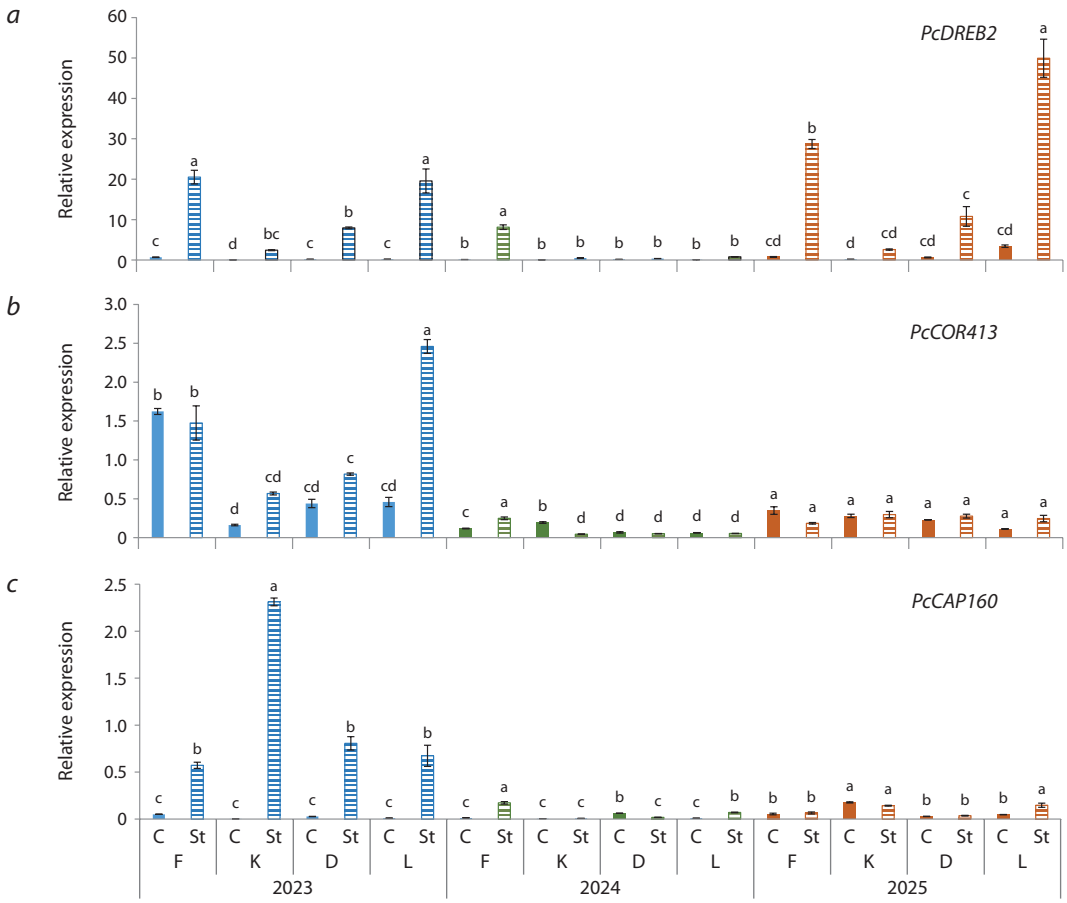


Fig. 5. Relative expression of the *PcDREB2*, *PcCOR413*, *PcCAP160* genes in pear flower buds under low-temperature stress conditions in 2023–2025.

same development stage of the generative structures, a more pronounced stress response was observed, taking the form of increased activity of peroxidase and superoxide dismutase, more intense accumulation of total phenols, flavonoids, MDA, and high expression of the *PcDREB2* gene. Similar results regarding the increase in antioxidant enzyme activity during the budding of pear *P. pyrifolia* were presented in the work by S. Hussain and colleagues (2015).

The "flower bud separation" stage is defined by the maximum accumulation of total phenols and MDA, the highest SOD activity, and a relatively high level of *PcCOR413* expression. Under stress, POX activity increased, and the expression of the *PcPPO1*, *PcDREB2*, *PcCOR413* and *PcCAP160* genes spiked. The "start of blossoming" stage, which was recorded only in Kieffer in 2023, is characterized by a more intense response to stress, including increased synthesis of flavonoids, PPO activity, and high levels of expression for all studied genes. When studying the response of flowers of different Asian pear genotypes to low-temperature stress from +2 to –4 °C for four hours, Chinese researchers observed an increase in SOD activity and in the proportion of gene transcripts related to flavonoid biosynthesis (Li et al., 2023; Lin et al., 2023). In another study, exposure of apple cell culture to low temperatures led to the activation of *DREB/CBF* gene expression (Du et al., 2015).

As shown above, the expression level of the *PcDREB2* gene in pear flower buds after stress was characterized by maximum growth (up to 100-fold increase), which corresponds to the data on the Ussuri pear *(P. ussuriensis* Maxim. ex Rupr.) – with the highest expression of the *DREB1* and *DREB2* genes in the first 12 hours under hypothermia (Yang, Huang, 2018). At the same time, in the fruits of the Williams cultivar, under exposure to low temperatures, no increase was detected in the expression of the *CBF1*, *CBF4*, *DREB2*, *COR413* genes, except for *CAP160* (Nham et al., 2017).

Based on the comparative analysis of the studied pear cultivars, it can be concluded that the autumn cultivar Kieffer has the ability to quickly emerge from a state of deep dormancy, which determines its earlier flowering period. As noted by G. Gabay and M.A. Flaishman, this is linked with the cultivar origin, since one of its parent forms is *P. pyrifolia*, which does not require low temperatures to enter a deep dormancy phase and quickly emerges from it after warm winters (Gabay, Flaishman, 2018). The relatively high stress levels of MDA are likely to contribute to its increased susceptibility to low temperatures during spring, resulting in a rapid response. Dzhankoyskaya Pozdnyaya, a winter-ripening cultivar, like Kieffer, is defined by a high content of phenolic compounds, as well as an increase in MDA content and SOD activity under stress conditions, but a less pronounced antioxidant defense system and lower

expression of cold resistance genes. Although Flamenco is a summer-ripening cultivar and Leven is winter-ripening, both cultivars had similar responses to low-temperature stress. On average, SOD activity and MDA content were at or below the reference values during the entire study period, while POX values and the expression of the *PcPOX1, PcDREB2*, *PcCAP160* and *PcCOR413* genes reached their maximum.

Conclusion

The conducted study revealed a difference in how pear cultivars of various ripening period and origin adapt to low-temperature stresses during winter and spring both at the ecological and at the physiological level. Pear plants with earlier flowering, such as Kieffer, are more susceptible to low temperatures due to increased SOD activity and high levels of MDA accumulation. In cultivars with later flowering (II–III decade of April), the protective mechanisms for containing oxidative stress in plant cells were activated more quickly: the maximum expression levels of cold resistance genes and peroxidase activity were observed at low MDA levels.

The obtained results will enable further evaluation of pear cultivars by cold resistance markers in order to create the latest inventory of promising cultivars and ensure their rational distribution in the southern regions of Russia that will result in stable and high-quality yields.

References

Ainsworth E.A., Gillespie K.M. Estimation of total phenolic content and other oxidation substrates in plant tissues using Folin–Ciocalteu reagent. *Nat Protoc*. 2007;2:875-877. doi 10.1038/nprot.2007.102

Asayesh Z.M., Arzani K., Mokhtassi-Bidgoli A., Abdollahi H. Enzymatic and non-enzymatic response of grafted and ungrafted young European pear (*Pyrus communis* L.) trees to drought stress. *Sci Hortic*. 2023;310:111745. doi 10.1016/j.scienta.2022.111745

Azarabadi S., Abdollahi H., Torabi M., Salehi Z., Nasiri J. ROS generation, oxidative burst and dynamic expression profiles of ROS-scavenging enzymes of superoxide dismutase (*SOD*), catalase (*CAT*) and ascorbate peroxidase (*APX*) in response to *Erwinia amylovora* in pear (*Pyrus communis* L.). *Eur J Plant Pathol*. 2017;147:279-294. doi 10.1007/s10658-016-1000-0

Bandurko I.A. Assessment of the gene pool of the Pear by resistance to adverse factors of the winter period in the foothill zone of the North-West Caucasus. *Science and Innovation.* 2024;21:88-92 (in Russian)

Bonyanpour A.R., Jamali B. Seasonal enzymatic and non-enzymatic antioxidant responses in seven Iranian pomegranate cultivars. *Adv Hortic Sci.* 2020;34(3):265-276

Boyarkin A.N. A quick method for determining the activity of peroxidase. *Biochemistry.* 1951;16(4):352-355 (in Russian)

Bradford M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal Biochem*. 1976;72(1-2):248-254. doi 10.1016/0003-2697(76)90527-3

Busatto N., Giné-Bordonaba J., Larrigaudière C., Lindo-Garcia V., Farneti B., Biasioli F., Vrhovsek U., Costa F. Molecular and biochemical differences underlying the efficacy of lovastatin in preventing the onset of superficial scald in a susceptible and resistant *Pyrus communis* L. cultivar. *Postharvest Biol Technol*. 2021;173:111435. doi 10.1016/j.postharvbio.2020.111435

Du F., Xu J.-N., Li D., Wang X.-Y. The identification of novel and differentially expressed apple-tree genes under low-temperature stress using high-throughput Illumina sequencing. *Mol Biol Rep*. 2015;42:569-580. doi 10.1007/s11033-014-3802-5

Dumanovic J., Nepovimova E., Natic M., Kuca K., Jacevic V. The significance of reactive oxygen species and antioxidant defense system in plants: a concise overview. *Front Plant Sci*. 2021;11:552969. doi 10.3389/fpls.2020.552969

Efimova M.V., Kolomeichuk L.V., Boyko E.V., Malofii M.K., Vidershpan A.N., Plyusnin I.N. Golovatskaya I.F., Murgan O.K., Kuznetsov Vl.V. Physiological mechanisms of *Solanum tuberosum* L. plants' tolerance to chloride salinity. *Russ J Plant Physiol.* 2018;65(3):394-403. doi 10.1134/S1021443718030020

Evers S.M., Knight T.M., Inouye D.W., Miller T.X., Salguero-Gómez R., Iler A.M., Compagnoni A. Lagged and dormant season climate better predict plant vital rates than climate during the growing season. *Glob Chang Biol.* 2021;27(9):1927-1941. doi 10.1111/gcb.15519

Gabay G., Flaishman M.A. Genetic and molecular regulation of chilling requirements in pear: breeding for climate change resilience. *Front Plant Sci.* 2024;15:1347527. doi 10.3389/fpls.2024.1347527

Hikmawanti N., Fatmawati S., Asri A.W. The effect of ethanol concentrations as the extraction solvent on antioxidant activity of katuk (*Sauropus androgynus* (L.) Merr.) leaves extracts. *IOP Conf Ser Earth Environ Sci.* 2021;755:012060. doi 10.1088/1755-1315/755/1/012060

Hussain S., Liu G., Liu D., Ahmed M., Hussain N., Teng Y. Study on the expression of dehydrin genes and activities of antioxidative enzymes in floral buds of two sand pear (*Pyrus pyrifolia* Nakai) cultivars requiring different chilling hours for bud break. *Turk J Agric For*. 2015;39(6):930-939. doi 10.3906/tar-1407-164

Klyukina A.V., Dragavtseva I.A., Oplachko R.A. Study of fruit crops varieties' needs in temperature regime of their development phases (on the example of pear varieties). *Fruit Growing and Viticulture of South Russia*. 2024;89(5):49-58. doi 10.30679/2219-5335-2024-5-89-49-58 (in Russian)

Kolomiyts I.A. Biological analysis of the development of flower buds in an apple tree. *Doklady Akademii Nauk SSSR.* 1952;84(4):821-824 (in Russian)

Lee J.C., Park Y.S., Jeong H.N., Kim J.H., Heo J.Y. Temperature changes affected spring phenology and fruit quality of apples grown in high-latitude region of South Korea. *Horticulturae*. 2023;9(7):794. doi 10.3390/horticulturae9070794

Li Y., Zhang J., Wang S., Zhang H., Liu Y., Yang M. Integrative transcriptomic and metabolomic analyses reveal the flavonoid biosynthesis of *Pyrus hopeiensis* flowers under cold stress. *Hortic Plant J*. 2023;9(3):395-413. doi 10.1016/j.hpj.2022.11.004

Lin S., Li Y., Zhao J., Guo W., Jiang M., Li X., Liu W., Zhang J., Yang M. Transcriptome analysis of biochemistry responses to low-temperature stress in the flower organs of five pear varieties. *Forests*. 2023;14(3):490. doi 10.3390/f14030490

Livak K.J., Schmittgen T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods*. 2001;25(4):402-408. doi 10.1006/meth.2001.1262

New Methods of Radical Increase in Yields of Fruit Crops Based on the Theory of the Environmental and Genetic Organization of Quantitative Features in the Context of Climate Fluctuation. Krasnodar, 2023 (in Russian)

Nham N.T., Macnish A.J., Zakharov F., Mitcham E.J. 'Bartlett' pear fruit (*Pyrus communis* L.) ripening regulation by low temperatures involves genes associated with jasmonic acid, cold response, and transcription factors. *Plant Sci*. 2017;260:8-18. doi 10.1016/j.plantsci.2017.03.008

Program and Methodology of Varietalization of Fruit, Berry and Walnut Crops. Orel, 1999 (in Russian)

Queiroz C., da Silva A.J.R., Lopes M.L.M., Fialho E., Valente-Mesquita V.L. Polyphenol oxidase activity, phenolic acid composition and browning in cashew apple (*Anacardium occidentale* L.) after processing. *Food Chem*. 2011;125(1):128-132. doi 10.1016/j.foodchem.2010.08.048

Şahin Ö., Dumanoğlu H., Sarikamiş G., Javadisaber J., Ergül A., Aydemir B.Ç. Tolerance of *Pyrus* spp. and *Cydonia oblonga* as pear rootstocks to iron chlorosis determined by in vitro growth, antioxidant and molecular responses. *Sci Hortic*. 2022;296:110911. doi 10.1016/j.scienta.2022.110911

Shitt P.G. The Study of the Growth and Development of Fruit and Berry Plants. Moscow, 1958 (in Russian)

Sotnik A.I., Babina R.D., Khoruzhy P.G. Comparative assessment of resistance of the generative organs of pear varieties (*Pirus communis* L.) to low-temperature stresses under the conditions of Crimea. *Izvestia Orenburg State Agrarian University*. 2017;3(65):72-74 (in Russian)

Sundyreva M.A., Stepanov I.V., Suprun I.I., Ushakova Y.V. A modified protocol of RNA isolation from mature leaves of grapes for RT-PCR. *Scientific Journal of Kuban State Agrarian University.* 2018;143:16-30. doi 10.21515/1990-4665-143-012 (in Russian)

Suzuki N., Koussevitzky S., Mittler R., Miller G. ROS and redox signalling in the response of plants to abiotic stress. *Plant Cell Environ*. 2012;35(2):259-270. doi 10.1111/j.1365-3040.2011.02336.x

Treml J., Mejkal K. Flavonoids as potent scavengers of hydroxyl radicals. *Compr Rev Food Sci Food Saf.* 2016;15(4):720-738. doi 10.1111/1541-4337.12204

Wei Z., Gao T., Liang B., Zhao Q., Ma F., Li C. Effects of exogenous melatonin on methyl viologen-mediated oxidative stress in apple leaf. *Int J Mol Sci*. 2018;19(1):316. doi 10.3390/ijms19010316

Xiao L.J., Asseng S., Wang X.T., Xia J.X., Zhang P., Liu L.L., Tang L., Cao W., Zhu Y., Liu B. Simulating the effects of low-temperature stress on wheat biomass growth and yield. *Agric For Meteorol.* 2022;326:109191. doi 10.1016/j.agrformet.2022.109191

Yang T., Huang X.-S. Deep sequencing-based characterization of transcriptome of *Pyrus ussuriensis* in response to cold stress. *Gene*. 2018;661:109-118. doi 10.1016/j.gene.2018.03.067

Zhai R., Liu J., Liu F., Zhao Y., Liu L., Fang C., Wang H., Li X., Wang Z., Ma F., Xu L. Melatonin limited ethylene production, softening and reduced physiology disorder in pear (*Pyrus communis* L.) fruit during senescence. *Postharvest Biol Technol*. 2018;139:38-46. doi 10.1016/j.postharvbio.2018.01.017

Zhang S. Recent advances of polyphenol oxidases in plants. *Molecules*. 2023;28(5):2158. doi 10.3390/molecules28052158

Zhang Y., Wu L., Liu L., Jia B., Ye Z., Tang X., Heng W., Liu L. Functional analysis of *PbbZIP11* transcription factor in response to cold stress in Arabidopsis and pear. *Plants*. 2023;13(1):24. doi 10.3390/plants13010024

Zhao Y., Hu M.Y., Wang Q., Yan X.K., Lu M.Y., Wu C.H., Li H.Y., Zhang M.J. Climate variation and its influence on the cold tolerance and phenology periods of pear cultivars in Jilin, China. *Int J Fruit Sci*. 2023;23(1):165-180. doi 10.1080/15538362.2023.2249995

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

Received July 4, 2025. Revised September 25, 2025. Accepted December 11, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):412-423

doi 10.18699/vjgb-26-45

Characteristics of galacturonate reductase (*GalUR*) genes in garlic (*Allium sativum* L.) and changes in their expression in response to abiotic stressors

M.A. Filyushin, T.M. Seredin, A.V. Shchennikova, E.Z. Kochieva

Skryabin Institute of Bioengineering, Federal Research Centre "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, Russia
michel7753@mail.ru

Abstract. In plants, the synthesis of L-ascorbic acid (Aa), in addition to the main L-galactose pathway, is carried out by three known alternative pathways. One of them, the D-galacturonic acid pathway, is thought to be specific for tissues with excess D-galacturonate, the substrate of D-galacturonate reductase (GalUR), which belongs to the Aldo-Keto Reductase (AKR) superfamily. In this study, the *AKR* gene family of garlic *Allium sativum* L. was identified and seven genes, *AsGalUR1–7*, presumably encoding GalUR enzymes, were determined. The structure and phylogeny of the *AsGalUR1–7* genes and the proteins they encode, as well as the *AsGalUR1–7* expression pattern in different organs of the garlic plant (*in silico* and qRT-PCR), were characterized. Based on the obtained data, the genes were conditionally divided into root (*AsGalUR1–4*) and leaf (*AsGalUR5–7*) groups depending on the highest expression level in the underground and aboveground parts of the plant, respectively. The *AsGalUR* expression in leaves and roots was analyzed in response to drought, salt and cold stresses, as well as exogenous phytohormones (abscisic acid, methyl jasmonate), accompanied by the AsA content measurement. It was shown that hormone treatment suppresses the expression of all analyzed genes in both organ types. Cold conditions stimulate the expression of root group genes and suppress that of leaf group genes in roots, and have the opposite effect in leaves. Osmotic stressors (NaCl, PEG) suppress the transcription of all genes in leaves, but do not change (NaCl) or stimulate (PEG) it in roots, which is accompanied by an increase in AsA accumulation in organs of both types. A positive correlation between the expression of the *AsGalUR1* and *4* genes and the AsA content is found in leaves under stress conditions. The data obtained can form the basis for further study of the mechanisms regulating AsA synthesis in garlic and other *Allium* species.

Key words: garlic; *Allium sativum* L.; L-ascorbic acid biosynthesis; D-galacturonic pathway; D-galacturonate reductase; GalUR

For citation: Filyushin M.A., Seredin T.M., Shchennikova A.V., Kochieva E.Z. Characteristics of galacturonate reductase (*GalUR*) genes in garlic (*Allium sativum* L.) and changes in their expression in response to abiotic stressors. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):412-423. doi 10.18699/vjgb-26-45

Funding. This study was supported by a grant from the Russian Science Foundation (No. 24-76-10005; gene characterization and analysis, expression analysis) and the Ministry of Science and Higher Education of the Russian Federation (gene expression analysis based on transcriptome data).

Характеристика генов галактуронатредуктаз (*GalUR*) чеснока (*Allium sativum* L.) и изменение их экспрессии в ответ на абиотические стрессоры

М.А. Филюшин, Т.М. Середин, А.В. Щенникова, Е.З. Кочиева

Институт биоинженерии им. К.Г. Скрябина, Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук, Москва, Россия
michel7753@mail.ru

Аннотация. Помимо основного L-галактозного пути синтеза L-аскорбиновой кислоты (AcK) в растениях, известно три альтернативных пути. Один из них, D-галактуроновый, считается специфичным для тканей с избытком D-галактуроната – субстрата D-галактуронатредуктазы (GalUR), относящейся к семейству альдегидкеторедуктаз (AKR). В настоящей работе идентифицировано семейство генов *AKR* чеснока *Allium sativum* L. и определено семь генов, *AsGalUR1–7*, предположительно кодирующих ферменты GalUR. Охарактеризованы структура и филогения генов *AsGalUR1–7* и кодируемых ими белков, а также профиль экспрессии генов *AsGalUR* в различных органах растения чеснока (*in silico* и ПЦР-РВ). На основе полученных данных гены условно разделены на корневую

(*AsGalUR1–4*) и листовую (*AsGalUR5–7*) группы по признаку наибольшей экспрессии в подземной и надземной частях растения соответственно. Проведен анализ экспрессии генов *AsGalUR* в листьях и корнях в ответ на воздействие засухи, солевого и холодового стрессоров, а также экзогенных фитогормонов (абсцизовая кислота, метилжасмонат), вкупе с измерением содержания АсК. Показано, что обработка гормонами подавляет экспрессию всех анализируемых генов в обоих типах органов. Холодовые условия в корнях стимулируют экспрессию корневых генов и подавляют – листовых и оказывают противоположный эффект в листьях. Осмотические стрессоры (NaCl, PEG) подавляют транскрипцию всех генов в листьях, но не меняют (NaCl) или стимулируют (PEG) ее в корнях, что сопровождается ростом накопления АсК. В листьях в стрессовых условиях выявлена положительная корреляция между экспрессией генов *AsGalUR1* и *4* и содержанием АсК. Полученные данные могут стать основой для дальнейшего изучения механизмов регуляции синтеза АсК у чеснока и других видов *Allium*.
Ключевые слова: чеснок; *Allium sativum* L.; биосинтез L-аскорбиновой кислоты; D-галактуроновый путь; D-галактуронатредуктаза; GalUR

Introduction

L-ascorbic acid (AsA, ascorbate) is a major component of the non-enzymatic antioxidant system in plants (Smirnoff and Wheeler, 2024). The required level of ascorbate in cells is maintained through a balance between the biosynthesis, degradation, and recycling (reduction of oxidized forms) of AsA, as well as ascorbate transport (Smirnoff, 2018). AsA is synthesized primarily in leaves (Franceschi, Tarlyn, 2002; Badejo et al., 2012) and transported via the phloem to various plant parts in the form of stable and oxidation-protected ascorbyl glycosides (Richardson et al., 2021; Huang et al., 2025).

In plants, AsA biosynthesis occurs primarily via the Smirnoff–Wheeler L-galactose pathway. However, there are three alternative pathways – D-galacturonic, L-gulose, and myo-inositol (Broad et al., 2020; Smirnoff, Wheeler, 2024), which also significantly contribute to ascorbate accumulation in some organs or at certain stages of plant development. For example, in strawberry (*Fragaria* × *ananassa* Duchesne ex Rozier), the L-galactose pathway is characteristic of leaves, while the D-galacturonic pathway is characteristic of ripe berries (Agius et al., 2003; Cruz-Rus et al., 2011; Liu et al., 2022). Similarly, in tomato (*Solanum lycopersicum* L.), leaves and growing fruits synthesize AsA via the L-galactose pathway, whereas ripe fruits, via the D-galacturonic pathway (Badejo et al., 2012). The transition from the main AsA biosynthetic pathway to an alternative variant in ripe fruits is associated with the active degradation of cell wall pectin, formed mainly by galacturonic acid residues, and the conversion of its main monomer, D-galacturonate, to AsA (Peltonen, Richard, 2022).

Of the alternative pathways, the D-galacturonic acid pathway is the best studied. It begins, as mentioned above, with the breakdown of pectin, resulting in the formation of D-galacturonic acid, which is converted to D-galacturonate (D-GalUA), followed by reduction to L-galactonate catalyzed by D-galacturonate reductase. L-galactonate is converted by aldonolactonase to L-galactono-1,4-lactone, which is oxidized to AsA by L-galactono-1,4-lactone dehydrogenase (Ishikawa et al., 2008; Peltonen, Richard, 2022).

Galacturonate reductases (EC 1.1.1.365; GalUR) structurally belong to the Aldo-Keto-Reductase (AKR) superfamily (Duan et al., 2020), one of 16 subfamilies of which, AKR4, includes AKR proteins of plant origin (Penning, 2015). The number of *AKR* genes identified in higher plants (Sengupta et al., 2015) varies significantly depending on the species. For example, the genome of *Arabidopsis thaliana* L. contains 22 *AKR* genes, that of *S. lycopersicum*, 25 genes, and that of *F.* × *ananassa*, 80 genes (Duan et al., 2020; Liu et al., 2022), and the functions of their protein products are not limited to participation in the synthesis of AsA. Plant AKRs also control the formation of other secondary metabolites and osmolytes, including vitamin B6, sorbitol, isoflavones, phytoestrogens, and others (Sengupta et al., 2015; Ha et al., 2019). Such functional differentiation of AKR proteins is considered to enhance plant adaptation to various environmental conditions, including stress factors. Regarding the AsA synthesis, it should be noted that the AKR family includes not only the GalUR enzymes of the D-galacturonic pathway, but also L-galactose dehydrogenase (L-GalDH) of the main L-galactose pathway (Vargas et al., 2022).

Garlic (*Allium sativum* L.) is an economically significant vegetable crop with an annual production of approximately 30 million tons (FAO; http://www.fao.org). The ascorbate content in leaves and bulbs has been extensively studied in various garlic varieties and accessions (Skoczylas et al., 2023; Šnirc et al., 2023; Popa et al., 2024; Yenealem et al., 2025). However, currently available information on the molecular regulation of AsA metabolism in *A. sativum* and equally economically valuable related *Allium* species is limited to research data from our laboratory. Namely, in *A. sativum*, individual genes of monodehydroascorbate reductases (MDHAR) of the ascorbate recycling pathway were identified and characterized (Anisimova et al., 2022), and changes in the expression of AsA biosynthesis and recycling genes in response to infection of garlic plants with the pathogenic fungus *Fusarium proliferatum* were shown (Shchennikova et al., 2025). In leek (*Allium porrum* L.), the variability of the GDP-L-galactose phosphorylase gene *GGP1* of the L-galactose pathway was determined (Anisimova et al., 2021a), and possible dependences of the AsA content on the expression level of the *MDHAR* genes (Anisimova et al., 2021b; Filyushin et al., 2021) and some other genes of the L-galactose and AsA recycling pathways (Filyushin et al., 2025) were identified.

The present study was aimed to identify and characterize D-galacturonate reductase (*GalUR*) genes in the genome of garlic *A. sativum* and to determine their expression pattern in various plant organs, as well as in response to abiotic stress factors and exogenous phytohormones.

Materials and methods

Identification and structural characterization of garlic *AsGalUR* genes. The D-galacturonate reductase gene sequences were searched for in the genomic and transcriptomic data of garlic *A. sativum* cv. Ershuizao (PRJNA606385, assembly Garlic.V2.fa) available in the AlliumDB database (https://allium.qau.edu.cn/). The *GalUR* sequences of tomato (LOC101256763 and LOC101250974 (=SlAKR4B)) (Suekawa et al., 2016) and strawberry (*FaGalUR*; AF039182.1) (Agius et al., 2003) were used as reference.

Sequence alignment and analysis were performed using MEGA 7.0 (https://www.megasoftware.net/). For the AsGalUR proteins, conserved domains and motifs (NCBI-CDD, http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; MEME 5.5.7, http://meme-suite.org/tools/meme), molecular weight (Mw) and isoelectric point (pI) (ExPASy, https://web.expasy.org/protparam/), molecular function and cellular localization (in terms of Gene Ontology, GO, PANNZER, http://ekhidna2.biocenter.helsinki.fi/sanspanz/) were determined. Phylogenetic analysis was performed using AKR sequences of *A. sativum*, *S. lycopersicum* and *A. thaliana* (MEGA 7.0, Neighbor-Joining method, bootstrap based on 1,000 replicates).

Analysis of *AsGalUR* gene expression in different garlic organs. The organ-specific expression pattern of the identified *AsGalUR* genes was determined using two methods: *in silico* and real-time PCR (qRT-PCR).

In silico analysis of *AsGalUR* gene expression was performed using available transcriptome data of different organs of the garlic cv. Ershuizao (roots, leaves, stems, seedlings, buds, flowers, and bulbs at developmental stages 1–8) (Sun et al., 2020). The data were visualized as a heatmap (Heatmapper, http://www.heatmapper.ca/expression/), where expression levels were presented as fragments per kilobase per million mapped reads (FPKM).

Using qRT-PCR, the expression of *AsGalUR* genes was determined in the roots, base, bulb, pseudostem, and leaves of garlic cv. Scorpion plants grown in soil under greenhouse conditions (ECCF, Federal Research Center of Biotechnology, Russian Academy of Sciences; 16-h photoperiod (light phase from 8:00); day/night – 22/16 °C; lighting intensity 190 μM/m^2/s). The plant material was ground in liquid nitrogen and used for the isolation of total RNA with purification from DNA impurities (RNeasy Plant Mini Kit and RNase-free DNasy set; QIAGEN, Germany) and cDNA synthesis (GoScript™ Reverse Transcription System, Promega, USA).

Based on the identified *AsGalUR* sequences, specific primers for qRT-PCR were developed (Table 1). *GAPDH* (MZ171220.1) and *UBQ* (MZ171222.1) were used as reference genes. The reaction mixture included 3 ng of cDNA and the "Reaction mixture for qRT-PCR in the presence of SYBR Green I and ROX" kit (Synthol LLC, Russia). The reaction was carried out in a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, USA), in six technical replicates of two biological replicates under the following conditions: 95 °C for 5 min; 40 cycles (95 °C for 15 s, 62 °C for 50 s). Data were statistically processed using one-way ANOVA with Bonferroni correction ("multiple comparisons, corrected with Bonferroni test") and visualized in GraphPad Prism v. 8 (https://www.graphpad.com).

Table 1. Primer sequences for qRT-PCR

Gene	Primer sequences (5′→3′)
AsGalUR1	TGCCATCCGAAGGAGTCTTGT AGCATTTGGCACTCCTCCATC
AsGalUR2	TGCCATCCGAAGGAGTCTTGT AGCATTTGGCACTCCTCCATG
AsGalUR3	GAGGAGTGCCAAATGCTTGGAT CGTCCACCCAAAGGAGAGTG
AsGalUR4	GTGTAACGATGCCCATCCTGAT CTACTGGAAATGGGTGTGATCC
AsGalUR5	CTGCTGTTAATCAGGTGGAGT ACACTTCCCCACGGAGCTC
AsGalUR6	CTGCTGTTAATCAGGTGGAGC ACACTTCCCCACGGAGCTC
AsGalUR7	TCGGTGTGAGCAATTTCTCATC ACACTTCCCCACGGAGCTC
GAPDH	CCATGTTTGTTGTTGGTGTGAATGAG TGGTGCAGCTAGCGTTGGAGAC
UBQ	AAGCCAAGATACAGGACAAG GCATACCACCTCTCAATCTC

Analysis of the *AsGalUR* gene expression dynamics in garlic seedlings exposed to various stressors (drought, salinity, cold) and exogenous phytohormones. Garlic cv. Scorpion plants were grown in transparent glass beakers in water until 15-day-old seedlings were obtained; cloves were fixed in a porous polyethylene substrate so that only the lower part of the clove (the root zone) was submerged in water. At 9:00 and 15:00 of the light growth phase, leaf and root samples were collected (stored at –80 °C) for subsequent analysis (qRT-PCR) of *AsGalUR* gene expression.

To imitate stressful conditions, experimental plants were transferred to the corresponding aqueous solutions for 24 h (100 mM NaCl for salinity; 10 % PEG-6000 for drought; 100 μM abscisic acid (ABA) and 100 μM methyl jasmonate (MeJA) for exogenous phytohotmones). Control plants remained in water. For cold exposure, experimental plants were placed in a climate chamber (+4 °C, without light), while the control plants were maintained in the dark at 22 °C. After 6 h and 24 h of exposure to the stressor/hormone, roots and leaves were collected from plants in the experimental and control groups and stored at –80 °C.

The collected plant material was ground in liquid nitrogen and used to obtain RNA/cDNA preparations and perform qRT-PCR as described above.

Determination of ascorbate content in plant tissues. The AsA content (mg/g fresh weight) was measured in the roots and leaves of garlic plants subjected to stress and treatment with phytohormones. Analysis was performed using the L-Ascorbic acid kit (R-Biopharm AG, Germany), and absorption spectra

were recorded on an Eppendorf BioSpectrometer® basic spectrophotometer (Eppendorf, Germany). Regression analysis of the data (search for correlations between the expression level of *AsGalUR* genes and ascorbate content) was performed using GraphPad Prism v. 8 (https://www.graphpad.com).

Results

Identification and structural characterization of garlic D-galacturonate reductase genes

Twenty-seven genes encoding AKR family proteins were identified in the *A. sativum* genome, seven of which were homologous to *GalUR* in *S. lycopersicum* (LOC101256763 and LOC101250974) and *F.* × *ananassa* (AF039182.1) and annotated in AlliumDB as "KEGG pathway: D-galacturonate reductase". These seven garlic genes were numbered *AsGalUR1–7* in the order of their localization on chromosomes 5, 6, and 7. The genes differed in the number of exons – 3 (*AsGalUR5–7*) or 4 (*AsGalUR1–4*), and in size – from 1,105 bp (*AsGalUR5*) to 4,085 bp (*AsGalUR7*). The genes were similar in terms of the length of the coding sequence (CDS) and the predicted physicochemical properties of the encoded proteins (Table 2).

To construct a phylogenetic dendrogram, the amino acid sequences of all 27 identified AKRs from *A. sativum* were compared with AKR homologs from *S. lycopersicum* and *A. thaliana*. In the resulting dendrogram, the proteins were divided into two subfamilies, AKR4A and AKR4B (Fig. 1).

Table 2. Characteristics of *AsGalUR* genes

Gene	Gene ID / Transcript ID	Genome localization	Gene, bp	Exon number	CDS, bp	Protein, aa	Mw, kDa	pI
AsGalUR1	Asa5G02985/ Asa3G05397.1	chr5: 764719171..764720824	1,654	4	945	314	35.0	5.73
AsGalUR2	Asa5G05896/ Asa5G00222.1	chr5: 1607364860..1607366708	1,849	4	969	322	36.2	5.48
AsGalUR3	Asa6G06508/ Asa1G06086.1	chr6: 1808233236..1808234881	1,646	4	945	314	35.1	5.73
AsGalUR4	Asa6G06509/ Asa1G06085.1	chr6: 1808295357..1808297099	1,743	4	951	316	35.5	5.21
AsGalUR5	Asa7G05242.1/ Asa5G04903.1	chr7: 1457744393..1457745497	1,105	3	963	320	36.1	6.39
AsGalUR6	Asa7G05243.1/ Asa5G04902.1	chr7: 1457746950..1457748486	1,537	3	975	324	36.7	5.96
AsGalUR7	Asa7G05244.1/ Asa4G05086.1	chr7: 1458214446..1458218530	4,085	3	963	320	36.1	6.18

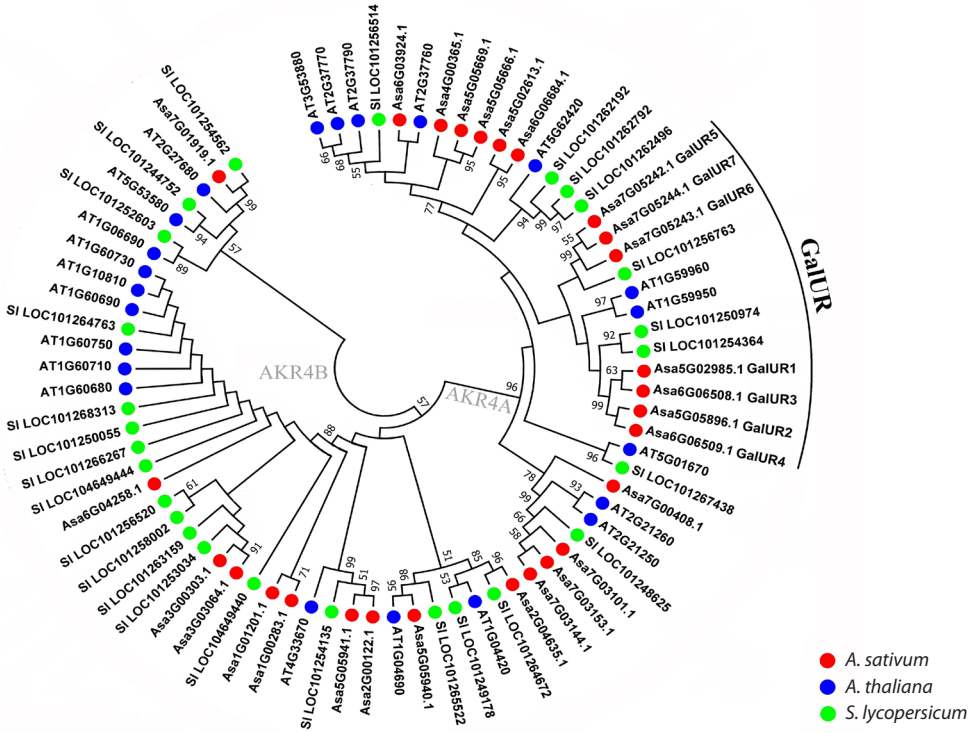


Fig. 1. Dendrogram constructed based on the amino acid sequences of AKR family proteins from garlic (*A. sativum*), Thale cress (*A. thaliana*), and tomato (*S. lycopersicum*) (MEGA 7.0, Neighbor-Joining method, 1,000 bootstrap replicates, significant bootstrap values (%) are indicated at the base of the branches).

The AsGalUR1–7 sequences formed a separate clade within AKR4A, dividing into subclades I (AsGalUR1–4) and II (AsGalUR5–7). The first (I) turned out to be homologs of D-galacturonate reductases from *A. thaliana* (AT1G59950 and AT1G59960 (Duan et al., 2020)) and *S. lycopersicum* SlAKR4B (LOC101250974 (Suekawa et al., 2016)), as well as of an uncharacterized *S. lycopersicum* AKR (LOC101254364). The second (II) clustered with known *S. lycopersicum* GalUR (LOC101256763 (Suekawa et al., 2016)) (Fig. 1). The homology of AsGalUR sequences was 86–99 % (I) and 89–95 % (II) within subclades, and 48–52 % between subclades I and II.

All the AsGalUR1–7 proteins were found to contain the conserved AKR domain PF00248.18 (the domain position was determined by comparison with homologs from *S. lycopersicum*, *Vitis vinifera* L., *F.* × *ananassa*, and *A. thaliana*) and, according to (Suekawa et al. 2016), 10 functionally important amino acid residues (aa), including the plant AKR cofactor binding sites (Fig. 2).

In the same proteins, conserved motifs were also identified (Fig. 3). The composition and position of the motifs were similar within the group of analyzed proteins, with the exception of slightly different C-terminal consensuses, which is associated with variability in the extra-domain region. Namely, individual proteins lacked motif 7 (AT1G59950 and AT1G59960) or 8 (Sl_LOC101256763, FaGalUR, and VvGalUR), or both motifs (AsGalUR2).

Analysis of AsGalUR1–7 sequences in PANNZER predicted that all seven proteins possess oxidoreductase activity (GO:0016616) and are localized in the cytosol (GO:0005829).

Determination of the *AsGalUR1–7* gene expression pattern in garlic plants

Using *A. sativum* cv. Ershuizao transcriptome data (Sun et al., 2020), the expression profiles of the *AsGalUR1–7* genes were determined in various garlic organs (including stages 1–8 of bulb growth and development) (Fig. 4). It was found that the *AsGalUR1*, *3*, and *4* genes (subclade I) were expressed in all analyzed organs, with the highest levels in roots (all three genes), leaves (*AsGalUR3* and *4*), stems (*AsGalUR1* and *3*), and buds (*AsGalUR3*). Transcripts of the fourth gene of subclade I, *AsGalUR2*, were not detected in any of the organs analyzed.

Transcripts of the *AsGalUR5–7* genes (subclade II) were present predominantly in aboveground organs (except flowers), with a maximum in seedlings (all three genes). *AsGalUR5* and *7* (but not *AsGalUR6*) were also highly expressed in leaves,

Fig. 2. Alignment of the amino acid sequences of D-galacturonate reductases from *A. sativum* (As), *F.* × *ananassa* (Fa), *V. vinifera* (Vv), *A. thaliana* (AT), and *S. lycopersicum* (Sl).

The background color corresponds to the level of amino acid conservation in the analyzed proteins (green – 100 %, blue – 80 %, pink – 60 %). The position of the AKR domain PF00248.18 is indicated by red underlining. Blue asterisks indicate active sites and cofactor-binding sites of plant AKRs (according to (Suekawa et al., 2016)).

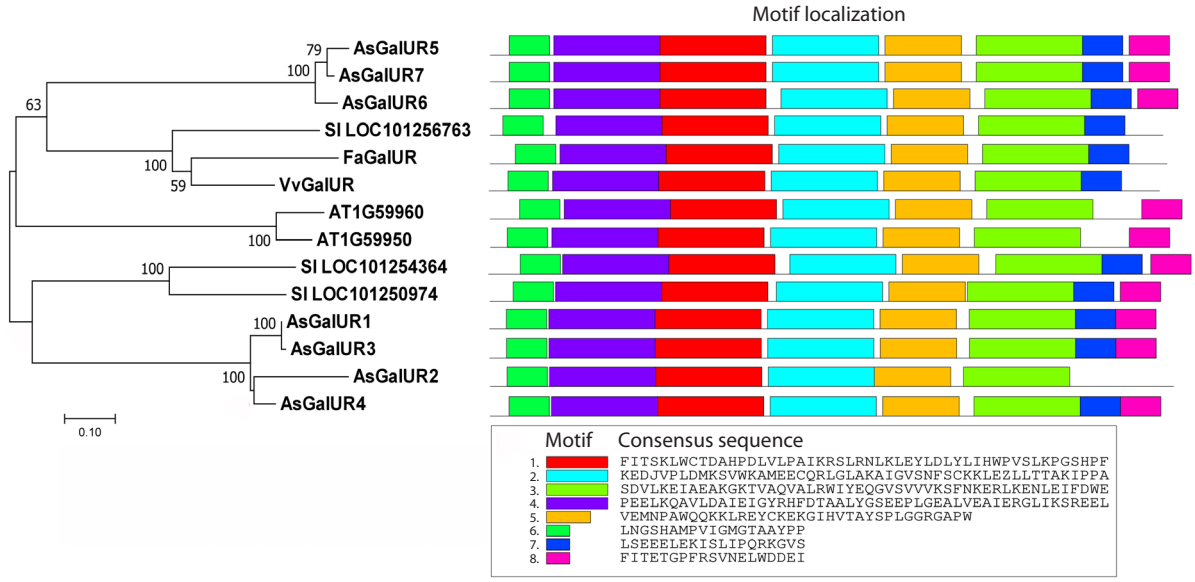


Fig. 3. Phylogenetic relationships and comparative profiling of conserved motifs of D-galacturonate reductases from *A. sativum* (As), *F.* × *ananassa* (Fa), *V. vinifera* (Vv), *A. thaliana* (AT) и *S. lycopersicum* (Sl).

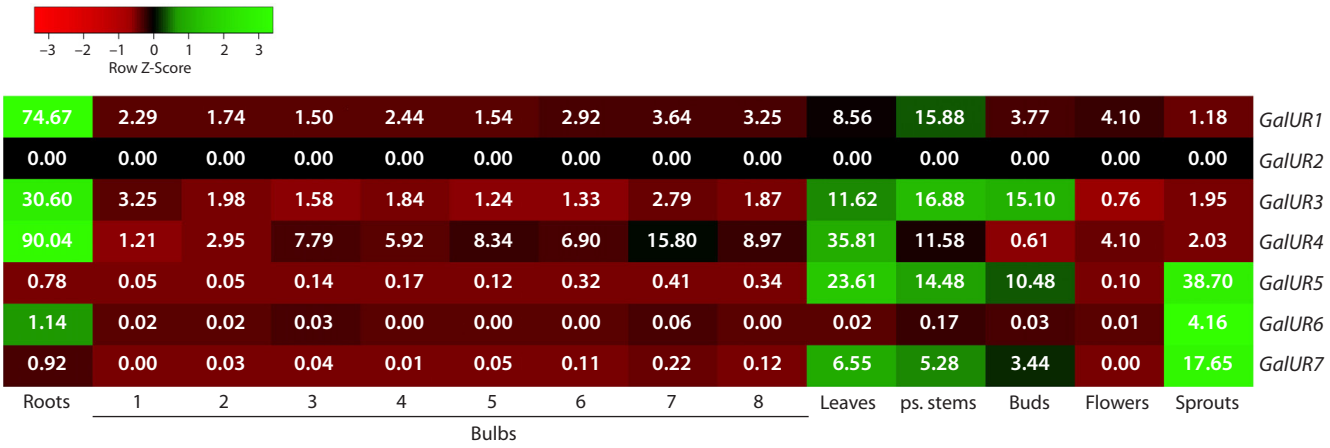


Fig. 4. Heatmap of *AsGalUR1–7* expression in different organs of garlic *A. sativum* cv. Ershuizao based on transcriptome data (Sun et al., 2020).

Numbers in boxes indicate the average FPKM values for three biological replicates. For bulbs, stages 1–8 (corresponding to 192, 197, 202, 207, 212, 217, 222, and 227 days of development) are shown.

stems, and buds. In bulbs and roots, the expression level of the *AsGalUR5–7* genes was low relative to the aboveground parts of the plant (Fig. 4).

Overall, based on the preferred expression profile (Fig. 4), the genes were conditionally assigned to the root (*AsGalUR1*, *3*, and *4*, despite a significant number of transcripts also being expressed in leaves/stems) and leaf (*AsGalUR5–7*) groups, which corresponded to their distribution among phylogenetic subclades I and II (Fig. 1). The exception was the *AsGalUR2* gene of subclade I (Fig. 1), for which no expression was detected in garlic organs (Fig. 4).

In the organs of garlic cv. Scorpion plants, *AsGalUR1–7* gene expression was determined using qRT-PCR (Fig. 5). In contrast to the transcriptome data (Fig. 4), transcripts of the *AsGalUR2* gene were detected everywhere, while *AsGalUR6* was not expressed. Subclade I genes (*AsGalUR1–4*) were expressed in all analyzed organs, with a maximum in the roots and clove base. Transcripts of subclade II genes (*AsGalUR5* and *7*) were detected only in the leaves (maximum) and pseudostem (Fig. 5).

Thus, the results of qRT-PCR (Fig. 5) made adjustments to the *in silico* expression profile of the *AsGalUR2* and *6* genes (Fig. 4) and confirmed the possibility of dividing the genes of subclades I and II into genes with predominant expression in the underground (I) and aboveground (II) parts of the plant.

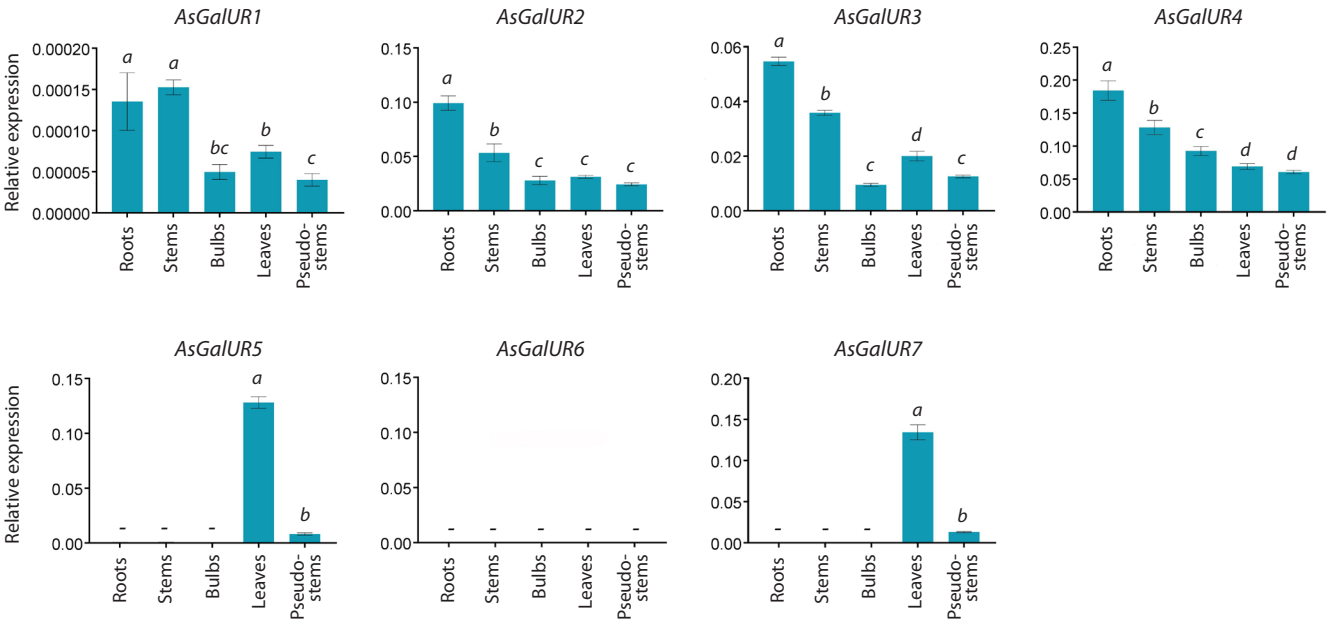


Fig. 5. Expression profile (qRT-PCR) of the *AsGalUR1–7* genes in cv. Scorpion garlic plants.
Significant differences in gene expression levels between different organs at $^{a-d}p < 0.05$.

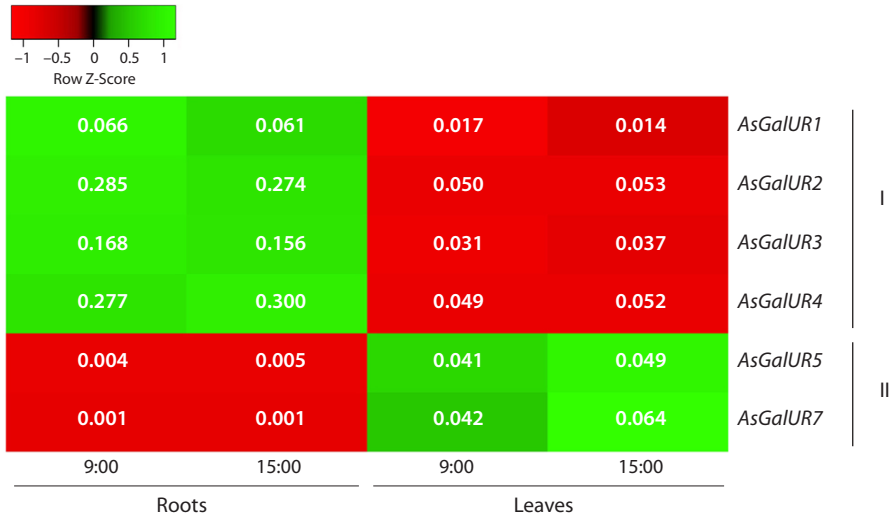


Fig. 6. Heatmap of the expression (based on qRT-PCR results) of the *AsGalUR1–5* and *7* genes in the roots and leaves of 15-day-old garlic cv. Scorpion seedlings at two time points of the plant growth light phase – 9:00 and 15:00.
The numbers in boxes indicate the average values of the *AsGalUR* gene expression levels, normalized to the expression of the reference genes *GAPDH* and *UBQ*.

Determination of the expression pattern of the *AsGalUR1–7* genes in garlic seedlings: 6-h dynamics during the light phase of plant growth and response to stress factors and exogenous phytohormones

The *AsGalUR1-5* and *7* gene expression levels were determined in roots and leaves of 15-day-old garlic cv. Scorpion seedlings at two time points (9:00 and 15:00) of the light growth phase (Fig. 6). It was confirmed that the genes of root subclade I (*AsGalUR1-4*) are most active in roots, while the genes of leaf subclade II (*AsGalUR5* and *7*) are most active in leaves. Moreover, over the 6-h light phase of plant growth, the gene expression level changed insignificantly, maintaining organ specificity. Interestingly, the transcripts of the "leaf" genes *AsGalUR5* and *7*, which in adult plants were present only in aboveground organs (Fig. 5), were also detected in the roots of seedlings, although with a significantly lower (20–25 times) expression level than in the leaves (Fig. 6).

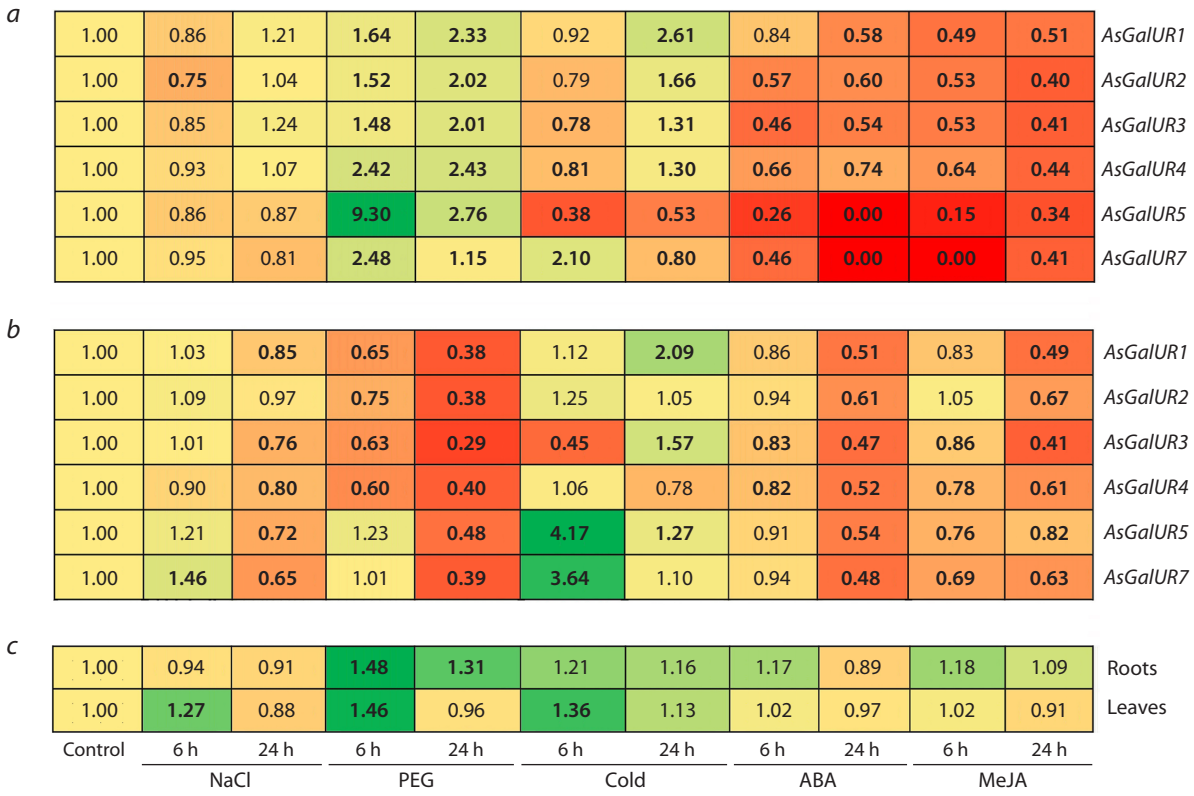


Fig. 7. Dynamics of *AsGalUR* gene expression in roots (*a*) and leaves (*b*) of garlic seedlings in response to abiotic stressors (salinity, 100 mM NaCl; drought, 10 % PEG-6000; cold stress, +4 °C) and exogenous phytohormones (100 μM ABA; 100 μM MeJa). Changes in the AsA content (*c*) in the roots and leaves of the same garlic seedlings in response to abiotic stressors and phytohormones.

Values significantly ($p < 0.05$) different from the control are shown in bold. Gradient coloring of cells corresponds to changes in gene expression level towards a decrease (shades of red) or an increase (shades of green) in relation to the control (yellow).

To determine the involvement of D-galacturonate reductases in the stress response of garlic plants, 15-day-old cv. Scorpion seedlings were exposed to various stressors (salinity, drought, cold, and the exogenous phytohormones ABA and MeJA). After 6 h and 24 h of exposure, gene expression of *AsGalUR1–5* and *7* was analyzed in the roots and leaves (Fig. 7).

In roots, it was shown that excess NaCl either did not change (*AsGalUR1*, *3–5*, and *7*) or suppressed (*AsGalUR2*, 6 h under stress conditions) gene expression. Dehydration (PEG) significantly stimulated gene expression (by 1.5–9.3 times) at both time points, with the exception of *AsGalUR7*, the transcript levels of which approached control values after 24 h. Six hours of cold exposure affected individual genes, decreasing (*AsGalUR3*, *4*, and *5*) or increasing (*AsGalUR7*) their expression. However, after 24 h, the expression of all genes changed in a subclade-specific manner: it increased (*AsGalUR1–4*, subclade I) or decreased (*AsGalUR5* and *7*, subclade II) compared to the control level. Phytohormone treatment significantly reduced the expression of all genes, reaching zero for *AsGalUR5* and *7* (Fig. 7*a*).

In leaves, 6 h of exposure to salt stress stimulated the expression of *AsGalUR7* (by 1.5 times), while 24 h of exposure suppressed the activity of all genes (except *AsGalUR2*). Six hours of drought did not change (*AsGalUR5* and *7*) or led to a decrease (*AsGalUR1–4*) in gene expression; after 24 h, the transcript levels of *AsGalUR1–5* and *7* decreased by 2.5–3.4 times. In response to 6 h of cold conditions, gene expression increased (*AsGalUR5* and *7*), decreased (*AsGalUR3*), or did not change (*AsGalUR1*, *2*, and *4*); 24 h of exposure led to an increase in the expression of *AsGalUR1*, *3*, and *5*. Treatment with ABA and MeJA, as in roots, negatively affected the expression of the *AsGalUR1–5* and *7* genes (Fig. 7*b*).

In garlic seedlings exposed to stress factors and exogenous phytohormones, the ascorbate content was determined to be 1.4–2.2 and 12.1–24.3 mg/100 g in roots and leaves, respectively (Fig. 7*c*). In roots, AsA content increased at both time points (6 h and 24 h of stress) in response to drought (PEG), whereas it did not change under the influence of other stressors and phytohormones. In leaves, increased ascorbate accumulation was observed after 6 h of exposure to NaCl, drought, and cold stress; at 24 h, the AsA content returned to control values (Fig. 7*c*).

In the leaves of garlic seedlings, the correlation analysis revealed a significant ($p < 0.05$) dependence of the ascorbate content on the expression level of the *AsGalUR1* (correlation coefficient $r = 0.63$) and *AsGalUR4* ($r = 0.66$) genes under stress conditions.

Discussion

The D-galacturonic pathway for L-ascorbic acid biosynthesis in plants was first discovered in the 1950s, and the enzyme D-galacturonate reductase was discovered at the same time. It catalyzes the reduction of D-galacturonic acid methyl ester to L-galactonic acid (Isherwood, Mapson, 1956), which is converted in several stages to AsA (Ishikawa et al., 2008; Peltonen, Richard, 2022). In 2003, the first gene of the *GalUR* family encoding D-galacturonate reductase was cloned in strawberry *F.* × *ananassa* (*FaGalUR*), in which ripe berries ascorbate is synthesized predominantly via the D-galacturonic pathway (Agius et al., 2003; Cruz-Rus et al., 2011).

In this study, the *AKR4* gene family (Fig. 1) was identified for the first time in the genome of garlic *A. sativum*. Seventeen of the 27 identified *AsAKR* genes belonged to the *AKR4A* subfamily, which was consistent with the higher number of *AKR4A* genes compared to *AKR4B* in higher plants (Duan et al., 2020). Seven *AKR4A* genes, *AsGalUR1–7* (Fig. 1, Table 2), encoded homologs of known GalUR proteins from *S. lycopersicum* (Suekawa et al., 2016), *A. thaliana* (Duan et al., 2020), and *F.* × *ananassa* (Agius et al., 2003).

Based on structural homology, it was suggested that AsGalUR1–7 proteins may function as D-galacturonate reductases in garlic plants. The presence of functionally important sites and a conserved domain specific for AKR in all AsGalUR1-7 proteins (Fig. 2), along with the results of GO analysis, also indicates the presence of oxidoreductase activity characteristic of AKR in AsGalUR1–7 when localized in the cell cytosol. The predicted localization of AsGalUR1–7 was consistent with the fact that the two-step conversion of D-galacturonate to L-galactono-1,4-lactone involving GalUR occurs precisely in the cytosol (Smirnoff et al., 2001).

Based on the number of exons, the genes were divided into two groups: *AsGalUR1–4* (4 exons) and *AsGalUR5–7* (3 exons) (Table 2), which suggested the existence of functional differences between them, possibly associated with the specificity of expression for individual organs/tissues/developmental stages and/or with the adaptive reactions of the garlic plant. This division was confirmed by phylogenetic analysis. In the dendrogram, the AsGalUR1–7 proteins formed two subclades: I (AsGalUR1-4) and II (AsGalUR5-7) (Fig. 1), although the set of conserved motifs looked identical, with the exception of the C-terminal part of AsGalUR2 (Fig. 3). The variability of the AsGalUR2 C-terminal sequence may indicate the presence of individual functional properties in this protein, but does not exclude its involvement in the synthesis of AsA. The different numbers of subclade I and II *GalUR* genes in *A. sativum* and, for example, *S. lycopersicum*, *A. thaliana* or *Brassica rapa* L. (Duan et al., 2020) (Fig. 1) suggests the emergence of precursors of these genes before the separation of the monocot and dicot classes and subsequent species-specific duplication evolution of the corresponding gene families.

As mentioned above, activation of D-galacturonic acid pathway of AsA synthesis as an alternative to the main L-galactose pathway occurs in ripe, juicy fruits, which is associated with the degradation of cell wall pectin in the pulp with the release of D-galacturonate (a substrate for GalUR) (Cruz-Rus et al., 2011; Badejo et al., 2012). Growing and ripening fruits obtain the required amount of ascorbate through internal synthesis via the main pathway and translocation of AsA from the leaves (Badejo et al., 2012). Softening of ripe fruits is accompanied by the destruction of cell walls. During this, pectin, which makes up ~35 % of the primary walls in dicotyledons and non-cereal monocots (including garlic) (Mohnen, 2008) and consists of ~70 % galacturonic acid residues (Mølhøj et al., 2004; Cruz-Rus et al., 2011; Badejo et al., 2012), is degraded. There is a significant increase in the amount of substrate for GalUR, which stimulates a switch to an alternative pathway for AsA synthesis in fruits (Peltonen, Richard, 2022).

Thus, to activate the D-galacturonic acid pathway, plants require D-galacturonate as a substrate for GalUR (Peltonen, Richard, 2022). However, in vegetative tissues, as well as in growing storage organs, this compound is strictly necessary for the synthesis of cell wall pectin, since plant growth and development are accompanied by active cell division (Mohnen, 2008). Nevertheless, the constant presence of high concentrations of D-galacturonate in the cytosol of vegetative tissue cells suggests that D-galacturonic acid pathway of AsA synthesis may also occur there as a minor addition to the main L-galactose pathway. This is indirectly confirmed by the positive correlation between the content of AsA and the expression level of D-galacturonate reductase genes found in the leaves of *B. rapa* (Duan et al., 2020) and tea bush (*Camellia sinensis* (L.) O. Kuntze) (Li et al., 2017).

Based on the above, in this study, we analyzed the expression of the *AsGalUR1–7* genes in various organs of the *A. sativum* cv. Ershuizao plant (*in silico*, using transcriptome data (Sun et al., 2020)) (Fig. 4), as well as in the leaves and roots (qRT-PCR) of the garlic cv. Scorpion, including in response to abiotic stressors and exogenous phytohormones (Fig. 5–7). In addition to vegetative organs, bulb samples (as a storage organ) at several stages of growth and maturation were included in the *in silico* analysis (Sun et al., 2020).

The analysis confirmed that the groups of *AsGalUR* genes belonging to subclades I and II differ in their expression profiles and, presumably, in their functional specialization. So, the transcript levels of genes *AsGalUR1–4* (I) were significantly higher in the underground part of the garlic plant compared to the aboveground organs, while the opposite pattern was observed for genes *AsGalUR5–7* (II) (Figs. 4–6). This suggests organ-specific involvement of these genes in the D-galacturonic pathway; subclade I and II genes were provisionally named as root (*AsGalUR1–4*) and leaf (*AsGalUR5–7*) genes. A question remains regarding the *AsGalUR2* and *6* genes, which showed different expression patterns depending on the type of analysis, but based on the totality of the data (Fig. 4 and 5), these genes correspond to the characteristics of their groups, although they require further study. In our further work, we relied on the results of qRT-PCR, that is, we excluded *AsGalUR6* from the study.

As expected, in the bulb, the root genes *AsGalUR1–4* were expressed significantly higher than the leaf genes; however,

no significant trends in transcript level changes were observed as the bulbs approached maturity (Fig. 4). This may be due to differences in the definition of full ripeness between juicy fruits and root vegetables. While in fruits biological ripeness is accompanied by softening of the pulp (Badejo et al., 2012), in root vegetables it means enlargement and vacuolization of cells for nutrient accumulation and the onset of physiological dormancy (Teper-Bamnolker et al., 2012). Softening of root tissue may be a sign of dormancy release, wilting, or rotting.

Consistent with the fact that 12 *GalUR* genes of *B. rapa* exhibit obvious divergence of expression under different stress conditions (Duan et al., 2020), in our study, the *AsGalUR1–5* and *7* genes also responded diversely and organ-dependently to both abiotic stressors and exogenous ABA and MeJA (Fig. 7*a*, *b*). Despite the conditional division of genes into root and leaf groups, in most cases their expression in response to stress changed in both roots and leaves (Fig. 7*a*, *b*). This indicates the involvement of all *AsGalUR1–5* and *7* genes in AsA synthesis via an alternative pathway under stress conditions throughout the plant, but with predominant specificity in the underground or aboveground parts.

Exogenous phytohormones suppressed the expression of genes of both group in both tissue types (Fig. 7*a*, *b*), which is consistent with the role of ABA and MeJA as plant growth regulators and cell division stimulators (Fattorini et al., 2009; Xie et al., 2020). It is possible that hormone treatment accelerates cell division, which requires increased pectin synthesis, reduces the amount of substrate for GalUR and, as a result, downregulates the expression of *AsGalUR* genes. On the other hand, exogenous ABA has been shown to stimulate AsA synthesis (Xu et al., 2022); however, this most likely relates to the main ascorbate synthesis pathway. In addition, we did not observe an increase in ascorbate content in response to phytohormones (Fig. 7*c*), which may suggest the absence of an effect or a delayed stimulatory effect of ABA on AsA synthesis.

The effects of abiotic stressors on gene expression were more diverse than those of hormones (Fig. 7*a*, *b*). The main difference between the reaction of root (*AsGalUR1–4*) and leaf (*AsGalUR5*, *7*) genes was the opposite expression dynamics under low temperature exposure. In roots (as expected from conditional gene specialization), after 24 h of treatment, the expression of root genes increased, while that of leaf genes decreased (Fig. 7*a*). In leaves, on the contrary, after 6 h of exposure, the expression of leaf *AsGalUR5* and *7* sharply increased, while the expression of root *AsGalUR1–4* decreased or remained unchanged (Fig. 7*b*). This is consistent with the positive dependence of plant cold resistance on AsA content (Fu et al., 2023) and emphasizes the organ specificity of the increase in *AsGalUR* gene expression in response to cold, which led to a significant (leaves) or weaker (roots) increase in ascorbate content (Fig. 7*c*).

Ascorbic acid promotes plant tolerance to salinity and drought (Younis et al., 2010). Accordingly, the expression of *AsGalUR* genes changed in response to osmotic/ionic stress (PEG, NaCl) with a more pronounced effect in the roots, which is associated with the specificity of stress conditions created (roots immersed in a PEG or NaCl solution). However, no functional differences were found between leaf and root *AsGalUR* genes: in leaves, the expression level of all genes decreased, while in roots it did not change (NaCl) or increased (PEG) (Fig. 7*a*, *b*), which was accompanied by an increase in AsA accumulation in organs of both types (Fig. 7*c*). It should be noted that the correlation between the gene expression level and the ascorbate content under stress conditions was found only in leaves and only for the *AsGalUR1* and *4* genes assigned to the root group, which confirms the conventionality of the functional division of *AsGalUR* genes in the adaptive responses of garlic plants.

The data we obtained are fully consistent with the activation of alternative pathways for ascorbate synthesis in plants under stressful conditions shown in other studies (Xu et al., 2012; Ruggieri et al., 2016). In particular, this may be due to a decrease in the flux of the main L-galactose pathway in response to stressors (due to the active involvement of this pathway precursor, D-glucose, in plant stress responses), which was demonstrated in the analysis of the *vtc1* and *vtc2* mutants of *A. thaliana* (Quiñones et al., 2024).

Conclusion

In this study, we identified the *A. sativum AKR* gene family and defined seven *AsGalUR1–7* genes, which presumably encode D-galacturonate reductases – key enzymes in the alternative D-galacturonic acid pathway of L-ascorbic acid biosynthesis. We characterized the structure and phylogeny of the genes and encoded proteins, as well as the organ-specific expression profile of the *AsGalUR1–7* genes. Based on this, the genes were conditionally divided into root (*AsGalUR1–4*) and leaf (*AsGalUR5–7*) groups. We analyzed gene expression in response to abiotic stress factors (salinity, drought, and cold) and exogenous phytohormones (ABA, MeJA), which revealed additional functional features of the genes in determining garlic plant stress tolerance.

References

Agius F., González-Lamothe R., Caballero J.L., Muñoz-Blanco J., Botella M.A., Valpuesta V. Engineering increased vitamin C levels in plants by overexpression of a D-galacturonic acid reductase. *Nat Biotechnol*. 2003;21(2):177-181. doi 10.1038/nbt777

Anisimova O.K., Shchennikova A.V., Kochieva E.Z., Filyushin M.A. Identification and variability of the GDP-L-galactose phosphosphorylase gene *ApGGP1* in leek cultivars. *Russ J Genet*. 2021a;57(3):311-318. doi 10.1134/S1022795421030030

Anisimova O.K., Seredin T.M., Shchennikova A.V., Kochieva E.Z., Filyushin M.A. Estimation of the vitamin C content and GDP-L-galactose phosphorylase gene (*VTC2*) expression level in leek (*Allium porrum* L.) cultivars. *Russ J Plant Physiol*. 2021b;68(1):85-93. doi 10.1134/S1021443720060023

Anisimova O.K., Shchennikova A.V., Kochieva E.Z., Filyushin M.A. Identification of monodehydroascorbate reductase (*MDHAR*) genes in garlic (*Allium sativum* L.) and their role in the response to *Fusarium proliferatum* infection. *Russ J Genet*. 2022;58(7):773-782. doi 10.1134/S1022795422070031

Badejo A.A., Wada K., Gao Y., Maruta T., Sawa Y., Shigeoka S., Ishikawa T. Translocation and the alternative D-galacturonate pathway contribute to increasing the ascorbate level in ripening tomato fruits together with the D-mannose/L-galactose pathway. *J Exp Bot*. 2012;63(1):229-239. doi 10.1093/jxb/err275

Broad R.C., Bonneau J.P., Hellens R.P., Johnson A.A.T. Manipulation of ascorbate biosynthetic, recycling, and regulatory pathways for improved abiotic stress tolerance in plants. *Int J Mol Sci.* 2020; 21(5):1790. doi 10.3390/ijms21051790

Cruz-Rus E., Amaya I., Sánchez-Sevilla J.F., Botella M.A., Valpuesta V. Regulation of L-ascorbic acid content in strawberry fruits. *J Exp Bot*. 2011;62(12):4191-4201. doi 10.1093/jxb/err122

Duan W., Huang Z., Li Y., Song X., Sun X., Jin C., Wang Y., Wang J. Molecular evolutionary and expression pattern analysis of *AKR* genes shed new light on GalUR functional characteristics in *Brassica rapa*. *Int J Mol Sci*. 2020;21(17):5987. doi 10.3390/ijms21175987

Fattorini L., Falasca G., Kevers C., Rocca L.M., Zadra C., Altamura M.M. Adventitious rooting is enhanced by methyl jasmonate in tobacco thin cell layers. *Planta*. 2009;231(1):155-168. doi 10.1007/s00425-009-1035-y

Filyushin M.A., Anisimova O.K., Kochieva E.Z., Shchennikova A.V. Correlation of ascorbic acid content and the pattern of monodehydroascorbate reductases (*MDHARs*) gene expression in leek (*Allium porrum* L.). *Russ J Plant Physiol*. 2021;68(5):849-856. doi 10.1134/S1021443721050034

Filyushin M.A., Seredin T.M., Shchennikova A.V., Kochieva E.Z. Vitamin C content and profile of ascorbate metabolism gene expression in green leaves and bleached parts of the pseudostem of leek (*Allium porrum* L.) F_1 hybrids. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2025;29(2):200-209. doi 10.18699/vjgb-25-23

Franceschi V.R., Tarlyn N.M. L-Ascorbic acid is accumulated in source leaf phloem and transported to sink tissues in plants. *Plant Physiol*. 2002;130(2):649-656. doi 10.1104/pp.007062

Fu Q., Cao H., Wang L., Lei L., Di T., Ye Y., Ding C., Li N., Hao X., Zeng J., Yang Y., Wang X., Ye M., Huang J. Transcriptome analysis reveals that ascorbic acid treatment enhances the cold tolerance of tea plants through cell wall remodeling. *Int J Mol Sci*. 2023; 24(12):10059. doi 10.3390/ijms241210059

Ha J., Kang Y.G., Lee T., Kim M., Yoon M.Y., Lee E., Yang X., Kim D., Kim Y.J., Lee T.R., Kim M.Y., Lee S.H. Comprehensive RNA sequencing and co-expression network analysis to complete the biosynthetic pathway of coumestrol, a phytoestrogen. *Sci Rep.* 2019; 9(1):1934. doi 10.1038/s41598-018-38219-6

Huang J., Wu H., Gao R., Wu L., Wang M., Chu Y., Shi Y., Xiang L., Yin Q. Integrated multi-omics analysis reveals glycosylation involving 2-O-β-D-Glucopyranosyl-L-ascorbic acid biosynthesis in *Lycium barbarum*. *Int J Mol Sci.* 2025;26(4):1558. doi 10.3390/ijms26041558

Isherwood F.A., Mapson L.W. Biological synthesis of ascorbic acid: the conversion of derivatives of D-galacturonic acid into L-ascorbic acid by plant extracts. *Biochem J.* 1956;64(1):13-22. doi 10.1042/bj0640013

Ishikawa T., Nishikawa H., Gao Y., Sawa Y., Shibata H., Yabuta Y., Maruta T., Shigeoka S. The pathway via D-galacturonate/L-galactonate is significant for ascorbate biosynthesis in *Euglena gracilis*: identification and functional characterization of aldonolactonase. *J Biol Chem*. 2008;283(45):31133-31141. doi 10.1074/jbc.M803930200

Li H., Huang W., Wang G.L., Wang W.L., Cui X., Zhuang J. Transcriptomic analysis of the biosynthesis, recycling, and distribution of ascorbic acid during leaf development in tea plant (*Camellia sinensis* (L.) O. Kuntze). *Sci Rep*. 2017;7:46212. doi 10.1038/srep46212

Liu H., Wei L., Ni Y., Chang L., Dong J., Zhong C., Sun R., Li S., Xiong R., Wang G., Sun J., Zhang Y., Gao Y. Genome-wide analysis of ascorbic acid metabolism related genes in *Fragaria × ananassa* and its expression pattern analysis in strawberry fruits. *Front Plant Sci*. 2022;13:954505. doi 10.3389/fpls.2022.954505

Mohnen D. Pectin structure and biosynthesis. *Curr Opin Plant Biol*. 2008;11(3):266-277. doi 10.1016/j.pbi.2008.03.006

Mølhøj M., Verma R., Reiter W.D. The biosynthesis of D-Galacturonate in plants. functional cloning and characterization of a membrane-anchored UDP-D-Glucuronate 4-epimerase from *Arabidopsis*. *Plant Physiol*. 2004;135(3):1221-1230. doi 10.1104/pp.104.043745

Peltonen K.E., Richard P. Identification of a D-galacturonate reductase efficiently using NADH as a cofactor. *Biotechnol Rep* (*Amst*). 2022; 35:e00744. doi 10.1016/j.btre.2022.e00744

Penning T.M. The aldo-keto reductases (AKRs): overview. *Chem Biol Interact*. 2015;234:236-246. doi 10.1016/j.cbi.2014.09.024

Popa P.-M., Băbeanu C., Cosmulescu S.-N. Evaluation of chemical compounds in local garlic genotypes from southwestern Romania. *Appl Sci.* 2024;14(16):6899. doi 10.3390/app14166899

Quiñones C.O., Gesto-Borroto R., Wilson R.V., Hernández-Madrigal S.V., Lorence A. Alternative pathways leading to ascorbate biosynthesis in plants: lessons from the last 25 years. *J Exp Bot*. 2024;75(9):2644-2663. doi 10.1093/jxb/erae120

Richardson A.T., McGhie T.K., Cordiner S.B., Stephens T.T.H., Larsen D.S., Laing W.A., Perry N.B. 2-O-β-d-Glucopyranosyl l-ascorbic acid, a stable form of vitamin C, is widespread in crop plants. *J Agric Food Chem*. 2021;69(3):966-973. doi 10.1021/acs.jafc.0c06330

Ruggieri V., Bostan H., Barone A., Frusciante L., Chiusano M.L. Integrated bioinformatics to decipher the ascorbic acid metabolic network in tomato. *Plant Mol Biol*. 2016;91(4-5):397-412. doi 10.1007/s11103-016-0469-4

Sengupta D., Naik D., Reddy A.R. Plant aldo-keto reductases (AKRs) as multi-tasking soldiers involved in diverse plant metabolic processes and stress defense: a structure-function update. *J Plant Physiol*. 2015;179:40-55. doi 10.1016/j.jplph.2015.03.004

Shchennikova A.V., Kochieva E.Z., Filyushin M.A. Ascorbate biosynthesis and recycling genes are involved in the responses of garlic *Allium sativum* L. plants to *Fusarium proliferatum* infection. *Dokl Biochem Biophys.* 2025;520(1):49-52. doi 10.1134/S1607672924601057

Skoczylas J., Jędrszczyk E., Dziadek K., Dacewicz E., Kopeć A. Basic chemical composition, antioxidant activity and selected polyphenolic compounds profile in garlic leaves and bulbs collected at various stages of development. *Molecules*. 2023;28(18):6653. doi 10.3390/molecules28186653

Smirnoff N. Ascorbic acid metabolism and functions: a comparison of plants and mammals. *Free Radic Biol Med*. 2018;22:116-129. doi 10.1016/j.freeradbiomed.2018.03.033

Smirnoff N., Wheeler G.L. The ascorbate biosynthesis pathway in plants is known, but there is a way to go with understanding control and functions. *J Exp Bot*. 2024;75(9):2604-2630. doi 10.1093/jxb/erad505

Smirnoff N., Conklin P.L., Loewus F.A. biosynthesis of ascorbic acid in plants: a renaissance. *Annu Rev Plant Physiol Plant Mol Biol*. 2001; 52:437-467. doi 10.1146/annurev.arplant.52.1.437

Šnirc M., Lidiková J., Čeryová N., Pintér E., Ivanišová E., Musilová J., Vollmannová A., Rybnikár S. Mineral and phytochemical profiles of selected garlic (*Allium sativum* L.) cultivars. *South Afr J Bot*. 2023;158:319-325. doi 10.1016/j.sajb.2023.05.024

Suekawa M., Fujikawa Y., Inada S., Murano A., Esaka M. Gene expression and promoter analysis of a novel tomato aldo-keto reductase in response to environmental stresses. *J Plant Physiol*. 2016;200:35-44. doi 10.1016/j.jplph.2016.05.015

Sun X., Zhu S., Li N., Cheng Y., Zhao J., Qiao X., Lu L., … Zhao X., Tian S., Su J., Cheng Z., Liu T. A chromosome-level genome assembly of garlic (*Allium sativum*) provides insights into genome evolution and allicin biosynthesis. *Mol Plant*. 2020;13(9):1328-1339. doi 10.1016/j.molp.2020.07.019

Teper-Bamnolker P., Buskila Y., Lopesco Y., Ben-Dor S., Saad I., Holdengreber V., Belausov E., Zemach H., Ori N., Lers A., Eshel D. Release of apical dominance in potato tuber is accompanied by pro-

grammed cell death in the apical bud meristem. *Plant Physiol*. 2012; 158(4):2053-2067. doi 10.1104/pp.112.194076

Vargas J.A., Leonardo D.A., D'Muniz Pereira H., Lopes A.R., Rodriguez H.N., Cobos M., Marapara J.L., Castro J.C., Garratt R.C. Structural characterization of L-galactose dehydrogenase: an essential enzyme for vitamin C biosynthesis. *Plant Cell Physiol*. 2022;63(8): 1140-1155. doi 10.1093/pcp/pcac090

Xie Q., Essemine J., Pang X., Chen H., Cai W. Exogenous application of abscisic acid to shoots promotes primary root cell division and elongation. *Plant Sci*. 2020;292:110385. doi 10.1016/j.plantsci.2019.110385

Xu X., Zhang Q., Gao X., Wu G., Wu M., Yuan Y., Zheng X., … Qi T., Li H., Luo Z., Li Z., Deng W. Auxin and abscisic acid antagonistically regulate ascorbic acid production via the SlMAPK8-SlARF4-SlMYB11 module in tomato. *Plant Cell*. 2022;34(11):4409-4427. doi 10.1093/plcell/koac262

Yenealem D., Eyayu D., Tibebe D., Mulugeta M., Kassa Y., Moges Z., Kerebih F., Fentie T., Amare A., Ayalew M. Electrochemical characterization and detection of ascorbic acid in garlic using activated glassy carbon electrode: a comprehensive study. *Food Anal Methods*. 2024;17:1473-1483. doi 10.1007/s12161-024-02660-3

Younis M.E., Hasaneen M.N., Kazamel A.M. Exogenously applied ascorbic acid ameliorates detrimental effects of NaCl and mannitol stress in *Vicia faba* seedlings. *Protoplasma*. 2010;239(1-4):39-48. doi 10.1007/s00709-009-0080-5

Conflict of interest. The authors declare no conflict of interest.
Received September 16, 2025. Revised January 13, 2026. Accepted January 13, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):424-434

doi 10.18699/vjgb-26-46

Transcriptional changes of aquaporin genes in leaves of black medic induced by arbuscular mycorrhizal fungal inoculation under water deficit

A.P. Yurkov, T.R. Kudriashova, A.I. Belyaeva, A.A. Kryukov

All-Russia Research Institute for Agricultural Microbiology, St. Petersburg, Russia
rainniar@rambler.ru

Abstract. One of the current research directions in plant-microbe interactions focuses on the mechanisms of plant adaptation to environmental stress through symbioses with various microorganisms. While the role of arbuscular mycorrhizal fungi in plant adaptation to drought is well-known, the underlying mechanisms of these processes remain poorly understood, particularly in leaf tissues. It is suggested that certain genes from the aquaporin family play a critical role both in adaptation to water deficit and in the development of an effective arbuscular mycorrhizal symbiosis. Thus, the important task in this study of plant-microbe symbioses is to assess the effect of arbuscular mycorrhizal fungal inoculation on the expression of aquaporin genes in leaves. This study utilizes the highly effective plant-microbe model system "*Medicago lupulina* + *Rhizophagus irregularis*" under drought stress conditions. A comparative assessment of gene transcription was carried out using the $2^{-\Delta\Delta CT}$ method based on real-time quantitative PCR results: normalization was performed relative to the actin reference gene with non-inoculated plants serving as the control. The study was conducted both at the initial development stage (the 2nd leaf stage), and at the stage of active plant-microbe interaction (the flowering stage). The study revealed genes with significant differential expression under drought conditions when comparing mycorrhizal and non-mycorrhizal *Medicago lupulina* plants: *NIP3;1*, *NIP4;2*, specific *NIP7;1*, *TIP5;1* at the 2nd leaf stage; genes *NIP3;1*, *NIP5;1*, *NIP6;4*, *NIP7;1* (specific), *PIP1;4*, *TIP2;3* and specific *XIP1;1* at the flowering stage. Previously, in a similar experiment, under well-watering conditions, the same genes did not have differential expression between mycorrhizal and non-mycorrhizal plants. Thus, the listed genes likely participate in the adaptation of the studied plants to drought conditions. The obtained information can be used to develop highly productive plant-microbe systems involving arbuscular mycorrhizal fungi, aimed at transitioning to organic farming, minimizing negative environmental impact, and enhancing plant resistance to water deficit.

Key words: aquaporins; AQP; arbuscular mycorrhiza; *Medicago lupulina*; *Rhizophagus irregularis*; drought

For citation: Yurkov A.P., Kudriashova T.R., Belyaeva A.I., Kryukov A.A. Transcriptional changes of aquaporin genes in leaves of black medic induced by arbuscular mycorrhizal fungal inoculation under water deficit. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):424-434. doi 10.18699/vjgb-26-46

Funding. The work was supported by Russian Science Foundation No. 24-26-00181.

Acknowledgements. The work was carried out using the equipment of the Center for Collective Use "Genomic Technologies, Proteomics and Cell Biology" at the ARRIAM.

Изменение транскрипции генов аквапоринов в листьях люцерны хмелевидной в результате микоризации грибом арбускулярной микоризы в условиях дефицита воды

А.П. Юрков, Т.Р. Кудряшова, А.И. Беляева, А.А. Крюков

Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии, Санкт-Петербург, Россия
rainniar@rambler.ru

Аннотация. Одним из актуальных направлений исследований растительно-микробных взаимодействий является изучение механизмов адаптации растений к стресс-факторам среды за счет развития симбиозов с различными микроорганизмами. Хорошо известна роль грибов арбускулярной микоризы в адаптации растений к засухе, но механизмы данных процессов до сих пор раскрыты не полностью, особенно в тканях листьев. Предполагается, что некоторые из генов аквапоринов могут играть важную роль как в адаптации растений к недостатку влаги, так и в развитии эффективного симбиоза с грибами арбускулярной микоризы. Таким образом, важной задачей в изучении растительно-микробных симбиозов является оценка влияния инокуляции грибом арбускулярной микоризы на экспрессию генов аквапоринов в листьях растений в высокоэффективной модельной растительно-микробной

системе "*Medicago lupulina* + *Rhizophagus irregularis*" в условиях засухи. Сравнительная оценка транскрипции генов выполнена методом $2^{-\Delta\Delta CT}$ по результатам количественной ПЦР в реальном времени: нормализация проведена по отношению к референсному гену – актину, контролем выступал вариант без инокуляции грибом. Исследование проведено как в начальную фазу развития – фазу развития 2-го листа, так и в фазу активного растительно-микробного взаимодействия – фазу цветения. В результате выявлены гены, имеющие достоверную дифференциальную экспрессию в условиях засухи при сравнении микоризованных и немикоризованных растений люцерны хмелевидной: *NIP3;1*, *NIP4;2*, специфический *NIP7;1*, *TIP5;1* – в фазу развития 2-го листа; *NIP3;1*, *NIP5;1*, *NIP6;4*, *NIP7;1* (специфический), *PIP1;4*, *TIP2;3* и *XIP1;1* (специфический) – в фазу цветения. Ранее в подобном эксперименте, но в условиях нормального полива эти же гены не имели дифференциальной экспрессии в сравнении микоризованных и немикоризованных растений. Вероятно, перечисленные гены принимают участие в адаптации изученных растений к условиям засухи. Полученные сведения могут быть использованы в разработке высокопродуктивных растительно-микробных систем с участием грибов арбускулярной микоризы с целью перехода к биологическому земледелию, минимизации негативного влияния на окружающую среду и повышению устойчивости растений к недостатку влаги.
Ключевые слова: аквапорины; AQP; арбускулярная микориза; *Medicago lupulina*; *Rhizophagus irregularis*; засуха

Introduction

The problem of water deficit in crop production has become increasingly acute due to the disruption of natural ecosystems by intensive agriculture, excessive water withdrawal, and the weather conditions in various regions. To solve the problem of drought, a number of measures are employed, including not only irrigation, the introduction of new methods of no-till farming, shelterbelts, and crop rotation but also biological methods aimed at enhancing plant adaptation to environmental stress factors using symbiotic microorganisms. The role of arbuscular mycorrhiza (AM) fungi is well known; they likely played a key role in the colonization of land by plants approximately 400 million years ago (Remy et al., 1994). During that period, these fungi performed part of the functions of the root system, actively supplying plants with water and minerals. AM fungi support the ionic homeostasis of the host plant, provide osmotic protection and increase water use efficiency (Mammadov et al., 2018; Luo et al., 2022). Currently, more than 80 % of terrestrial plant species form AM with fungi from the class Glomeromycetes (Smith, Read, 2008). In this regard, a highly relevant area of research is to identify the mechanisms that control the adaptation of plants to soil moisture deficiency and the role of AM fungi in this process.

It should be assumed that plant adaptation to drought is closely related to the regulation of water transporters in plant tissues, among which the most represented group are aquaporins, small membrane proteins of the Major Intrinsic Proteins (MIP) family that form channels for transporting molecules across biological membranes (Maurel et al., 2015). In plants, they play a key role in adaptation to drought, salinity, and growth regulation (Daneliia et al., 2024; Kudriashova et al., 2025). The greatest diversity of aquaporins is characteristic of angiosperms, resulting from gene duplication via polyploidization (Singh et al., 2020), a process common to nearly all angiosperms. Polyploidization leads to gene duplication and the emergence of new isoforms. In particular, 35 aquaporin genes were found in *Arabidopsis thaliana*, 46 in *Medicago truncatula*, and 120 in rapeseed *Brassica napus* (Min et al., 2019; Daneliia et al., 2024). Aquaporins have a conservative structure. An important feature in their structure is that they form tetramers where each monomer functions as an independent channel (Kudoyarova et al., 2022). Aquaporin activity can be regulated by phosphorylation, pH, and redox reactions.

Not all aquaporins are equally effective in water transport. Thus, aquaporins of angiosperms are divided into the five subfamilies: (1) NIP (nodulin 26-like intrinsic proteins) with low water permeability participate in the exchange of metabolites with microsymbionts (Kruse et al., 2006), and are localized in the plasma membrane and membranes of the endoplasmic reticulum (Ma et al., 2006; Mizutani et al., 2006; Lopez et al., 2016); (2) PIP (plasma membrane intrinsic proteins), permeable to water, hydrogen peroxide, and carbon dioxide, are localized in the plasma membrane, the inner membrane of chloroplasts, thylakoid, and the endoplasmic reticulum (Zhou et al., 2024); (3) SIP (small basic intrinsic proteins) with low permeability to water are localized on the membrane of the endoplasmic reticulum, and are poorly studied (Hussain et al., 2020; Zhou et al., 2024); (4) TIP (tonoplast intrinsic proteins), permeable to water, hydrogen peroxide, ammonium, and urea, are localized in the tonoplast (Maurel et al., 2008; Zhou et al., 2024); (5) XIPs (uncharacterized/X intrinsic proteins) with low water permeability are localized on the plasma membrane, and are poorly studied (Lopez et al., 2016; Noronha et al., 2016).

An analysis of the literature data suggests that the functions of aquaporins require significantly more investigation. Some studies indicate that mycorrhization increases plant resistance to drought. However, the mechanisms of aquaporin regulation remain unclear, especially regarding the species and tissue specificity. In particular, their expression in leaves is still under-researched (Daneliia et al., 2024; Kudriashova et al., 2025). Changes in aquaporin regulation mediated by mycorrhization remain an enigma (Sharma et al., 2021). Promising research directions include the identification of marker genes for effective symbioses (including AM symbiosis) that ensure plant adaptation to a lack of water in the substrate, the analysis of aquaporin gene expression under water stress, as well as the study of posttranslational modifications. From a practical perspective, such research will enable the development of symbiotically highly efficient and productive plant-microbial

systems (PMS) necessary for the implementation of soil-protective resource-saving agriculture and, as a result, the production of environmentally friendly agricultural products.

Based on the above, the purpose of this study is to identify aquaporin genes in leaves that serve as markers of effective symbiosis. We employed a highly efficient model – PMS "*Medicago lupulina* + *Rhizophagus irregularis*" – under conditions of substrate moisture deficiency at the early and late stages of symbiosis development. For this purpose, the authors selected the highly responsive to mycorrhization line MlS-1 of *Medicago lupulina*. Previous transcriptomic analysis using the Massive Analysis of cDNA Ends (MACE-Seq) on this line identified over 4,500 differentially expressed genes ($p_{adj} < 0.01$) in leaves upon colonization by the highly effective strain of the AM fungus *Rhizophagus irregularis* RCAM00320 (Yurkov et al., 2023).

Materials and methods

Plant and fungal materials. Black medic (*Medicago lupulina* L.), which is a widespread species of the genus *Medicago*, diploid and self-pollinating, was used as a model plant. The MlS-1 line, which was bred from the VIK32 cultivar population and is characterized by its high sensitivity to mycorrhization, was selected as the object of the study (Yurkov et al., 2015). In the absence of arbuscular mycorrhizal (AM) fungal inoculation, plants of this line show signs of dwarfism under conditions of low plant-available phosphorus in the substrate (Yurkov et al., 2020). The RCAM00320 *Rhizophagus irregularis* strain from the ARRIAM collection, which has high symbiotic efficiency, was used for inoculation.

Pot experiment. The experimental procedure was based on the protocol described by A.P. Yurkov et al. (2015). To prevent spontaneous infection by nodule bacteria and other microorganisms, the soil-sand mixture was sterilized (the substrate ratio was 2:1) by double autoclaving at 134 °C and 2 atm for 1 hour with an interval of 2 days. No substrate toxicity was detected after treatment of the mixture. Two *Medicago* seedlings were planted per pot, each containing 210 g of a soil-sand mixture. Half of the pots were inoculated with *Plectranthus verticillatus* roots colonized by *R. irregularis*, and in the other half (in the control), no inoculation was performed. Agrochemical parameters of the used soil were reported in the work of A.P. Yurkov et al. (2020): P_2O_5 content was 23 mg/kg and the pH was 6.44. The humidification conditions were selected based on a preliminary drought trial. The experimental design included variants with different moistening of plants, both with and without AM *R. irregularis* inoculation:

– "normal", watering maintained at 0.6 of the total water-holding capacity (WHC);
– "drought, variant 1", transition from normal watering in the first 8 days to 0.4 of WHC for 16 days (until the 2nd leaf stage) and for 40 days (until the flowering stage);
– "drought, variant 2", transition from normal watering to 0.4 of WHC, for 7 days before harvesting at the 2nd leaf stage and the flowering stage (i. e., the drought regime was maintained for a plant for 7 days before harvest).

The third variant of drought (continuous 0.4 of WHC) was excluded due to high mortality rates in preliminary tests. In this preliminary experiment (Table S1)[1], shoot fresh weight, symbiotic efficiency, plant mortality rate, and mycorrhizal infection frequency in the root were determined. Based on these indicators, drought variant 2 was selected for the analysis of the relative level of aquaporin gene expression in leaves.

In the main experiment, the first harvest was carried out 24 days after planting and inoculation, during the stage of the 2nd true leaf, and the second harvest – on day 48, during the flowering stage. Productivity parameters (shoot and root dry weight) and the effectiveness of symbiosis were evaluated according to previously described methods (Yurkov et al., 2015). The roots were stained with trypan blue (Phillips, Hayman, 1970). The calculation of mycorrhizal parameters, including the intensity of mycorrhizal infection in the root (*M*, %) and the abundance of arbuscules in the mycorrhizal part of the root (*a*, %), were carried out according to (Trouvelot et al., 1986) using specialized software (Vorobyev et al., 2016). To analyze gene expression levels, plant material was frozen in liquid nitrogen immediately after harvesting and stored at –80 °C until RNA extraction.

Isolation of RNA and analysis of gene expression. The selection of genes and primers for their amplification (Table S2) was carried out on the basis of *M. truncatula* data from the Phytozome genetic sequence database (https://phytozome-next.jgi.doe.gov/) and transcriptomic analysis of *M. lupulina* (Yurkov et al., 2023). The study included 30 genes from 46 known orthologs in *M. truncatula*, reported by K. Min et al. (2019). 16 genes had very low expression levels across all experimental variants, ranging from three to six orders of magnitude lower than the reference gene (with quantification cycles appearing more than 10 cycles later). Such low expression levels in the analysis are associated with a high error rate; therefore, these genes were excluded from further analysis. The absence of non-target PCR products was controlled by gel electrophoresis and analysis of melting curves. The efficiency of the primers was calculated using serial cDNA dilutions, primers with an efficiency close to 100 % (>95 %) were included in the work. The primers were tested at both sampling time points.

Total RNA from plant material was isolated using TRIzol reagent (Thermo Fisher Scientific, USA) with modifications (MacRae, 2007). Before cDNA synthesis, the quality of DNAase treatment was checked by PCR, cDNA synthesis was carried out from ~1 microgram of total RNA per sample using the Maxima First Strand cDNA Synthesis Kit with dsDNase according to manufacturer's instructions (Thermo Fisher Scientific, USA). The quality of the obtained cDNA was checked by ubiquitin gene amplification. The level of gene expression was assessed by PCR-RT using a C1000 thermal cycler with a CFX-96 module (BioRad, USA) and a SYBR Green I reagent kit (Syntol, Russia). The amplification conditions were:

[1] Supplementary Tables S1 and S2 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Yurkov_Engl_30_3.pdf

initial denaturation – 95 °C, 5 min; 40 cycles – 95 °C, 15 s; 60 °C, 30 s; 72 °C, 30 s; followed by melt curve analysis. To compare the expression levels of the analyzed aquaporin genes, the $2^{-\Delta\Delta CT}$ method was used: the expression level in the variant with mycorrhization was compared against the control without AM. Normalization was performed relative to the reference gene, actin, according to (Yurkov et al., 2020). The PCR mix (volume – 10 µl) included: 1 µl of 10x B + SYBR Green buffer, 1 µl of 2.5 mM dNTP, 1 µl of $MgCl_2$ (25 mM), 0.3 µl of each primer (10 mM), 0.125 µl (0.625 units) SynTaq DNA polymerase (Syntol, Russia), 4.275 µl of deionized water and 2 µl of cDNA. Each sample (both with and without AM) was analyzed in three biological and four technical replicates.

Statistical data processing was performed using ANOVA followed by Tukey's post-hoc test with significance defined at $p < 0.05$. Student's *t*-test ($p < 0.05$) was used to compare the expression levels between the "+AM" and "–AM" groups.

Results

The indicators of mycorrhization of *M. lupulina* and the effect of inoculation by the AM fungus on the growth, development of the host plant, and expression of aquaporin genes in leaves under drought conditions were evaluated. To simulate water stress, a drought regime was selected (based on a preliminary experiment; Table S1), consisting of watering at the rate of 0.4 of the total water-holding capacity (WHC) for one week prior to each experiment harvest, conducted 24 and 48 days post-inoculation.

M. lupulina productivity analysis (Fig. 1*A*, *B*) showed that plants inoculated with the fungus *R. irregularis* showed a significant increase in the shoot dry biomass at both the 2nd leaf stage (day 24) and the flowering stage (day 48). On the other hand, an increase in the root dry weight during mycorrhization was observed only at a late stage of development (at 48 days). Notably, the symbiotic efficiency of AM (MGR, Mycorrhizal Growth Response) was significant and high during the flowering stage (more than 100 %; Fig. 1*C*, *D*).

Microscopic analysis of *M. lupulina* roots showed the active development of mycorrhiza and arbuscules in the mycorrhized part of the root (Fig. 1*E*, *F*). There was an increase in the intensity of AM in the root (*M*) from 24 days to 48 days, and the abundance of arbuscules in the mycorrhized part of the root (*a*) was already high starting from the 2nd leaf stage; these results indicate a high activity of AM symbiosis under simulated drought conditions and provide a reliable basis for evaluating the effect of mycorrhization on the expression of aquaporin family genes in the leaves. Water stress was sufficient to induce mechanisms of adaptation to drought, but not critical for the survival of *M. lupulina* and maintaining a high level of mycorrhization.

The indicators of AM efficiency and activity in conditions of lack of moisture (watering 0.4 MC of WHC) were comparable to the results obtained earlier with standard humidification (0.6 of WHC), in the model PMS "MlS-1 *M. lupulina* line + RCAM00320 *R. irregularis* strain" (Yurkov et al., 2021). This confirms that even under drought stress, it is possible to maintain a highly functional mycorrhizal symbiosis; furthermore, moderate drought stress stimulated a change in the transcription of aquaporin genes without suppressing the effectiveness of AM symbiosis.

To address the study's objectives, primers were selected and tested (Table S2). The analysis revealed that a subset of 30 aquaporin genes showed differential expression in *M. lupulina* leaves in response to mycorrhization under drought conditions (Fig. 2 and 3). Among these, genes with specific expression patterns were identified (marked with "+" for presence and "n. d." for not detected, Fig. 2 and 3).

On day 24 (the stage of the 2nd true leaf), the following expression changes were observed in the leaves of mycorrhized plants (light columns) compared to the control (dark gray columns):

– significant upregulation during mycorrhization (expression during mycorrhization was significantly ($p < 0.05$) higher than that in the control without inoculation by the AM fungus) for the *NIP3;1*, *NIP4;2*, *TIP2;2* genes, as well as specific expression of *NIP7;1*;
– significant downregulation during mycorrhization (expression during mycorrhization was significantly ($p < 0.05$) lower than that in the control without inoculation with the AM fungus) for the *PIP2;3*, *TIP1;1*, *TIP1;4*, *TIP2;3*, *TIP4;1* genes.

On day 48 (the flowering stage), the following was detected in the leaves of plants inoculated with the AM fungus:

– significant upregulation during mycorrhization of the *NIP3;1*, *NIP5;1*, *NIP6;4*, *PIP1;2*, *PIP1;4*, *PIP2;1*, *TIP2;3* genes, as well as specific expression of *NIP7;1*, *XIP1;1*;
– significant downregulation during mycorrhization of the *NIP1;2*, *NIP1;5*, *NIP2;1*, *NIP4;2*, *PIP2;3*, *SIP1;3*, *TIP1;1*, *TIP2;2*, *TIP4;1* genes.

Thus, the most significant aquaporin genes involved in the development of effective AM symbiosis under water deficit have been identified. In particular, *NIP7;1* demonstrated specific or induced expression during mycorrhization in both the vegetative and reproductive phases of the host plant. Special attention should also be paid to the *XIP1;1* gene with specific expression during the flowering stage.

The genes of the NIP subfamily generally exhibited upregulation (increased expression level ($p < 0.05$) relative to the level in the control without inoculation with AM fungus) in the early stage of AM symbiosis development, but by the late stage of development, the flowering stage, both up- and downregulation were observed for the genes of this family. This suggests complex regulation of aquaporins depending on the vegetative/reproductive stage of the host plant and with a restructuring of water metabolism under drought conditions during mycorrhization.

Discussion

The results indicate that aquaporins of the NIP and TIP subfamilies are actively involved in the development of AM symbiosis under drought conditions. However, according to research, only TIP aquaporins significantly affect the transmembrane

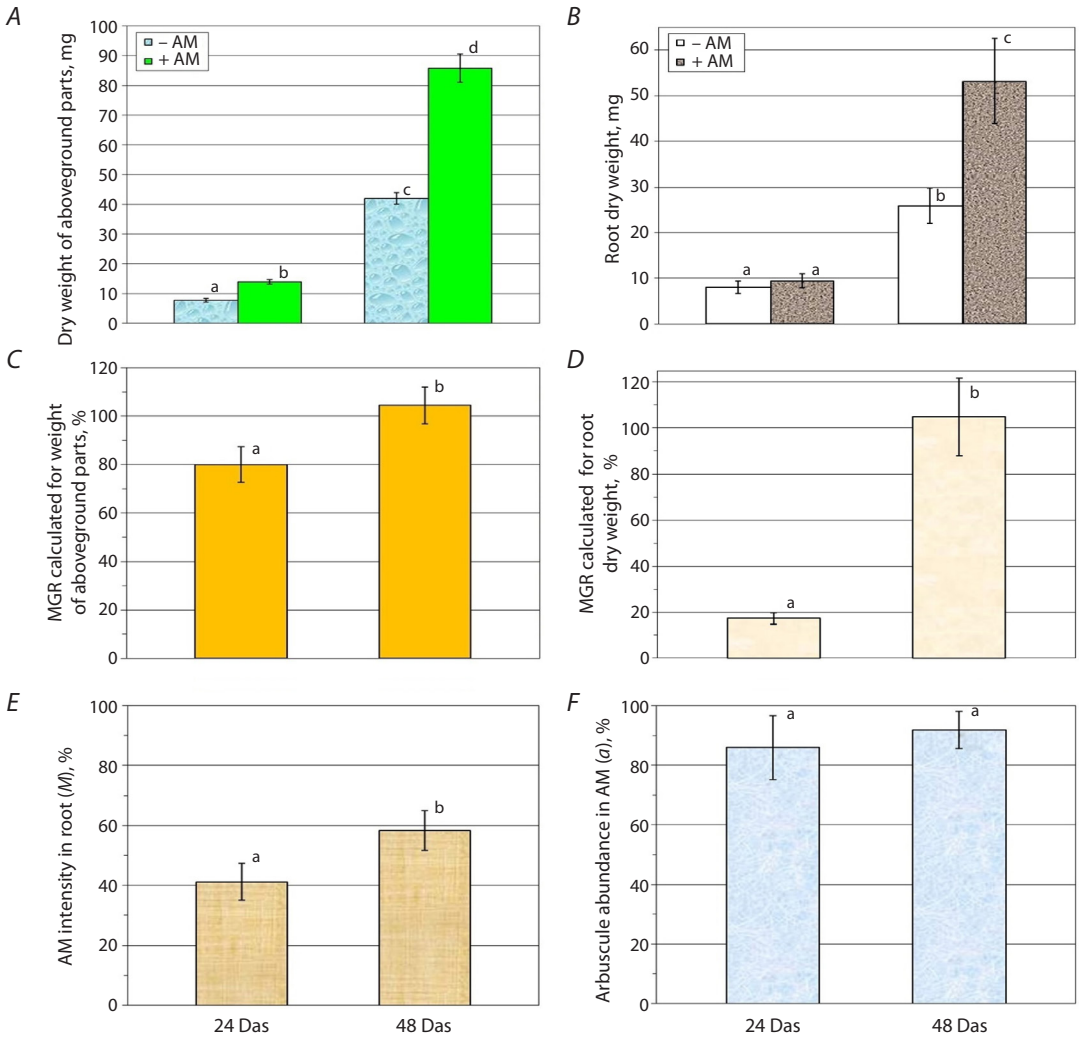


Fig. 1. Indicators of productivity (*A* – shoot dry weight, mg, *B* – root dry weight, mg), symbiotic efficiency calculated by shoot weight, % (*C*) and root weight, % (*D*), and indicators of mycorrhization: *E* – intensity of mycorrhizal infection in the root (*M*); *F* – abundance of arbuscules in mycorrhizal part of the root (*a*) on days 24 and 48 after planting and inoculation.
a, b, c – different letters above the columns indicate significant ($p < 0.05$) differences.

water transport (Zhou et al., 2024). At the same time, NIP aquaporins transport diverse substrates, including metalloids, but have low water permeability. It is assumed that NIP proteins are involved in the exchange of metabolites between the host plant and symbiotic microorganisms (Kruse et al., 2006). It is also known that plants may downregulate aquaporin gene expression under drought to conserve water; therefore, special attention should be paid to genes, the expression of which is suppressed under conditions of water deficit (Quiroga et al., 2019). At the same time, according to other data, water deficit can trigger the activation of certain aquaporin genes that play a key role in plant resistance to drought (Jia, Liu, 2020; Zhou et al., 2024). Thus, the relationship between the expression of these genes, the effectiveness of plant symbiosis with AM fungi, and drought resistance remains poorly understood (Sharma et al., 2021). Further investigation could elucidate the mechanisms of adaptation of AM plants to water stress.

The results of this study indicate that the selected irrigation regime (0.4 of WHC for 7 days prior to harvest) was sufficient for the development of symbiosis, as evidenced by the significant response to mycorrhization and the high frequency of AM fungal colonization in the roots of the host plant (refer to "drought, variant 2", Table S1). Meanwhile, productivity indicators were reduced compared to those under standard watering conditions at 0.6 of WHC (Yurkov et al., 2020). It was shown that the symbiotic efficiency of the model system "*M. lupulina* + *R. irregularis*" under water deficit (0.4 of WHC) was high at both the early and late developmental stages. At 24 days (2nd leaf stage), a significant increase in shoot biomass due to mycorrhization was observed. At 48 days (flowering stage), the system was characterized by higher symbiotic efficiency in terms of both shoots and roots dry weight, as well as a higher activity of AM development, reflected in increased root mycorrhization intensity (*M*, %).

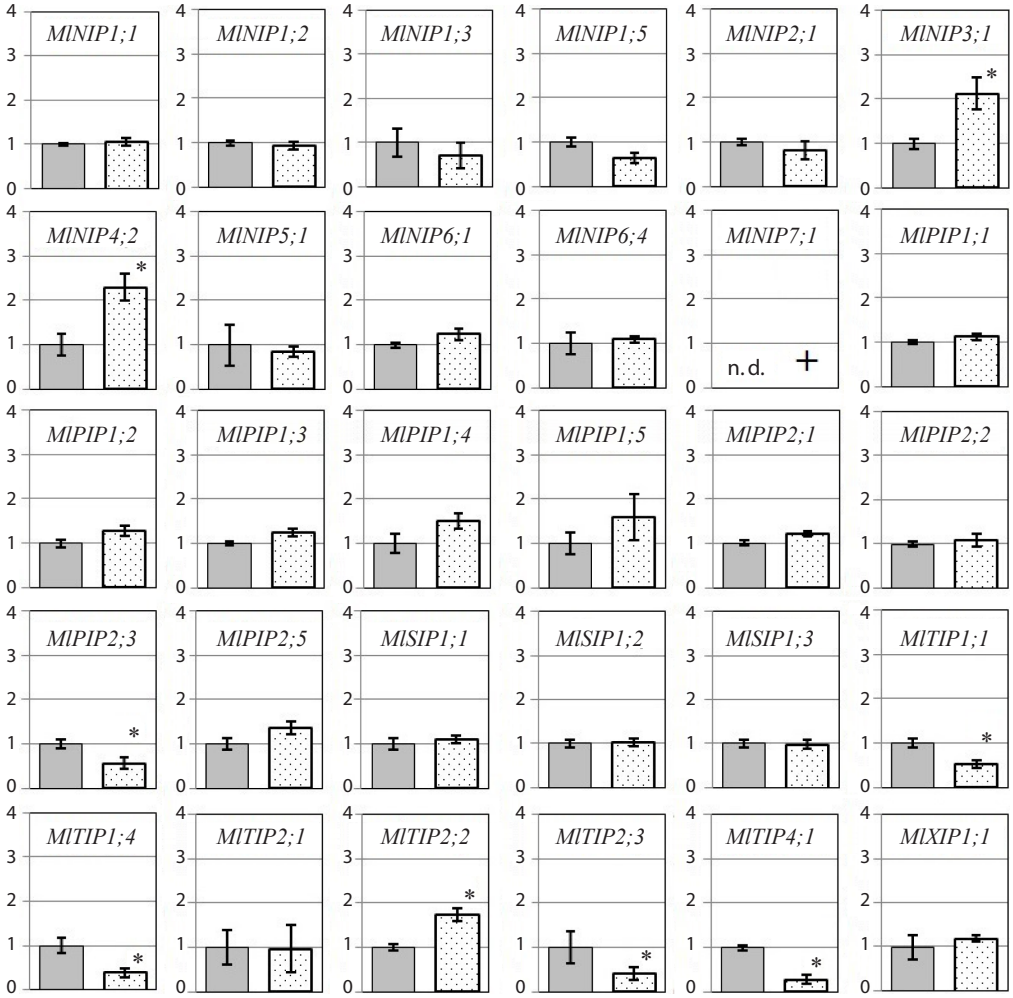


Fig. 2. The relative level of transcripts (normalized value $2^{-\Delta\Delta CT}$) in *M. lupulina* leaves 24 days after sowing and inoculation with the AM fungus *R. irregularis* (during the 2nd leaf development stage) under drought condition.

Here and in Figure 3: * Significant ($p < 0.05$) differences in the variant with and without AM. "+" the presence of specific expression in the variant. "n. d." not detected.

Meanwhile, the abundance of arbuscules in the mycorrhizal part of the root (*a*, %) was already high at the early stage of symbiosis development.

It was previously shown that under well-watered conditions (0.6 of WHC) (Yurkov et al., 2021), the indicators of AM efficiency and activity were also high. Thus, the selected drought regime caused stress sufficient to change the expression of a number of genes, but not critical for the functioning of the AM symbiosis itself. However, the role of aquaporins in the development of effective PMS under water deficit, as well as their regulation during mycorrhiza in general, has not been sufficiently studied. This is due to the broad substrate-specificity of these transporters, and their diverse functions depending on the type of tissue and subcellular localization (Daneliia et al., 2024; Kudriashova et al., 2025). The data obtained in this study revealed multidirectional changes in gene expression depending on the subfamily and phase of development: in the early phase (24 days), upregulation of NIP subfamily genes (*NIP3;1*, *NIP4;2*, and specific expression of *NIP7;1*) was observed against the background of general downregulation of TIP subfamily genes (*TIP1;1*, *TIP1;4*, *TIP2;3*, *TIP4;1*). At the late stage (48 days), the parity between up- and downregulation was found in the NIP subfamily (four genes each, including one gene with specific expression during mycorrhiza – *NIP7;1*). The TIP subfamily, as in the early phase, was characterized by the predominance of downregulation (three genes). Furthermore, a gene from the XIP subfamily (*XIP1;1*) with specific expression upon mycorrhization during the flowering stage was also identified. It should be noted that the XIP subfamily remains poorly understood, though. XIP proteins are known to be localized on the plasma membrane (Lopez et al., 2016; Noronha et al., 2016).

It is of particular interest to compare the data obtained in this study under drought conditions (at 0.4 of WHC) with the transcriptomic analysis of mycorrhizal and non-mycorrhizal *M. lupulina* plants under normal watering conditions (at 0.6 of WHC; Yurkov et al., 2023). Only one gene, *TIP2;2* (one of four upregulated genes under drought conditions), also had

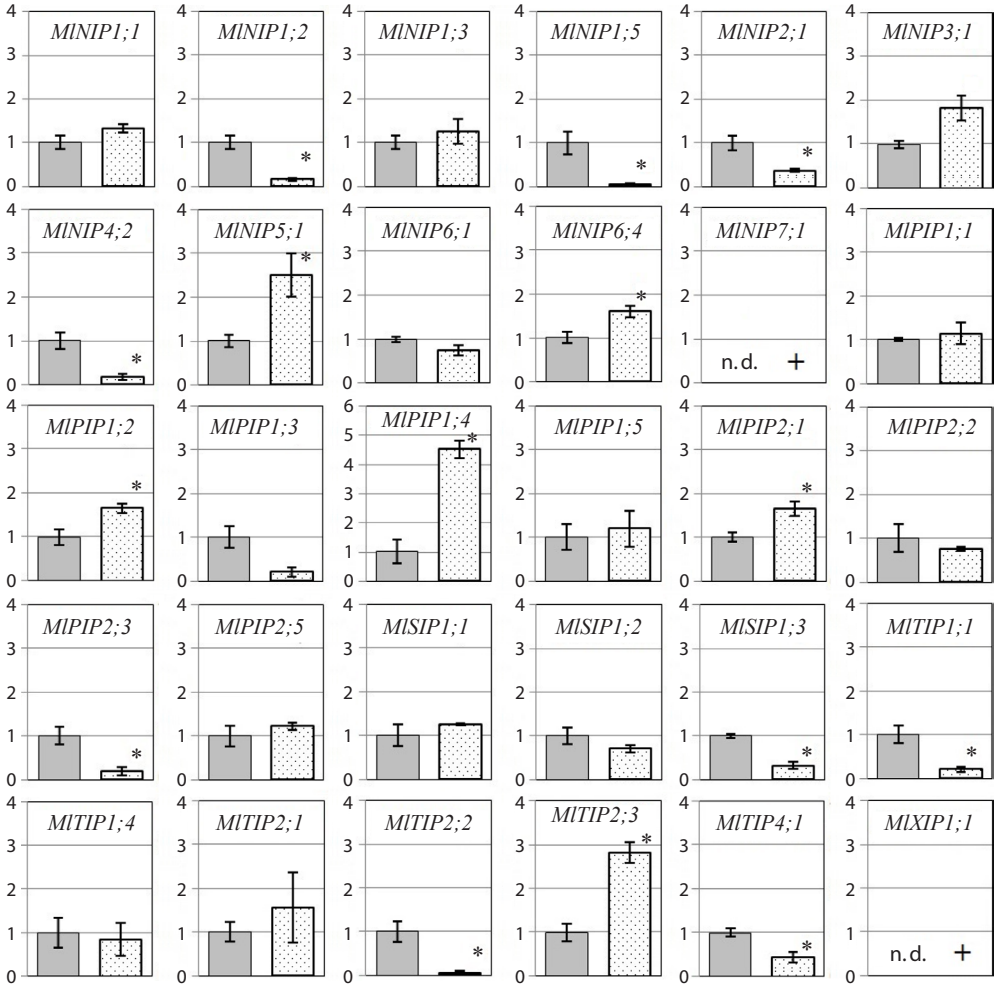


Fig. 3. The relative level of transcripts (normalized value $2^{-\Delta\Delta CT}$) in *M. lupulina* leaves 48 days after sowing and inoculation with the AM fungus *R. irregularis* (during the flowering stage) under drought condition.

upregulation under normal watering conditions during the 2nd leaf stage, as well as *PIP1;2* and *PIP2;1* (two of eight) during the flowering stage (Yurkov et al., 2023). Among the genes characterized by a decrease in expression under drought, only the *NIP1;5* gene (one out of nine) had a similar reduction in transcription under normal watering (Yurkov et al., 2023). Thus, the remaining genes that exhibited differential expression under drought conditions in the current study, but showed no such response under normal watering conditions in the 2023 experiment, may be considered candidate genes for the adaptation of *M. lupulina* plants to drought conditions. This list includes: *NIP3;1*, *NIP4;2*, specific *NIP7;1*, *TIP5;1* – at the 2nd leaf stage; *NIP3;1*, *NIP5;1*, *NIP6;4*, *NIP7;1* (specific), *PIP1;4*, *TIP2;3* and *XIP1;1* (specific) – at the flowering stage.

A comparative analysis with other PMS revealed several parallels (Asadollahi et al., 2023; Daneliia et al., 2024; Wang et al., 2024; Kudriashova et al., 2025). In the PMS "*Zea mays* + *R. irregularis*", the *ZmTIP2;3* gene was also upregulated during mycorrhization and presumably played an important role in increasing the resistance of corn to drought stress. However, this PMS was characterized by a lower responsiveness to inoculation by the AM fungus, since only a few maize genes (*ZmPIP1;6*, *ZmPIP2;2*, *ZmTIP2;3*, *ZmTIP4;1*) exhibited significant expression changes in response to mycorrhization. On the other hand, downregulation was observed in maize leaves during mycorrhization at the development stages of the 2nd and 13th leaves for some TIP subfamily genes (*ZmTIP3;1*, *ZmTIP4;4* – in both phases; *ZmTIP4;3* – only in 13th leaf stage) (Wang et al., 2024). The downregulation of certain TIP subfamily genes (*PtTIP2;1* and *PtTIP 5;1*) under drought conditions was also shown in the roots of the PMS "*Poncirus trifoliata* + *F. mosseae*" (He J.-D. et al., 2019). So, while the results for PMS "*Z. mays* + *R. irregularis*" (on day 56) are consistent with our results for PMS "*M. lupulina* + *R. irregularis*" (on day 48) for the *TIP2;3* gene, it is necessary to note the phase-dependent nature of the effects (dependence on the stage of plant development), as well as species- and tissue-specific patterns in aquaporin expression.

Supporting this high tissue specificity, previous assessment of expression in the roots of the PMS "*M. lupulina* + *R. irregularis*" (Kryukov et al., 2025) showed that during the vegetative stage (2nd leaf stage), mycorrhization led to

the upregulation of the *NIP1;2*, *NIP1;*5, *TIP2;1* genes, the specific expression of *NIP4;2*, and the downregulation of *NIP2;1*, *NIP3;1*, *PIP1;4*, *TIP3;1*, *XIP1;1*. At the same time, during the generative phase (flowering), only two genes with specific expression were identified (*NIP4;1*, *NIP7;1*), whereas a large group of 14 genes across the NIP, PIP, SIP, TIP, and XIP subfamilies was downregulated (Kryukov et al., 2025). Thus, a comparative analysis revealed tissue-specific expression of the aquaporin family genes: only the *NIP4;2* gene had upregulation in both roots and leaves in the PMS "*M. lupulina* + *R. irregularis*" in the vegetative stage; and only the *NIP7;1* gene had specific expression during mycorrhization in the generative phase. However, a spectrum of genes with downregulation in both leaves and roots during mycorrhization in the flowering stage was broader – *NIP2;1*, *SIP1;3*, *TIP2;2*, *TIP4;1*. The phase-dependent effect of marker genes for the development of effective AM symbiosis has also been identified in other gene families, for example, in the SWEET sugar transporters (Kryukov et al., 2023; Kudryashova et al., 2024; Kudriashova et al., 2025). Notably, *TIP2;3* was the only gene in the TIP subfamily that switched from downregulation to upregulation between the early (24 days) and the late (48 days) stages of mycorrhization. It is possible that *TIP2;2* functionally compensates at the stage of the 2nd leaf, since the latter showed a reverse switch from up- to downregulation. Similar hypotheses about the compensatory mechanism of transcription of other aquaporins have been proposed by R. Porcel et al. (Porcel et al., 2005; Sharma et al., 2021). The tissue-specific expression of TIP and PIP aquaporins is often linked to increased water exchange and subsequent plant growth stimulation (He F. et al., 2016). For example, the expression levels of *RpTIP2;1* and *RpPIP2;1* consistently increased in different tissues – leaves, stem, and roots under mycorrhization and drought, directing water flows toward plant tissues, whereas *RpTIP1;3* and *RpPIP1;3* were downregulated in the leaves of the PMS "*Robinia pseudoacacia* + *R. irregularis*" (He F. et al., 2016).

In the present study, downregulation of the PIP subfamily (*PIP2;3* gene) was shown in the PMS "*M. lupulina* + *R. irregularis*" both at the early (2nd leaf stage) and late (flowering) stages of development. Interestingly, according to literature data, higher activity of PIP proteins should enhance water conductivity under drought conditions. However, this hypothesis was challenged by experiments in the PMS "*Glycine max* + *F. mosseae*" and "*Lactuca sativa* + *F. mosseae*": in those studies, downregulation of PIP genes (*GmPIP1*, *GmPIP2*, *LsPIP1*, *LsPIP2*) was observed (Porcel et al., 2006). This AM-mediated effect represents a regulatory mechanism that allows the host plant to conserve metabolic resources and minimize water loss under stress conditions. According to J.M. Ruiz-Lozano and R. Aroca (2010), the AM fungi help the plant retain absorbed water by reducing membrane permeability through the suppression of PIP aquaporin expression. On the other hand, since the hyphae of the fungus can supply the host plant with water themselves, this may be a strategy that allows AM plants to depend less on the intensive expression of their own aquaporins and thus save energy (Sharma et al., 2021). Based on the functional studies of aquaporins, it can be assumed that the TIP subfamily includes the main water transporters through the vacuolar membrane (tonoplast) (Zhou et al., 2024). Their suppression at the late stages of symbiosis likely contributes to minimizing water loss and optimizing cellular homeostasis (Quiroga et al., 2019). In other words, the downregulation of TIP genes at the flowering stage may reflect the strategy of water conservation in the PMS "*M. lupulina* + *R. irregularis*".

Aquaporins of the NIP subfamily are involved in the transport of various substrates, including toxic metalloids and plant glycerol (Dean et al., 1999). They play an important role in the development of interaction between host plant and symbiotic microorganisms (Kruse et al., 2006). Aquaporins research began with the study of the NLM protein from the NIP subfamily, discovered in the peribacteroid membrane of symbiotic soybean root nodules (Sandal, Marcker, 1988). Aquaporins of the NIP subfamily are also found in non-leguminous plants (Kruse et al., 2006). Thus, it can be assumed that aquaporins of the NIP subfamily play an active role in the formation and development of effective symbiosis with the AM fungus in the PMS "*M. lupulina* + *R. irregularis*". It is known that NIP aquaporins are closer in structure to bacterial AqpZ than to glycerol transporters from the GLP subfamily (Heymann, Engel, 1999). This indicates that the ancestral plant aquaporin probably did not have the ability to transport glycerol, and this function arose later in the evolution as compensation for the lack of GLP in plants (Murata et al., 2000). During adaptation to drought, the transport of glycerol in the leaves of the host plant in the PMS "*M. lupulina* + *R. irregularis*" may play an important role.

The upregulation of *NIP3;1*, *NIP4;2* as well as the specific expression of *NIP7;1* suggest that they play a role in the early stages of the symbiosis of "*M. lupulina* + *R. irregularis*". The parity between up- and downregulation in the NIP subfamily in the leaves at the late stage of the symbiosis "*M. lupulina* + *R. irregularis*" indicates a phase-dependent effect of mycorrhization on the expression of aquaporins. At the same time, aquaporin genes – *NIP3;1* and the specific *NIP7;1* – maintained upregulation across the change of developmental stages. Thus, mycorrhization by the AM fungus *R. irregularis* modulates the expression of aquaporins in a phase-specific manner. We propose that the *NIP3;1* and specific *NIP7;1* may serve as positive markers for effective AM development under drought. This activation of NIP is probably related to the active metabolic exchange in AM symbiosis. The proportion of downregulated aquaporins under drought during mycorrhization may depend on the duration of stress. For example, in the experiments of G. Bárzana et al. (2014), in the PMS "*Zea mays* + *R. intradices*", the downregulation of only *ZmNIP2;1*, and *ZmNIP2;2* was observed during short-term drought (4 days), whereas extended drought (up to 12 days) led to the suppression of a significant part of aquaporin genes (*ZmPIP1;1*, *ZmPIP1;2*, *ZmPIP1;3*, *ZmPIP1;4*, *ZmPIP2;2*, *ZmPIP2;4*, *ZmNIP2;1*, *ZmNIP2;2*, *ZmTIP1;1*, *ZmTIP1;2*). A subsequent increase in the duration of drought (up to 42 days) in the PMS "*Poncirus trifoliata* + *F. mosseae*" resulted in the suppression of all

studied aquaporin genes in the roots (Zou et al., 2019). Variations in the expression of the same isoforms of aquaporin may stem not only from the type of plant tissue, but also from the specific features of plant interaction with different AM fungal species (Bárzana et al., 2014; Sharma et al., 2021).

On the other hand, it is possible to assume the existence of other mechanisms for drought resistance in mycorrhizal plants due to the upregulation of the PIP and TIP subfamilies (in this study, upregulation of *PIP1;2*, *PIP1;4*, *PIP2;1*, and *TIP2;3* in the flowering stage was observed). So, according to K. Sharma et al. (2021), the increased activity of PIP and TIP proteins is directly related to the resistance of plants to drought through several pathways: (1) increasing the efficiency of water transport (root hydraulic conductivity); (2) intensification of nutrient metabolism, in particular, nitrogen absorption; (3) enhancing photosynthesis (by the participation of aquaporins in CO_2 diffusion) and increased production of photosynthetic substances (Uehlein et al., 2003; Bárzana et al., 2012); (4) maintaining osmotic pressure and turgor, including through regulation of stomata, improving the aquatic status of plants by increasing water use efficiency (Sharma et al., 2021). It is noteworthy that there are few studies elucidating the reasons for the downregulation of aquaporins under drought in mycorrhizal conditions (Asadollahi et al., 2023), which underlines the relevance of our analysis.

Thus we demonstrate that mycorrhizal symbiosis significantly influenced the water status of the host plant in the PMS "*M. lupulina* + *R. irregularis*" by modulating aquaporins expression under drought conditions.

Conclusion

This study showed that, despite the water deficit conditions, the plant-microbial system "*M. lupulina* + *R. irregularis*" demonstrated high rates of mycorrhization and symbiotic efficiency. In the early stage (day 24, 2nd leaf), a multi-directional regulation was observed: activation of the *NIP3;1*, *NIP4;2*, *TIP2;2* genes, specific expression of *NIP7;1*, and suppression of other genes (*PIP2;3*, *TIP1;1*, *TIP1;4*, *TIP2;3*, *TIP4;1*). The SIP subfamily showed no significant changes. In the late phase (day 48, flowering), a more pronounced downregulation of the expression of most genes was revealed (*NIP1;2*, *NIP1;5*, *NIP2;1*, *NIP4;2*, *PIP2; 3*, *SIP1;3*, *TIP1;1*, *TIP2;2*, *TIP4;1*). The upregulation among the main water transporters was maintained for *PIP1;2*, *PIP1;4*, *PIP2;1*, *TIP2;3*. Two genes (*NIP7;1*, *XIP1;1*) remained active, which may indicate their specific role in the symbiosis. However, the *TIP2;2*, *PIP1;2*, *PIP2;1* and *NIP1;5* genes cannot be markers for effective AM development under drought due to the fact that they have similar regulatory patterns under well-watered conditions based on a comparative analysis with previously published transcriptomic profiles of *M. lupulina* leaves.

Functionally, the TIP subfamily genes, which encoded proteins that regulate water transport across the vacuolar membrane, showed the most significant downregulation, both at an early and late stages. They are probably involved in osmoregulation during water stress. The NIP genes are presumably involved in the processes of interaction with the symbiotic fungus, as evidenced by their selective activation. The role of the genes of the XIP subfamily in the response to drought during mycorrhization has yet to be studied. Our findings expand the current understanding of the mechanisms of adaptation of mycorrhizal plants to drought, in particular: (1) confirm the key role of aquaporins of the TIP subfamily in the regulation of water balance; (2) indicate the possible specificity of NIP genes for mycorrhizal symbiosis; (3) demonstrate the phase-dependent nature of regulation of aquaporin expression, which emphasizes the complexity of their interaction with the symbiont fungus. Thus, the study contributes to the understanding of the molecular basis of drought tolerance in mycorrhizal plants and opens new perspectives for further analysis of the role of individual aquaporin genes in symbiotic systems.

References

Asadollahi M., Iranbakhsh A., Ahmadvand R., Ebadi M., Mehregan I. Synergetic effect of water deficit and arbuscular mycorrhizal symbiosis on the expression of aquaporins in wheat (*Triticum aestivum* L.) roots: insights from NGS RNA-sequencing. *Physiol Mol Biol Plants.* 2023;29(2):195-208. doi 10.1007/s12298-023-01285-w

Bárzana G., Aroca R., Paz J.A., Chaumont F., Martinez-Ballesta M.C., Carvajal M., Ruiz-Lozano J.M. Arbuscular mycorrhizal symbiosis increases relative apoplastic water flow in roots of the host plant under both well-watered and drought stress conditions. *Ann Bot.* 2012;109(5):1009-1017. doi 10.1093/aob/mcs007

Bárzana G., Aroca R., Bienert G.P., Chaumont F., Ruiz-Lozano J.M. New insights into the regulation of aquaporins by the arbuscular mycorrhizal symbiosis in maize plants under drought stress and possible implications for plant performance. *Mol Plant Microbe Interact.* 2014;27(4):349-363. doi 10.1094/MPMI-09-13-0268-R

Daneliia G.V., Yemelyanov V.V., Shishova M.F. The role of water-transporting aquaporins of the PIP and TIP subfamilies in plant development and adaptation to stress factors. *Ecological Genetics*. 2024;22(4):343-368. doi 10.17816/ecogen637037

Dean R.M., Rivers R.L., Zeidel M.L., Roberts D.M. Purification and functional reconstitution of soybean nodulin 26. An aquaporin with water and glycerol transport properties. *Biochemistry*. 1999;38(1): 347-353. doi 10.1021/bi982110c

He F., Zhang H., Tang M. Aquaporin gene expression and physiological responses of *Robinia pseudoacacia* L. to the mycorrhizal fungus *Rhizophagus irregularis* and drought stress. *Mycorrhiza.* 2016; 26(4):311-323. doi 10.1007/s00572-015-0670-3

He J.-D., Dong T., Wu H.-H., Zou Y.-N., Wu Q.-S., Kuča K. Mycorrhizas induce diverse responses of root *TIP* aquaporin gene expression to drought stress in trifoliate orange. *Sci Hortic*. 2019;243:64-69. doi 10.1016/j.scienta.2018.08.010

Heymann J.B., Engel A. Aquaporins: phylogeny, structure, and physiology of water channels. *News Physiol Sci*. 1999;14:187-193. doi 10.1152/physiologyonline.1999.14.5.187

Hussain A., Tanveer R., Mustafa G., Farooq M., Amin I., Mansoor S. Comparative phylogenetic analysis of aquaporins provides insight into the gene family expansion and evolution in plants and their role in drought tolerant and susceptible chickpea cultivars. *Genomics*. 2020;112(1):263-275. doi 10.1016/j.ygeno.2019.02.005

Jia Y., Liu X. Polyploidization and pseudogenization in allotetraploid frog *Xenopus laevis* promote the evolution of aquaporin family in higher vertebrates. *BMC Genomics*. 2020;21(1):525. doi 10.1186/s12864-020-06942-y

Kryukov A.A., Gorbunova A.O., Kudriashova T.R., Ivanchenko O.B., Shishova M.F., Yurkov A.P. SWEET transporters of *Medicago lupulina* in the arbuscular-mycorrhizal system in the presence of medium

level of available phosphorus. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2023;27(3):189-196. doi 10.18699/VJGB-23-25

Kryukov A.A., Kudriashova T.R., Belyaeva A.I., Gorenkova A.I., Yurkov A.P. Effect of *Rhizophagus irregularis* inoculation on aquaporin gene expression in the roots of *Medicago lupulina* in drought conditions. *Ecological Genetics*. 2025;23(3):263-275. doi 10.17816/ecogen643544 (in Russian)

Kruse E., Uehlein N., Kaldenhoff R. The aquaporins. *Genome Biol.* 2006;7(2):206. doi 10.1186/gb-2006-7-2-206

Kudoyarova G., Veselov D., Yemelyanov V., Shishova M. The role of aquaporins in plant growth under conditions of oxygen deficiency. *Int J Mol Sci.* 2022;23(17):10159. doi 10.3390/ijms231710159

Kudryashova T.R., Kryukov A.A., Gorbunova A.O., Ivanchenko O.B., Yurkov A.P. SWEET genes of *Medicago lupulina* during the development of symbiosis with arbuscular mycorrhiza under conditions of low phosphorus levels. *Butlerov Communications.* 2024;78(5):119-127. doi 10.37952/ROIjbc-01/24-78-5-119/ROI-jbc-C/24-7-2-8 (in Russian)

Kudriashova T.R., Kryukov A.A., Gorenkova A.I., Yurkov A.P. Aquaporins and their role in plant-microbial systems. *Vavilov Journal of Genetics and Breeding = Vavilov J Genet Breed.* 2025;29(2):238-247. doi 10.18699/vjgb-25-27

Lopez D., Amira M.B., Brown D., Muries B., Brunel-Michac N., Bourgerie S., Porcheron B., … Fumanal B., Label P., Pujade-Renaud V., Auguin D., Venisse J.-S. The *Hevea brasiliensis* XIP aquaporin subfamily: genomic, structural and functional characterizations with relevance to intensive latex harvesting. *Plant Mol Biol.* 2016;91(4-5):375-396. doi 10.1007/s11103-016-0462-y

Luo Y., Ma L., Du W., Yan S., Wang Z., Pang Y. Identification and characterization of salt- and drought-responsive AQP family genes in *Medicago sativa* L. *Int J Mol Sci*. 2022;23(6):3342. doi 10.3390/ijms23063342

Ma J.F., Tamai K., Yamaji N., Mitani N., Konishi S., Katsuhara M., Ishiguro M., Murata Y., Yano M. A silicon transporter in rice. *Nature.* 2006;440(7084):688-691. doi 10.1038/nature04590

MacRae E. Extraction of plant RNA. In: Hilario E., Mackay J. (Eds) Protocols for Nucleic Acid Analysis by Nonradioactive Probes. Methods in Molecular Biology. Vol. 353. Humana Press, 2007;15-24. doi 10.1385/1-59745-229-7:15

Mammadov J., Buyyarapu R., Guttikonda S., Parliament K., Abdurakhmonov I.Y., Kumpatla S.P. Wild relatives of maize, rice, cotton, and soybean: treasure troves for tolerance to biotic and abiotic stresses. *Front Plant Sci.* 2018;9:886. doi 10.3389/fpls.2018.00886

Maurel C., Verdoucq L., Luu D.T., Santoni V. Plant aquaporins: membrane channels with multiple integrated functions. *Annu Rev Plant Biol*. 2008;59(1):595-624. doi 10.1146/annurev.arplant.59.032607.092734

Maurel C., Boursiac Y., Luu D.T., Santoni V., Shahzad Z., Verdoucq L. Aquaporins in plants. *Physiol Rev*. 2015;95(4):1321-1358. doi 10.1152/physrev.00008.2015

Min X., Wu H., Zhang Z., Wei X., Jin X., Ndayambaza B., Wang Y., Liu W. Genome-wide identification and characterization of the aquaporin gene family in *Medicago truncatula*. *J Plant Biochem Biotechnol*. 2019;28(3):320-335. doi 10.1007/s13562-018-0484-4

Mizutani M., Watanabe S., Nakagawa T., Maeshima M. Aquaporin *NIP2;1* is mainly localized to the ER membrane and shows root-specific accumulation in *Arabidopsis thaliana*. *Plant Cell Physiol*. 2006;47(10):1420-1426. doi 10.1093/pcp/pcl004

Murata K., Mitsuoka K., Hirai T., Walz T., Agre P., Heymann J.B., Engel A., Fujiyoshi Y. Structural determinants of water permeation through aquaporin-1. *Nature*. 2000;407(6804):599-605. doi 10.1038/35036519

Noronha H., Araújo D., Conde C., Martins A.P., Soveral G., Chaumont F., Delrot S., Gerós H. The grapevine uncharacterized intrinsic protein 1 (VvXIP1) is regulated by drought stress and transports glycerol, hydrogen peroxide, heavy metals but not water. *PLoS One.* 2016;11(8):e0160976. doi 10.1371/journal.pone.0160976

Phillips J.M., Hayman D.S. Improved procedures for clearing roots and staining parasitic and vesicular-arbuscular mycorrhizal fungi for rapid assessment of infection. *Trans Br Mycol Soc.* 1970;55(1):158-161. doi 10.1016/S0007-1536(70)80110-3

Porcel R., Gómez M., Kaldenhoff R., Ruiz-Lozano J.M. Impairment of *NtAQP1* gene expression in tobacco plants does not affect root colonization pattern by arbuscular mycorrhizal fungi but decreases their symbiotic efficiency under drought. *Mycorrhiza.* 2005;15:417-423. doi 10.1007/s00572-005-0346-5

Porcel R., Aroca R., Azcon R., Ruiz-Lozano J.M. PIP aquaporin gene expression in arbuscular mycorrhizal *Glycine max* and *Lactuca sativa* plants in relation to drought stress tolerance. *Plant Mol Biol*. 2006;60:389-404. doi 10.1007/s11103-005-4210-y

Quiroga G., Erice G., Aroca R., Chaumont F., Ruiz-Lozano J.M. Contribution of the arbuscular mycorrhizal symbiosis to the regulation of radial root water transport in maize plants under water deficit. *Environ Exp Bot*. 2019;167:103821. doi 10.1016/j.envexpbot.2019.103821

Remy W., Taylor T.N., Hass H., Kerp H. Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc Natl Acad Sci USA*. 1994;91(25):11841-11843. doi 10.1073/pnas.91.25.11841

Ruiz-Lozano J.M., Aroca R. Host response to osmotic stresses: stomatal behaviour and water use efficiency of arbuscular mycorrhizal plants. In: Koltai H., Kapulnik Y. (Eds) Arbuscular Mycorrhizas: Physiology and Function. Dordrecht: Springer, 2010;239-256. doi 10.1007/978-90-481-9489-6_11

Sandal N.N., Marcker K.A. Soybean nodulin 26 is homologous to the major intrinsic protein of the bovine lens fiber membrane. *Nucleic Acids Res*. 1988;16:9347-9348. doi 10.1093/nar/16.19.9347

Sharma K., Gupta S., Thokchom S.D., Jangir P., Kapoor R. Arbuscular mycorrhiza-mediated regulation of polyamines and aquaporins during abiotic stress: deep insights on the recondite players. *Front Plant Sci*. 2021;12:642101. doi 10.3389/fpls.2021.642101

Singh S., Bhatt V., Kumar V., Kumawat S., Khatri P., Singla P., Shivaraj S.M., Nadaf A., Deshmukh R., Sharma T.R., Sonah H. Evolutionary understanding of aquaporin transport system in the basal eudicot model species *Aquilegia coerulea*. *Plants*. 2020;9(6):799. doi 10.3390/plants9060799

Smith S.E., Read D.J. (Eds) Mycorrhizal Symbiosis. San Diego, CA: Academic Press, 2008. doi 10.1016/B978-0-12-370526-6.X5001-6

Trouvelot A., Kough J.L., Gianinazzi S. Mesure du taux de mycorhization VA d'un système radiculaire. Recherche de méthodes ayant une signification fonctionnelle. In: Gianinazzi-Pearson V., Gianinazzi S. (Eds) Physiological and Genetical Aspects of Mycorrhizae. Paris: INRA, 1986;217-221

Uehlein N., Lovisolo C., Siefritz F., Kaldenhoff R. The tobacco aquaporin NtAQP1 is a membrane CO_2 pore with physiological functions. *Nature.* 2003;425(6959):734-737. doi 10.1038/nature02027

Vorobyev N.I., Yurkov A.P., Provorov N.A. Certificate N2010612112 about the Registration of the Computer Program "Program for Calculating the Mycorrhization Indices of Plant Roots" (Dated December 2, 2016). Moscow: The Federal Service for Intellectual Property, 2016 (in Russian)

Wang D., Ni Y., Xie K., Li Y., Wu W., Shan H., Cheng B., Li X. Aquaporin *ZmTIP2;3* promotes drought resistance of maize through symbiosis with arbuscular mycorrhizal fungi. *Int J Mol Sci*. 2024;25(8):4205. doi 10.3390/ijms25084205

Yurkov A.P., Jacobi L.M., Gapeeva N.E., Stepanova G.V., Shishova M.F. Development of arbuscular mycorrhiza in highly responsive and mycotrophic host plant – black medick (*Medicago lupulina* L.). *Russ J Dev Biol*. 2015;46(5):263-275. doi 10.1134/S1062360415050082

Yurkov A., Kryukov A., Gorbunova A., Sherbakov A., Dobryakova K., Mikhaylova Y., Afonin A., Shishova M. AM-induced alteration in the expression of genes, encoding phosphorus transporters and enzymes of carbohydrate metabolism in *Medicago lupulina. Plants.* 2020;9(4):486. doi 10.3390/plants9040486

Yurkov A., Puzanskiy R., Avdeeva G., Jacobi L., Gorbunova A., Kryukov A., Kozhemyakov A., … Yemelyanov V., Shavarda A., Kir-

pichnikova A., Smolikova G., Shishova M. Mycorrhiza-induced alterations in metabolome of *Medicago lupulina* leaves during symbiosis development. *Plants*. 2021;10(11):2506. doi 10.3390/plants10112506

Yurkov A.P., Afonin A.M., Kryukov A.A., Gorbunova A.O., Kudryashova T.R., Kovalchuk A.I., Gorenkova A.I., … Zhukov V.A., Puzanskiy R.K., Mikhailova Y.V., Yemelyanov V.V., Shishova M.F. The effects of *Rhizophagus irregularis* inoculation on transcriptome of *Medicago lupulina* leaves at early vegetative and flowering stages of plant development. *Plants.* 2023;12:3580. doi 10.3390/plants12203580

Zhou X., Yi D., Ma L., Wang X. Genome-wide analysis and expression of the aquaporin gene family in *Avena sativa* L. *Front Plant Sci*. 2024;14:1305299. doi 10.3389/fpls.2023.1305299

Zou Y.N., Wu H.H., Giri B., Wu Q.S., Kuča K. Mycorrhizal symbiosis down-regulates or does not change root aquaporin expression in trifoliate orange under drought stress. *Plant Physiol Biochem*. 2019; 144:292-299. doi 10.1016/j.plaphy.2019

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

Received August 31, 2025. Revised November 27, 2025. Accepted December 3, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):435-443

doi 10.18699/vjgb-26-47

A production strain of soybean nodule bacteria RZ300 *Bradyrhizobium japonicum* resistant to drying on the seed surface: cultural properties and genomic features

Yu.V. Kosulnikov, A.A. Kryukov, K.N. Berdysheva, A.I. Kovalchuk, A.P. Yurkov, Yu.V. Laktionov

All-Russian Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia
laktionov@arriam.ru

Abstract. Pre-sowing treatment of cultivated legume seeds with nodule bacteria preparations is a standard agronomic practice. This is particularly important in soybean cultivation, as effective microsymbionts of soybeans are often absent from the soil. However, as many studies have shown, the efficacy of biopreparations depends largely on the survival of rhizobial cells on seeds during drying. In this study, we analyzed the viability of three production strains of *Bradyrhizobium japonicum* Kirchner (634b, 640 and RZ300) on soybean (*Glycine max* L.) seeds of various origins (varieties: EN Argenta, Bara and Prudence). The experiments evaluated several parameters: inoculant concentrations (10 and 100 %), drying temperatures (5, 15, and 25 °C), and protective polymer-carbohydrate formulations. The experiments revealed that the soybean variety had no noticeable effect on the viability of the studied rhizobial strains, while the strains themselves differed significantly in this regard. The RZ300 strain demonstrated the highest resistance to drying on soybean seeds. A comparative genomic analysis of this strain and the less resistant *B. japonicum* strain 634b revealed the presence of the *opg*C gene in the RZ300 strain (encodes the OpgC protein involved in the biosynthesis of osmoregulated periplasmic glucans (OPGs)). This gene is absent in strain 634b and may potentially determine the increased resistance of nodule bacteria to drying on seeds. An evaluation of various protective formulations demonstrated that formulations based on 50 % sucrose provide the best protection, with rhizobia showing the highest resistance to drying at +5 °C. The results obtained in this study can be used both in the selection of effective inoculant strains and for providing technological support in the development of biological products. The genomic data support the development of genetic screening systems to identify promising strains and the potential introduction of the *opg*C gene into promising rhizobial strains to improve their manufacturability, i. e. to enable effective early seed inoculation.

Key words: nodule bacteria; soy *Glycine max* L.; *Bradyrhizobium japonicum*; microbial biologics; osmotic stress; genome-wide sequencing; osmotic stress resistance genes

For citation: Kosulnikov Yu.V., Kryukov A.A., Berdysheva K.N., Kovalchuk A.I., Yurkov A.P., Laktionov Yu.V. A production strain of soybean nodule bacteria RZ300 *Bradyrhizobium japonicum* resistant to drying on the seed surface: cultural properties and genomic features. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):435-443. doi 10.18699/vjgb-26-47

Funding. The work was supported by Russian Science Foundation No. 25-26-00242.

Acknowledgements. The work is carried out using the equipment of the Center for Collective Use "Genomic Technologies, Proteomics and Cell Biology" at the ARRIAM.

Производственный штамм клубеньковых бактерий сои RZ300 *Bradyrhizobium japonicum*, устойчивый к высыханию на поверхности семян: культуральные свойства и специфические особенности генома

Ю.В. Косульников, А.А. Крюков, К.Н. Бердышева, А.И. Ковальчук, А.П. Юрков, Ю.В. Лактионов

Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии, Пушкин, Санкт-Петербург, Россия
laktionov@arriam.ru

Аннотация. Предпосевная обработка семян возделываемых бобовых культур препаратами клубеньковых бактерий является стандартной агрономической практикой, активно используемой при выращивании сои, эффективные микросимбионты которой часто отсутствуют в почве. В то же время, как показали многие

исследования, эффект от применения биопрепарата во многом зависит от выживаемости ризобиальных клеток на семенах при высыхании. В настоящей работе проведен анализ жизнеспособности у трех производственных штаммов *Bradyrhizobium japonicum*, 643б, 640 и RZ300, на семенах сои *Glycine max* L. разного происхождения (сорта ЭН Аргента, Бара и Пруденс) в опытах, различающихся по разным параметрам: концентрациям инокулянта (10 и 100 %), температурам высушивания (5, 15, 25 °С) и составам защитных полимерно-углеводных композиций. В результате экспериментов обнаружено, что сорт сои не оказывал заметного влияния на жизнеспособность изучаемых штаммов ризобий, в то время как штаммы существенно различались по этому признаку. Наибольшую устойчивость к высыханию на семенах сои показал штамм RZ300. Сравнительный анализ генома этого штамма с геномом слабо устойчивого к высыханию штамма *B. japonicum* 634б позволил выявить наличие у штамма RZ300 гена *opg*C (кодирует белок OpgC, участвующий в биосинтезе осморегулируемых периплазматических глюканов (OPGs)), который отсутствует у штамма 634б и, возможно, может определять повышенную устойчивость клубеньковых бактерий к высыханию на семенах. При изучении эффекта различных защитных композиций было отмечено, что лучшими защитными свойствами обладают составы на основе 50 % раствора сахарозы, а наибольшая устойчивость ризобий к высыханию проявляется при температуре +5 °С. Полученные в этой работе результаты могут быть использованы как в селекции эффективных штаммов-инокулянтов, так и в технологическом сопровождении при создании биопрепаратов. Данные, полученные при изучении геномов штаммов, представляют интерес как для разработки систем генетического скрининга при поиске перспективных штаммов, так и для изучения возможности введения генетических конструкций с геном *opg*C в перспективные штаммы ризобий для улучшения их технологичности, т.е. обеспечения возможности эффективной заблаговременной инокуляции семян.

Ключевые слова: клубеньковые бактерии; соя *Glycine max* L.; *Bradyrhizobium japonicum*; микробные биопрепараты; осмотический стресс; полногеномное секвенирование; гены устойчивости к осмотическому стрессу

Introduction

An important agrobiological feature of soybeans is the ability to form a nitrogen-fixing legume-rhizobial symbiosis with nodule bacteria (Vavilov, Posypanov, 1983; Regar et al., 2017). At the same time, an effective symbiosis only occurs when active virulent soybean symbiont bacteria are present in the soil in sufficient quantities, which is rarely the case under field conditions and, thus, reduces the legume yield (Lamptey et al., 2014). To fully realize the potential of legume-rhizobial symbiosis, it is necessary to artificially introduce symbiotically effective nodule bacteria strains into the rhizosphere, which in agricultural practice is achieved through the pre-sowing inoculation of soybean seeds with rhizobial-based biopreparations. A biopreparation based on the *Bradyrhizobium japonicum* 634b strain under the trade name Rhizotorphin is widely used in domestic agriculture and provides high yield increases across various soybean varieties (Vasilchikov, Gurev, 2018; Volobueva et al., 2023). Strain 634b is frequently used as a reference for assessing new rhizobial candidates, including strain 640 (Magomedov et al., 2011).

When analyzing the effectiveness of rhizobial strains, seed inoculation is carried out on the day of sowing, as it is known that nodule bacteria are sensitive to drying. However, in agricultural practice, for both technological and economic reasons, pre-sowing seed treatment is carried out well in advance, which can lead to the death of the applied rhizobia even before the seeds are planted. A review of this problem was provided by J. Vriezen et al. (2007). As early as 1932, E.B. Fred et al. reported a decrease in the number of viable nodule bacteria cells on seeds and suggested that the nutrient medium composition, pH and temperature are factors determining the resistance of cells to desiccation.

Later, J.M. Vincent et al. (1961) studied the culture of *Rhizobium trifolii* during its drying on glass beads and proposed that the decrease in the number of viable rhizobia was due to both "seed factors" and the drying factor itself. It was shown that the negative effect of drying could be partially offset by the addition of saccharides, particularly maltose, which indicates that the availability of nutrients, and potentially other dissolved substances, affects the survival of bacterial cells.

Several studies have shown the difference in the survival dynamics for nodule bacteria on various matrices such as glass beads, seeds, soil, nitrocellulose filters, etc. (Vriezen et al., 2006). One possible reason for the observed differences is that dry inoculated seeds have a water activity ranging from 0.45 to 0.6 (Smith, 1992) and, thus, still contain a relatively large amount of water compared with completely dry surfaces of glass beads or nitrocellulose filters.

From a practical point of view, given the high sensitivity of nodule bacteria to drying, it is important either to sow the treated seeds on the day of treatment (Vasilchikov, Gurev, 2018), or to use special polymer and carbohydrate protectors that increase the resistance of bacteria to osmotic stress (Skorupska et al., 2006; Deaker et al., 2007; Reina-Bueno et al., 2012). In particular, we have previously shown that the water-soluble polymer polyvinylpyrrolidone combined with activated carbon significantly improves rhizobial survival on inoculated seeds. This combination is more ef-

fective that polyvinylpyrrolidone alone and it reduces bacterial mortality on inoculated seeds by 20–30 % after the first 5–7 days of seed storage (Laktionov et al., 2019).

Finally, it should be noted that studies on the survival of rhizobia during pre-sowing inoculation should become an important part in the selection of effective inoculant strains. A key element of such research should be the investigation of the mechanisms of rhizobia resistance to desiccation. While a number of studies have explored the molecular and genetic aspects of osmotic resistance in nodule bacteria (Vriezen et al., 2007), especially in the context of climate change (Zhang et al., 2024), relatively few studies have evaluated the viability of rhizobial collection strains on legume seeds or explored methods to enhance the resistance of cells to osmotic stress (Laktionov et al., 2019).

The aim of this study was to evaluate the resistance of three production strains of soybean nodule bacteria: strain 634b (used to produce inoculants for Risotorphin), strain 640 (identified in several studies as more effective than strain 634b) and a new promising strain RZ300. The evaluation was conducted on various soybean varieties under different storage temperatures and using protective polymer-carbon formulations of varying compositions. Furthermore, the study aimed to elucidate the mechanisms underlying the desiccation resistance of these rhizobial strains using comparative genomics.

Materials and methods

The study used *B. japonicum* strains 634b, 640 and RZ300 from the network bioresource collection in the field of genetic technologies for agriculture (Russian Collection of Agricultural Microorganisms RCAM).

Strain 634b was isolated from four soybean (cv. Kolkhida 4) nodules at the Natshakai Station in Georgia. This strain is used to produce inoculants for soy under the trademark Risotorphin (Vasilchikov, Gurev, 2018; Volobueva et al., 2023). Strain 640 was isolated in 1976 from soybean (cv. Smena) nodules in the meadow-chernozem soil of the Amur region at the All-Russian Soybean Research Institute. A number of studies have identified the strain as more effective than 634b (Magomedov et al., 2011). The RZ300 strain was isolated in 2022 in the Krasnodar Territory from soybean (cv. Bara) nodules and has hardly been studied to date. Strain RZ300 exhibits stable nodulation in both chamber and field experiments under normal and extreme conditions (including presence of chemical pesticides, and lack of moisture). Patent No. 2806593 "Method of cultivation of nodule bacteria of soy *Bradyrhizobium japonicum* RZ300" was obtained for strain RZ300.

B. japonicum strains 634b, 640 and RZ300 were grown for inoculation in a liquid semi-synthetic medium (Yadav et al., 2011): Mannit – 10.0 g/L; yeast extract – 1 g/L; K_2HPO_4 – 0.5 g/L; $MgSO_4 \cdot 7H_2O$ – 0.2 g/L; NaCl – 0.1 g/L; $CaCO_3$ – traces. Cultures were grown for 7 days at 28 °C in 250 mL glass flasks with cotton stoppers on an orbital shaker at 180 rpm. After cultivation, the flasks were stored at +5 °C. To obtain experimental liquid cultures, a 2 % (v/v) inoculum from the flasks was aseptically transferred to a BIORUS laboratory fermenter using the same medium composition (Yadav et al., 2011).

The microorganisms were cultured to the stationary phase, in which the cells are most resistant to osmotic stress (Soria et al., 2006), in a periodic regime for 7 days at 28 °C with mechanical mixing (150 rpm) and aeration (1 L air/1 L medium per min). The resulting experimental preparations were aseptically poured into sterilized glass flasks with cotton stoppers and stored in a refrigerator. To determine the titer of the resulting bacterial suspensions, a series of consecutive 10-fold dilutions were prepared, followed by inoculation on Petri dishes with an agar medium (Yadav et al., 2011); the number of colonies was counted after 10 days of incubation at 28 °C.

To study the resistance of strains to drying, glass beads and *Glycine max* L. soybean seeds from three varieties of different origin – EN Argenta (Deriglazova, Morozov, 2022), Bara (Parakhin et al., 2017), Prudence (Bobkova, 2020) – were used.

The dynamics of nodule bacteria viability during drying were monitored for 24 hours following inoculation with rhizobium preparations. To determine the effect of the "seed factor", glass beads of a similar size to the seeds were used. The preparations were applied in concentrations of 10 % (aqueous solution) and 100 % (undiluted inoculant) at the rate of 10 L per 1 ton of seeds, following standard agricultural practice (Kincharova, Matvienko, 2021). An undiluted preparation was applied to assess the effect of exopolysaccharides and residual media, as well as a high cell density. The treatment of seeds and beads with tank solutions followed by an assessment of the dynamics of bacterial viability over time was carried out according to the author's methodology (Laktionov et al., 2019). Experiments were conducted in triplicate.

To evaluate the effectiveness of polymer-carbohydrate compositions as bacterial protectors on seeds drying at various temperatures, the following formulations were prepared: 1 % solution of carbokimethylcellulose (CMC); 50 % solution of sucrose; a mixture of 1 % solution of carbokimethylcellulose (CMC), 50 % solution of sucrose; a mixture of 1 % solution of carbokimethyl-

cellulose (CMC), 50 % solution of sucrose and 1 % activated carbon.

The purity of microbial cultures was determined by both morphological characteristics (morphology of colonies on plates and rhizobium cells in a fixed smear) and molecular genetic identification (16S rRNA gene sequence analysis).

Rhizobium DNA was isolated by the CTAB method (Doyle J.J., Doyle J.L., 1987, 1990). Whole-genome sequencing for strains RZ300 (BioProject PRJNA1266151) and 634b (BioProject PRJNA1334995) was performed using 3rd generation sequencing (Oxford Nanopore Technologies, UK). Genome assembly and annotation were conducted using bioinformatic tools including: flye, SPAdes, mauveAligner, GeneMark, Prokka, and EggNOG-mapper. Annotated whole-genomic sequences were deposited GenBank.

Statistical analysis was performed using the SciPy library, and data visualization was performed using the matplotlib library in Python.

Results

At the first stage of the research, the cell titers of the bacterial suspensions were determined at the time of the working solution preparation. The titers were: $1.8 \pm 0.24 \cdot 10^9$ CFU/mL for strain 634b; $1.77 \pm 0.29 \cdot 10^9$ CFU/mL for strain 640; $2.0 \pm 0.19 \cdot 10^9$ CFU/mL for strain RZ300.

To evaluate the specific effects of drying on the survival of the studied strains, rhizobial viability on soybean seeds (cv. EN Argenta) and glass beads was compared (the results are shown in Figure 1). It has been shown that viable cells on glass beads remained at levels of several thousand CFU per bead only for three hours after inoculation and almost completely died within a day after treatment. In contrast, tens of thousands of viable cells remained on the seeds a day after treatment. At the same time, the differences in cell viability between different strains were not statistically significant, the calculated F-statistic (2.67) was less than the critical F-value (3.89) at a 5 % significance level.

In the second stage of the research, the potential influence of three soybean varieties on the dynamics of nodule bacteria viability was studied without the addition of protective formulations (Fig. 2). When seeds were treated with a 10 % inoculant solution at a rate of 10 L/t, 168 hours post-inoculation, the cells of the RZ300 strain remained on soybean seeds of the EN Argenta, Bara and Prudence varieties in the amount of 16, 14 and 21 thousand CFU per seed, respectively. At the same time, for strains 634b and 640, these values did not exceed 3, 6 and 2 thousand CFU per 1 seed, respectively.

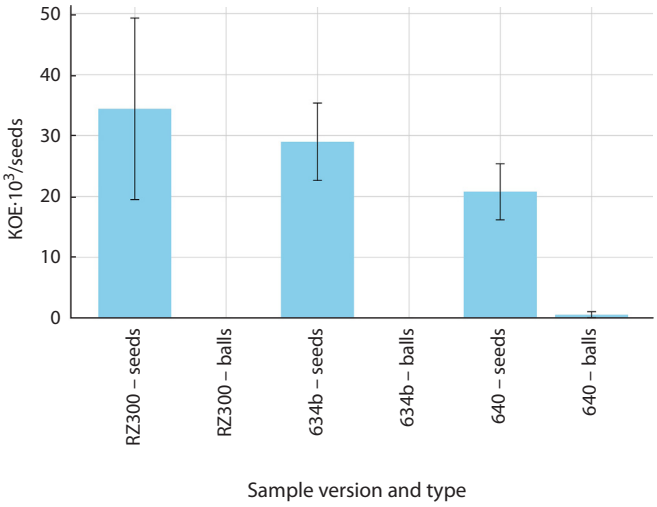


Fig. 1. Cell viability of *B. japonicum* strains on EN Argenta soybean seeds and glass beads 24 hours after inoculation.

Viability analysis of the three *B. japonicum* strains across different soybean varieties 168 hours after treatment showed that for the EN Argenta variety, the number of viable cells of the RZ300 strain ($\overline{X}$ = 15.67 thousand CFU/seed) was significantly higher (LSD = 5.19, $p < 0.05$) than that of strains 634b ($\overline{X}$ = 2.00 thousand CFU/seed) and 640 ($\overline{X}$ = 2.67 thousand CFU/seed). Similar patterns were observed for the Bara and Prudence varieties. The RZ300 strain demonstrated statistically higher cell viability in the Bara ($\overline{X}$ = 14.00 thousand CFU/seed; LSD = 4.85) and Prudence ($\overline{X}$ = 21.33 thousand CFU/seed; LSD = 3.20) varieties compared to strains 634b and 640, between which no significant differences were found. After 336 hours, the number of bacteria of the RZ300 strain on seeds of all varieties was estimated at thousands of CFU/seed, while viable cells of strains 634b and 640 could not be isolated from seeds. A similar pattern was observed for the 100 % inoculant solutions, accounting for the higher initial cell densities.

In the third stage of the study, the effect of various protective formulations and temperature conditions on cell viability of the RZ300 strain, which demonstrated the best survival rates in inoculated seeds, was studied. Figure 3 shows that the rate of cell mortality increased with increasing temperature. At the same time, carboxymethylcellulose was a relatively ineffective cell protector, while sucrose and sucrose-based compositions significantly increased the resistance of bacteria to drying on seeds. At +5 °C, the temperature exhibiting the least contrast between treatment means, it was shown that the sucrose variants provided significantly higher cell viability (at 5 % significance level) compared with

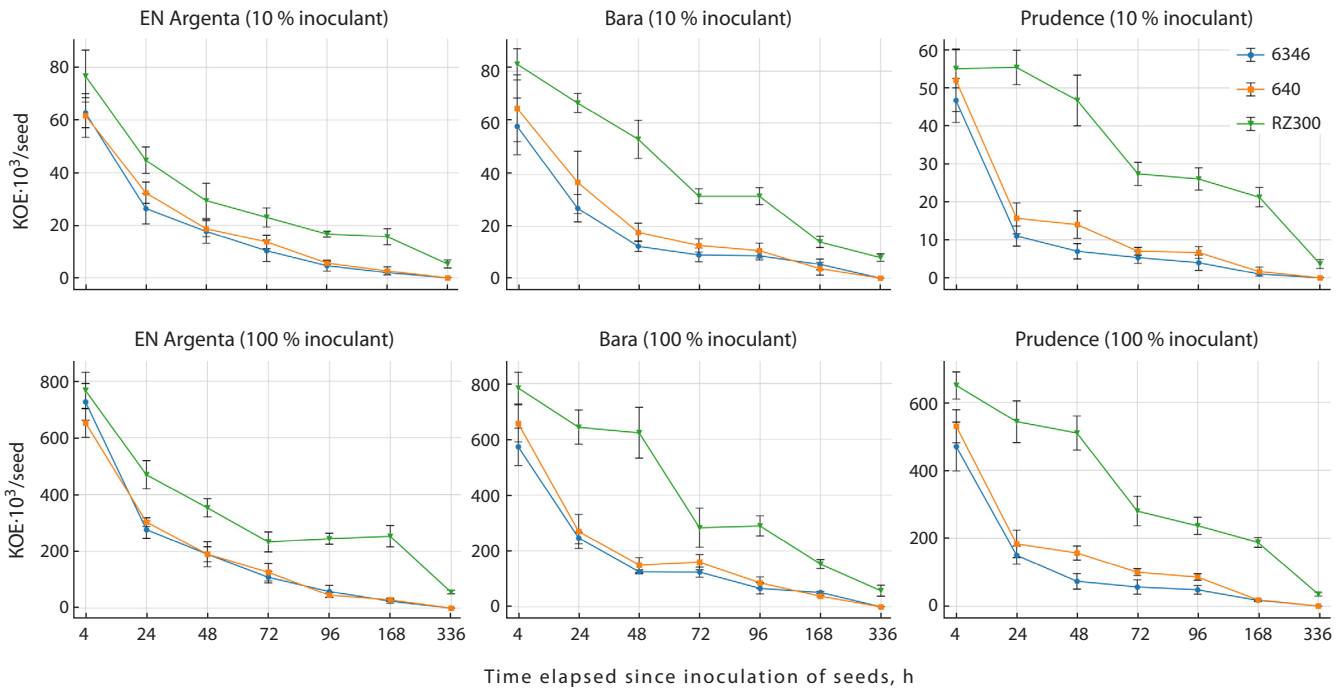


Fig. 2. Survival dynamics of *B. japonicum* strains on soybean seeds (cv. EN Argenta, Bara, Prudence), inoculated with 10 and 100 % solutions of preparations without the addition of protective formulations.

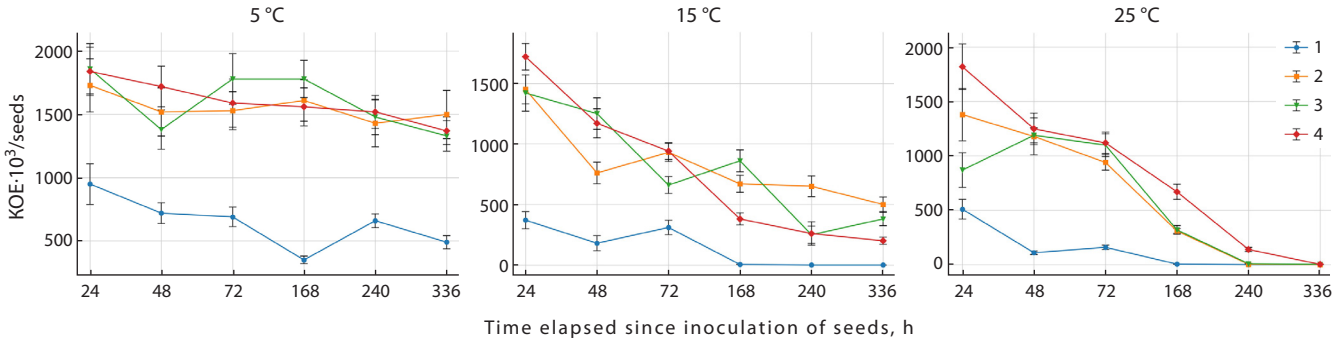


Fig. 3. Dynamics of the viability of RZ300 cells on soybean seeds of the EN Argenta variety during two weeks of storage at different temperatures (+5, +15, +25 °C) and with various protective formulations (1 – 1 % CMC; 2 – 50 % sucrose, 3 – 1 % CMC + 50 % sucrose; 4 – 1 % CMC + 50 % sucrose + 0.5 % activated carbon).

1 % carboxymethylcellulose, while the differences among various sucrose-based formulations were not significant. Thus, it can be assumed that sucrose has the greatest protective effect for bacteria undergoing drying on seeds. Consequently, sucrose-based compositions can further enhance the resistance of the RZ300 strain to drying on seeds.

Finally, to identify the genomic features that may be responsible for resistance of strain RZ300 to desiccation, a comparative analysis of its whole genome was performed against the genome of the less resistant strain 634b.

The results of whole-genome sequencing for the RZ300 strain have been deposited in NCBI (BioProject PRJNA1266151). The size of the circular chromosome is 9,199,961 bp. The total number of genes is 9,643, of which there are 5,736 coding ones, and a significant part is identified as pseudogenes – 3,842. The results of whole-genome sequencing for strain 634b have also been deposited in NCBI (BioProject PRJNA1334995). The genomes of the RZ300 and 634b strains were compared using the RAST (Rapid Annotation using Subsystem Technology) server. The Table shows the differentially represented genes specific to each studied strain. The genes putatively associated with the increased stress resistance of the RZ300 strain are highlighted in bold.

Comparative analysis of functional genes groups in drying-resistant RZ300 and drying-sensitive 634b of *B. japonicum* strains

Functional group of genes	Represented in the	
	RZ300 strain	634b strain
Stress response	***opg*C**, ***opg*G2**, *bph*P, ***hfl*X**	*LTC*4S, *kat*A, *teh*A
Carbohydrates	*nag*E, *nag*B, *nag*A, *hat*A, *liv*K, *mal*F, *thi*H, *hbd*, *bdh*, *rib*B, *gadh*, *gcd*C, *gcd*, *deo*C	*bkd*A, *bkd*B, *bkd*C, *bkd*D, *act*P, *hpr*, *pps*A, *mal*K, *fix*X, *xyl*R
Subsystem-based clustering	*tse*T, *rny*, *fix*O, *trm*E, ***dgk*A**	*aep*A, *rtx*C, *min*E, *rad*50, *mre*11
DNA metabolism	*vsr*	*yie*H, *yrr*C, *rep*, *pcr*A, *hup*
RNA metabolism	*rpo*E, *muc*D	*rpo*H3, *rnp*A, *mbl*
Nitrogen metabolism	*nar*B, *nar*K, *ntr*C, *nor*B	*nos*D, *nos*F, *nos*R, *nos*L, *nos*Y, *nos*Z, *cyn*D, *cyn*B, *cyn*A
Protein metabolism	***lgt***, *trn*V	*hypC*, *ere*
Membrane transport	*opp*F, *opp*D, *opp*A, *dpp*C, *liv*H, *liv*M, *liv*G, *liv*F, ***nha*A**	*mgt*A, *trb*B, *trb*C, *trb*D, *trb*E, *trb*J, *lap*B, *lap*C, *lap*E, *flp*A, *hrc*U
Amino acids and derivatives	*pot*A, *pot*I, *urt*A, *ure*J, *tdh*	*prp*A, *atz*A, *arc*A, *arc*D, *odc*, *ant*A, *hmg*
Breath	*cyd*A, *res*A	*hup*V, *hup*U, *hox*X, *hox*T, *hox*A, *hox*J
Fatty acids, lipids, isoprenoids	*hbd*	–
Metabolism of aromatic compounds	*muc*I, *est*A, *hba*P, *mao*A	*bkd*A
Virulence, disease and defense	*lod*B, *lod*A, *czc*D, *cus*A, *czc*A, *ncc*B, *mer*R, *cue*O	–
Sulfur metabolism	*gsh*T, *trk*A	–
Iron metabolism	*ton*B, *fhu*A	–
Cofactors, vitamins, prosthetic groups, pigments	*pdh*A, *btu*B	*bcd*
Phages, prophages	*ter*L	–
Mobility and chemotaxis	*che*A	*fli*S
Various	***cdh*A**, ***gpm*A**	–

Note. Genes presumably related to the increased stress resistance of strain RZ300 are highlighted in bold.

Discussion

In agricultural practice, the productivity potential of legume-rhizobial symbiosis can be realized only if high bacterial survival is maintained on the seeds prior to sowing. At the same time, despite the high practical importance of the issue of nodule bacteria resistance to drying on seeds, this area of research remains poorly studied in both Russian and international scientific literature.

The use of glass beads has shown that the surface of seeds is a more favorable environment for nodule bacteria than the inert surface of glass. This indicates that the evolutionary history of symbiosis, in addition to the well-known functions related to virulence and nitrogen fixation, may have included the development of specific functions related to maintaining the viability of rhizobia on the seed surface, which is obviously of great importance for ensuring the stability of symbiosis.

Notably, our findings indicate that the seed variety does not significantly affect the dynamics of the decrease in the number of cells, whereas bacterial strains differ significantly in their resistance to desiccation. This highlights two points: first, the conservative nature of the plant-side mechanisms that ensure the preservation of rhizobia on the surface of seeds, and second, the potential for

selecting inoculant strains based on resistance to desiccation as a way to increase their manufacturability, that is, integration with existing agrotechnological schemes, which has received very little attention so far.

In this study, we have shown that the selection of effective protective formulations and seed treatment methods significantly contributes to ensuring the survival of rhizobia on inoculated seeds. Thus, for instance, a 10-fold increase in the consumption rate of the inoculant led to a proportional increase in the number of viable cells on the seeds throughout the study period. However, the overall dynamics of cell death remained similar to that when using 10 % solutions. This suggests that there is no pronounced protective effect of exopolysaccharides applied with an increased volume of the inoculant on the osmotic resistance of the strains.

It has been shown that the dynamics of reducing the number of viable bacteria on seeds slows down significantly with decreasing temperature, as well as with the addition of protective polymer-carbohydrate compositions, in particular, based on 50 % sucrose. Thus, the combination of lowering the temperature and adding sucrose during the processing of soybean seeds can ensure a significantly higher number of viable cells per seed at the time of planting.

Comparative analysis of the genomes of strain RZ300 and strain 634b revealed the following differences. According to the group of "stress response" genes, the drying-resistant strain RZ300 has the *opg*C and *opg*G2 genes responsible for the synthesis of osmoregulated periplasmic glucans, which directly explains its osmotic stress tolerance, while the osmosensitive 634b lacks these genes. OpgC is a protein involved in the biosynthesis of osmoregulated periplasmic glucans (OPGs), occurring in a number of bacteria. These glucans play an important role in adaptation to osmotic stress, cell envelope integrity, biofilm formation, and pathogenicity. It is likely that the role of this gene in the formation of biofilms determines the resistance of bacteria to desiccation on seeds. Mutations in the *opg*C gene can lead to changes in OpgC synthesis, which, in turn, affects bacterial survival under osmotic stress (Bontemps-Gall, Lacroix, 2015).

Among other genes that could potentially be involved in the control of stress resistance, the RZ300 strain contains a gene encoding a GTP-binding protein related to HflX. The results of recent studies indicate the role of HflX in determining bacterial resistance to macrolides and lincosamides (Rudra et al., 2020). Another important candidate gene for stress control represented in strain RZ300 is *dgk*A. It is known that the expression of the *dgk*A gene is associated with accelerated bacterial growth and survival in response to adverse environmental factors (Baker et al., 2021). In the context of membrane transport, the drying-resistant strain RZ300 demonstrates more advanced import systems, including transporters for oligopeptides (*opp*A, *opp*D, *opp*F genes), dipeptides (DDPPC genes) and amino acids (*liv*H, *liv*M, *liv*G, *liv*F), and also has the *nha*A gene for regulating Na+/H+-antiport. It is noteworthy that the regulation of the HAA gene of *Escherichia coli* is used to produce transgenic rice plants (*Oryza sativa* L. ssp. *japonica*) with increased resistance to salinity, but reduced one to drought (Wu et al., 2005).

A notable feature of the RZ300 is the presence of the *lgt* gene encoding an enzyme involved in lipoprotein biosynthesis, which is critically important for maintaining the integrity of the cell wall and, consequently, for resistance to stress factors. It has been shown that a decrease in Lgt levels in a clinical strain of uropathogenic *E. coli* leads to increased permeability of the outer membrane and increased sensitivity to serum and antibiotics (Diao et al., 2021). Finally, the *cdh*A gene encoding choline dehydrogenase and the *gpm*A gene responsible for phosphoglycerate mutase were found in the RZ300 strain, which indicate its ability to maintain the integrity of cell membranes.

A number of studies have shown that pathogenic bacteria mutated by the *cdh*A gene have reduced virulence and resistance to antibiotics acting on the cell wall (Pancholi et al., 2010). It is known that a deletion in the *gpm*A gene in *E. coli* leads to a specific hypersensitivity to H_2O_2, comparable to the deletion of the main H_2O_2 scavenger gene *kat*G. Exposure to H_2O_2enhances the transcription of *gpm*A, which highlights its role in protecting against oxidative stress. Thus, the *gpm*A gene can be defined as an element of the bacterial defense mechanism against oxidative stress (Roth et al., 2022).

Thus, the analysis of the genetic differences between the osmosis-resistant strain RZ300 and the osmosensitive 634b identified a large group of differentially represented genes, putatively associated with resistance to drying control. Among these, *opg*C and *opg*G2 are of the greatest interest, responsible for the synthesis of osmoregulated periplasmic glucans, which can play an important role in determining the resistance of nodule bacteria to desiccation on inoculated seeds. These data not only clarify the mechanisms of rhizobia resistance to desiccation, but also provide future research directions such as transferring of genetic constructs containing these genes into symbiotically effective rhizobia strains to meet the requirements of modern agricultural technologies.

Conclusion

In this study, the problem of ensuring the resistance of rhizobial inoculant strains to drying on the surface of inoculated soybean seeds is investigated in detail. We demonstrated that the survival rate of rhizobia strains on soybean seeds, firstly, is significantly higher than on a neutral carrier (glass beads), and secondly, does not depend on the soybean variety. It has been shown that the main differences in survival are related to the strains themselves, which indicates the importance of selecting promising inoculant strains based on this feature. Optimal technologies have been developed for the use of protective compositions that increase the resistance of strains to drying on seeds. To clarify the underlying mechanisms of strain resistance to desiccation, a comparative whole-genome analysis was performed on resistant and less resistant strains. Candidate genes potentially involved in the control of this trait were identified.

Of particular interest is the presence of the *opg*C gene in the resistant RZ300 strain, which is absent in the less resistant 634b strain and likely plays a role in bacterial resistance to desiccation on seeds. We assumed that this gene may serve as a molecular marker of resistance to desiccation of bacteria on seeds and can be used in the future for targeted breeding or genetic modification of rhizobial strains for microbial biopreparations.

References

Baker B.R., Ives C.M., Bray A., Caffrey M., Cochrane S.A. Undecaprenol kinase: function, mechanism and substrate specificity of a potential antibiotic target. *Eur J Med Chem*. 2021;210:113062. doi 10.1016/j.ejmech.2020.113062

Bashan Y., de-Bashan L., Prabhu S.R., Hernandez J. Advances in plant growth-promoting bacterial inoculant technology: formulations and practical perspectives (1998-2013). *Plant Soil*. 2014;378:1-33. doi 10.1007/s11104-013-1956-x

Bobkova Yu.A. Reaction of OAK Prudence soybeans to foliar fertilization with macro- and micro fertilizers in the conditions of the Orel region. *Vestnik Agrarnoy Nauki = Bulletin of Agrarian Science*. 2020;5(86):11-18. doi 10.17238/issn2587-666X.2020.5.11 (in Russian)

Bontemps-Gallo S., Lacroix J.M. New insights into the biological role of the osmoregulated periplasmic glucans in pathogenic and symbiotic bacteria. *Environ Microbiol Rep*. 2015;7(5):690-697. doi 10.1111/1758-2229.12325

Deaker R., Roughley R.J., Kennedy I.R. Desiccation tolerance of rhizobia when protected by synthetic polymers. *Soil Biol Biochem*. 2007; 39(2):573-580. doi 10.1016/j.soilbio.2006.09.005

Deriglazova G.M., Morozov A.N. Competitiveness of the domestic soybean variety EN Argenta compared with the Canadian breeding variety of OAK Prudence in the conditions of the Central Black Earth region. *Zernobobovye i Krupyanye Kultury = Legumes Groat Crops.* 2022;4(44):49-57. doi 10.24412/2309-348X-2022-4-49-57 (in Russian)

Diao J., Komura R., Sano T., Pantua H., Storek K.M., Inaba H., Ogawa H., … Yanagida H., Nishikawa J., Reid P.C., Cunningham C.N., Kapadia S.B. Inhibition of *Escherichia coli* lipoprotein diacylglyceryl transferase is insensitive to resistance caused by deletion of Braun's lipoprotein. *J Bacteriol*. 2021;203(13):e00149-21. doi 10.1128/JB.00149-21

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

Doyle J.J., Doyle J.L. Isolation of plant DNA from fresh tissue. *Focus.* 1990;12(1):13-15

Fred E.B. Some factors which influence the growth and longevity of the nodule bacteria. In: Fred E.B., Baldwin I.L., Mccoy E. (Eds) Root Nodule Bacteria and Leguminous Plants. Madison; Wisconsin: University of Wisconsin, 1932;104-117

Kincharova M.N., Matvienko E.V. The effectiveness of pre-sowing seed treatment in the fight against diseases of grain sorghum. *Agrarnyy Vestnik Urala = Agrarian Bulletin of the Urals.* 2021;9(212):2-10. doi 10.32417/1997-4868-2021-212-09-2-10 (in Russian)

Laktionov Yu.V., Kosulnikov Yu.V., Dudnikova D.V., Yahno V.V., Kojemyakov A.P. Pre-sowing protection of inoculated soybean *Glycine max* (L.) Merr. seeds by water-soluble polymer compositions and their solid-phase modification. *Sel'skokhozyajstvennaya Biologiya = Agricultural Biology.* 2019;54(5):1052-1059. doi 10.15389/agrobiology.2019.5.1052rus (in Russian)

Lamptey S., Ahiabor B.D.K., Yeboah S., Asamoah C. Response of soybean (*Glycine max*) to rhizobial inoculation and phosphorus application. *J Exp Biol Agric Sci.* 2014;2(1):72-77

Magomedov R.D., Tsekhmeistruk N.G., Shelyakin V.A., Ryabukha S.S., Didovich S.V. The influence of various strains *Rhizobium japonicum* (Kircher) Buchanan on soybean yield. *Maslichnyye Kul'tury = Oil Crops*. 2011;2(148-149):159-162 (in Russian)

Pancholi V., Boël G., Jin H. *Streptococcus pyogenes* Ser/Thr kinase-regulated cell wall hydrolase is a cell division plane-recognizing and chain-forming virulence factor. *J Biol Chem*. 2010;285(40):30861-30874. doi 10.1074/jbc.M110.153825

Parakhin N.V., Lysenko N.N., Petrova S.N., Kuzmicheva Yu.V., Ryzhov I.A. Evaluation of herbicide system efficacy in agrocenosis of different soybean varieties depending on tillage method. *Zemledeliye*. 2017;2:39-43 (in Russian)

Regar M.K., Meena R.H., Jat G., Mundra S.L. Effect of different rhizobial strains on growth and yield of soybean [*Glycine max* (L.) Merrill]. *Int J Curr Microbiol App Sci.* 2017;6(11):3653-3659. doi 10.20546/ijcmas.2017.611.427

Reina-Bueno M., Argandoña M., Nieto J.J., Hidalgo-García A., Iglesias-Guerra F., Delgado M.J., Vargas C. Role of trehalose in heat and desiccation tolerance in the soil bacterium *Rhizobium etli*. *BMC Microbiol*. 2012;12(1):207. doi 10.1186/1471-2180-12-207

Roth M., Goodall E.C.A., Pullela K., Jaquet V., François P., Henderson I.R., Krause K.-H. Transposon-directed insertion-site sequencing reveals glycolysis gene *gpmA* as part of the H_2O_2 defense mechanisms in *Escherichia coli*. *Antioxidants* (*Basel*). 2022;11(10):2053. doi 10.3390/antiox11102053

Rudra P., Hurst-Hess K.R., Cotten K.L., Partida-Miranda A., Ghosh P. Mycobacterial HflX is a ribosome splitting factor that mediates antibiotic resistance. *Proc Natl Acad Sci USA*. 2020;117(1):629-634. doi 10.1073/pnas.1906748117

Skorupska A., Janczarek M., Marczak M., Mazur A., Król J. Rhizobial exopolysaccharides: genetic control and symbiotic functions. *Microb Cell Fact*. 2006;5(1):7. doi 10.1186/1475-2859-5-7

Smith R.S. Legume inoculant formulation and application. *Can J Microbiol.* 1992;38(6):485-492. doi 10.1139/m92-080

Soria M.A., Pagliero F.E., Correa O.S., Kerber N.L., Garcia A.F. Tolerance of *Bradyrhizobium japonicum* E109 to osmotic stress and the stability of liquid inoculants depend on growth phase. *World J Microbiol Biotechnol*. 2006;22(1):1235-1241. doi 10.1007/s11274-006-9166-9

Vasilchikov A.G., Gurev G.P. Adaptation of soybean varieties with different growing seasons to the soil and climatic conditions of the Oryol region. *Zernobobovye i Krupyanye Kultury = Legumes Groat*

Crops. 2018;4(28):49-53. doi 10.24411/2309-348X-2018-00001 (in Russian)

Vavilov P.P., Posypanov G.S. Legumes and the Issue of Vegetable Protein. Moscow, 1983 (in Russian)

Vincent J.M., Thompson J.A., Donovan K.O. Death of root nodule bacteria on drying. *Aust J Agric Res.* 1961;13(2):258-270. doi 10.1071/AR9620258

Volobueva O.G., Trukhachev S., Belopukhov C., Seregina I. Comparative study of symbiotic activity of legumes when using Risotorphin and Epin-extra. *Braz J Biol*. 2023;83(3):e264218. doi 10.1590/1519-6984.264218

Vriezen J.A., de Bruijn F.J., Nüsslein K. Desiccation responses and survival of *Sinorhizobium meliloti* USDA 1021 in relation to growth phase, temperature, chloride and sulfate availability. *Lett Appl Microbiol*. 2006;42(2):172-178. doi 10.1111/j.1472-765X.2005.01808.x

Vriezen J., Bruijn F., Nüsslein K. Responses of rhizobia to desiccation in relation to osmotic stress, oxygen, and temperature. *Appl Environ Microbiol.* 2007;73(11):3451-3459. doi 10.1128/AEM.02991-06

Wu L., Fan Z., Guo L., Li Y., Chen Z., Qu L. Over-expression of the bacterial *nhaA* gene in rice enhances salt and drought tolerance. *Plant Sci*. 2005;168(2):297-302. doi 10.1016/j.plantsci.2004.05.033

Yadav J., Verma J.P., Rajak V.K., Tiwari K.N. Selection of effective indigenous *Rhizobium* strain for seed inoculationof chickpea (*Cicer aritenium* L.) production. *Bacteriol J*. 2011;1(1):24-30. doi 10.3923/bj.2011.24.30

Zhang Y., Ku Y.S., Cheung T.Y., Cheng S.S., Xin D., Gombeau K., Cai Y., Lam H.M., Chan T.F. Challenges to rhizobial adaptability in a changing climate: genetic engineering solutions for stress tolerance. *Microbiol Res*. 2024;288:127886. doi 10.1016/j.micres.2024.127886

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

Received July 11, 2025. Revised December 4, 2025. Accepted December 4, 2025.

ANIMAL GENETICS
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):444-450

doi 10.18699/vjgb-26-48

Comparative cytogenetic analysis of the germline-restricted chromosome in Fringillidae species (Passeriformes, Aves)

L.P. Malinovskaya [1, 2], K.V. Tishakova [1, 3], P.M. Borodin [2]

[1] Novosibirsk State University, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
borodin@bionet.nsc.ru

Abstract. An additional germline-restricted chromosome (GRC) has been found in the germline cells of all studied passerine bird species. It is eliminated from somatic cells during early embryogenesis and from spermatocytes after the first or second division of male meiosis. The GRC is transmitted across generations predominantly via the maternal line. It contains amplified and rearranged copies of genomic regions from the standard chromosome set. Some of these genes are expressed in the gonads of both males and females. However, the function and evolutionary dynamics of the GRC remain unknown. We conducted a comparative cytogenetic analysis of the GRC in five closely related finch species – the Eurasian bullfinch *Pyrrhula pyrrhula*, the common greenfinch *Chloris chloris*, the European goldfinch *Carduelis carduelis*, the common redpoll *Acanthis flammea*, and the pine grosbeak *Pinicola enucleator* – using fluorescent *in situ* hybridization (FISH) with a whole-chromosome DNA probe derived from the bullfinch GRC on spread spermatocytes of these species and immunolocalization of synaptonemal complex (SC) and centromere proteins. We described for the first time the SC karyotype of the pine grosbeak ($2n$ = 82 + GRC). The standard chromosome set consists of nine submetacentric bivalents (seven macro- and two microbivalents) and 32 acrocentric microbivalents. All acrocentric microbivalents contain centromeres composed of multiple centromeric domains (metapolycentromeres). The grosbeak GRC is a large acrocentric macrounivalent. Cross-species *in situ* hybridization of the bullfinch GRC DNA probe showed only weak signals on the GRC of the grosbeak and redpoll, whereas no signal was detected on the greenfinch and goldfinch GRCs. These data are consistent with published results for two other representatives of this family and indicate rapid divergence and high species specificity of GRC sequences within the family Fringillidae. We also detected interspecies differences in the localization of sequences homologous to the bullfinch GRC on the bivalents of the standard set of these species. Thus, our data indicate rapid evolution of the GRC's genetic composition and reveal species-specific dynamics of increase and decrease in the copy number of detected sequences in the standard chromosome set during the evolution of songbird species.

Key words: germline-restricted chromosome; GRC; chromosome evolution; fluorescent *in situ* hybridization; centromere; finches

For citation: Malinovskaya L.P., Tishakova K.V., Borodin P.M. Comparative cytogenetic analysis of the germline-restricted chromosome in Fringillidae species (Passeriformes, Aves). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):444-450. doi 10.18699/vjgb-26-48

Funding. This research was supported by the Russian Science Foundation, grant No. 23-14-00182.

Acknowledgements. We thank the Microscopy Center of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia (Regulation No. 3054), supported by grants from the Ministry of Science and Higher Education of the Russian Federation No. FSUS-2024-0018 and FWNR-2022-0015.

Сравнительный цитогенетический анализ хромосомы, ограниченной клетками зародышевой линии, у вьюрковых птиц (Passeriformes, Aves)

Л.П. Малиновская [1, 2], К.В. Тишакова [1, 3], П.М. Бородин [2]

[1] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
borodin@bionet.nsc.ru

Аннотация. Добавочная хромосома, ограниченная клетками зародышевой линии (germline-restricted chromosome, GRC), обнаружена у всех исследованных воробьинообразных птиц. Она элиминируется из клеток соматической линии в раннем эмбриогенезе и из сперматоцитов после первого или второго деления мейоза. GRC передается в ряду поколений преимущественно по материнской линии, содержит амплифицированные и перестроенные копии последовательностей хромосом основного набора. Некоторые из них экспрессируются в гонадах самцов и самок. Однако функция и эволюционная динамика GRC остаются неизвестными. Мы провели сравнительный цитогенетический анализ GRC пяти видов птиц семейства Вьюрковые – обыкновенного снегиря *Pyrrhula pyrrhula*, обыкновенной зеленушки *Chloris chloris*, обыкновенного щегла *Carduelis carduelis*, обыкновенной чечётки *Acanthis flammea* и обыкновенного щура *Pinicola enucleator* – с использованием флуоресцентной гибридизации *in situ* ДНК-зонда к целой GRC обыкновенного снегиря с материалом ядер распластанных сперматоцитов данных видов и иммунолокализации белков синаптонемного комплекса и центромеры. Мы впервые описали кариотип синаптонемных комплексов обыкновенного щура ($2n = 82 + \text{GRC}$). Биваленты основного набора состоят из девяти субметацентрических (семь макро- и два микробивалента) и 32 акроцентрических микробивалентов. Все акроцентрические микробиваленты основного набора содержат центромеры, состоящие из нескольких центромерных доменов (метаполицентромеры). GRC щура представляет собой крупный акроцентрический макроунивалент. Перекрестная гибридизация *in situ* ДНК-зонда к GRC снегиря показала только слабые сигналы на GRC щура и чечётки, тогда как на GRC зеленушки и щегла сигналы отсутствовали. Эти данные согласуются с опубликованными результатами, полученными для двух других представителей этого семейства, и свидетельствуют о высокой видовой специфичности последовательностей GRC в пределах семейства Вьюрковые. Мы также обнаружили межвидовые различия в локализации последовательностей, сходных с последовательностями GRC снегиря, на бивалентах основного набора исследуемых видов. Таким образом, наши данные указывают на быструю эволюцию генетического состава GRC и видоспецифичную динамику увеличения и сокращения копийности выявленных последовательностей на хромосомах основного набора в ходе эволюции певчих птиц.

Ключевые слова: germline-restricted chromosome; GRC; хромосомная эволюция; флуоресцентная гибридизация *in situ*; центромера; вьюрки

Introduction

The Germline-Restricted Chromosome (GRC) has been identified in all studied passerine birds, indicating a monophyletic origin (Torgasheva et al., 2019). The GRC in all examined species contains amplified and rearranged copies of sequences from the standard (A) chromosomes. Some of these sequences are expressed in the gonads of both males and females (Biederman et al., 2018; Kinsella et al., 2019). The GRC is eliminated from somatic cell lineages during early embryogenesis and from spermatocytes after the first or second meiotic division (Pigozzi, Solari, 2005). The GRC is transmitted predominantly through the maternal line (Pei et al., 2022).

The elimination of the GRC from somatic cells allows genes located on it to escape the selective pressure characteristic of genes on the standard chromosomes. This is expected to lead to rapid changes in the GRC's genetic composition. Indeed, the size and genetic content of the GRC vary widely among species. In some species, the GRC is one of the largest macrochromosomes (macro-GRC), while in others, it is one of the smallest microchromosomes (micro-GRC) (Borodin et al., 2022).

Results of cross-species fluorescent *in situ* hybridization (FISH) using whole-chromosome DNA probes derived from the GRC of the zebra finch (*Taeniopygia guttata*), the sand martin (*Riparia riparia*), the Eurasian siskin (*Spinus spinus*), and the great tit (*Parus major*) demonstrated an astonishingly low level of similarity between the GRCs of different species (the divergence time between the most distant species being ~25–30 million years) (Torgasheva et al., 2019, 2021). This indicates extremely rapid evolution of this chromosome's genetic content. This conclusion is supported by comparative sequencing data of the micro-GRC in closely related species, such as the thrush nightingale (*Luscinia luscinia*) and the common nightingale (*L. megarhynchos*), which revealed substantial differences in GRC composition despite their relatively recent divergence (approximately 1.8 million years ago) (Schlebusch et al., 2023).

Many genes on the GRC are present in a fragmented, likely non-functional state, with the exception of a small number of highly conserved and presumably essential genes. This rapid change in genetic composition and the abundance of duplications, deletions, and pseudogenes stand in sharp contrast to the generally conserved avian karyotype, making the GRC the most rapidly evolving chromosome in the passerine bird genome (Borodin et al., 2022).

The rapid changes in the size and genetic composition of the GRC and the consequent diversity of its gene content – including genes involved in reproductive system development – have prompted the formulation of several hypotheses. One posits that the GRC resolves germline-soma conflict by isolating genes with mutually antagonistic effects. Another hypothesis views the GRC as a highly efficient genomic parasite. A third considers this chromosome a potential driver of speciation, as rapid divergence of its genetic content in isolated populations is expected to lead to genetic incompatibilities. In our view, these hypotheses are not mutually exclusive. It is highly probable that the GRC fulfills all of these roles simultaneously (Borodin et al., 2022; Borodin, 2023; Vontzou et al., 2023).

The scale and patterns of GRC variability at the cytogenetic level within narrow taxonomic groups, such as a family, remain insufficiently studied. Cross-species FISH experiments comparing GRC genetic content within a single family have so far encompassed only two or three species (Torgasheva

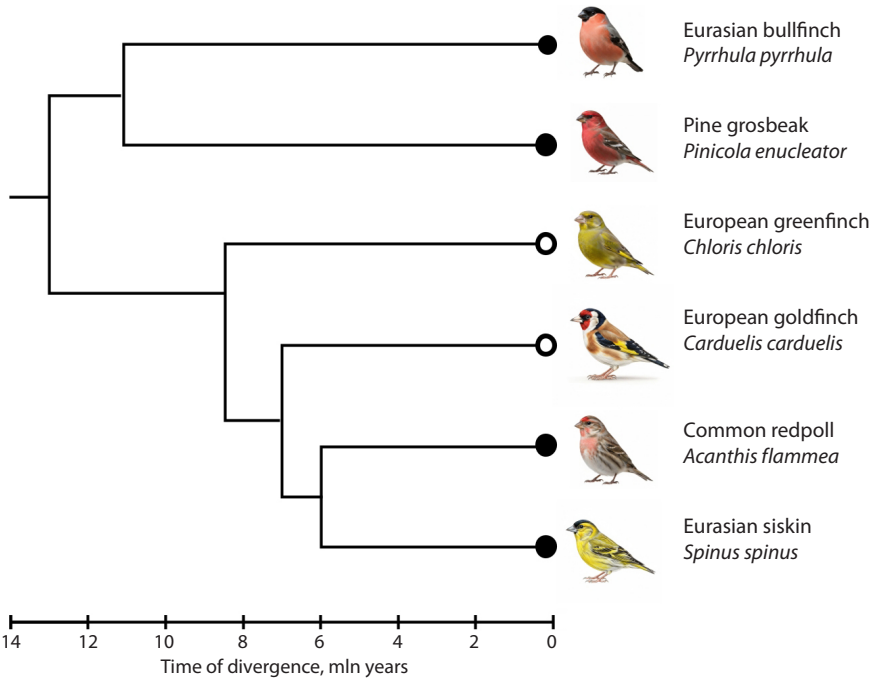


Fig. 1. A cladogram of bird species from the family Fringillidae used for the comparative analysis of the GRC.

A black circle denotes a macro-GRC, a white circle denotes a micro-GRC. The cladogram was constructed using the Timetree.org resource (Kumar et al., 2017) (last accessed December 3, 2025).

et al., 2019). A more systematic comparative analysis of the GRC across multiple members of a single clade has not been previously conducted.

The aim of the present work was a comparative cytogenetic analysis of the GRC in five species of the family Fringillidae and an assessment of the pattern of *in situ* hybridization of a DNA probe derived from the whole macro-GRC of the Eurasian bullfinch (*Pyrrhula pyrrhula*) with nuclear spreads of spermatocytes of several species with varying degrees of phylogenetic relatedness and different GRC morphology. For this purpose, we selected four species representing different genera within this family (Fig. 1): the European greenfinch (*Chloris chloris*) (micro-GRC), the European goldfinch (*Carduelis carduelis*) (micro-GRC), the common redpoll (*Acanthis flammea*) (macro-GRC), and the pine grosbeak (*Pinicola enucleator*), the karyotype and GRC morphology of which had not been described until now.

The selected species represent different evolutionary lineages within the family: the divergence time between the bullfinch and the pine grosbeak is approximately 11 million years, and between the bullfinch and the greenfinch, goldfinch, and redpoll, it is approximately 13 million years (Price et al., 2014; Hooper, Price, 2017). Integrating the results of this experiment with data previously obtained from the analysis of two other species in this family – the Eurasian siskin (*Spinus spinus*) (macro-GRC) and the European goldfinch (*Carduelis carduelis*) (micro-GRC) (Torgasheva et al., 2019), which diverged approximately 8.5 million years ago (Price et al., 2014; Hooper, Price, 2017), – will allow for a more detailed assessment of GRC variability within a single family.

Materials and methods

Biological material was obtained from birds delivered with fatal injuries to the Wildlife Rehabilitation Center (Novosibirsk) between April and May 2022–2023. This study examined one male of each species: the Eurasian bullfinch (*Pyrrhula pyrrhula*), pine grosbeak (*Pinicola enucleator*), European greenfinch (*Chloris chloris*), European goldfinch (*Carduelis carduelis*), and common redpoll (*Acanthis flammea*). Euthanasia was performed by isoflurane overdose (Laboratories Karizoo, S.A., Barcelona, Spain). Handling of birds and euthanasia were conducted in accordance with national regulations for the care and use of laboratory animals. The experimental protocol was reviewed and approved by the Bioethics Commission of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (Protocols No. 114 dated December 17, 2021, and No. 199 dated November 21, 2024).

Synaptonemal complex (SC) spreads were prepared according to the method of A.H. Peters et al. (1997). Immunocytochemical detection of SC proteins and centromeres was performed following the protocol of L.K. Anderson et al. (1999) using the following primary antibodies: rabbit polyclonal antibodies to SYCP3 (dilution 1:500; Abcam, UK; cat. No. ab15093), anti-centromere antibodies from the serum of patients with CREST syndrome (dilution 1:100; Antibodies Inc., USA; cat. No. 15-234). The following secondary antibodies were used: goat anti-rabbit IgG conjugated with Cy3 (dilution 1:500; Jackson ImmunoResearch, USA; cat. No. 111-165-144), donkey anti-human IgG conjugated with AMCA (dilution 1:100; Jackson ImmunoResearch, USA; cat. No. 709-155-149).

Slides were incubated overnight at +4 °C with primary antibodies and for one hour at +37 °C with secondary antibodies in a humid chamber. To prevent photobleaching, Vectashield mounting medium (Vector Laboratories, USA; cat. No. H-1000-10) was applied to the slides.

The DNA probe for the whole bullfinch GRC was obtained by microdissection of five micronuclei copies from meiotic chromosome spreads, as described by A.A. Torgasheva et al. (2019). Slides were preliminarily stained with a 0.1 % Giemsa solution (Sigma, USA) for 3–5 minutes at room temperature. DNA from the microdissected micronuclei was amplified and labeled with biotin-11-dUTP (Sigma, USA) using the GenomePlex Whole Genome Amplification Kit (Sigma-Aldrich, USA; cat. No. WGA1).

FISH using the bullfinch GRC DNA probe was performed according to a standard protocol (Liehr et al., 2017) with some modifications. The hybridization mixture (32 µl) contained hybridization buffer (50 % formamide, 2× SSC), 0.2 % Tween 20, and 40 ng of labeled probe. DNA on slides pretreated with RNase A was denatured in 70 % formamide with 2× SSC at +72 °C for 3 minutes. The probe was denatured at +95 °C for 5 minutes. Hybridization was carried out overnight at +39 °C in a humid chamber. The biotin-labeled probe was detected using avidin-FITC (dilution 1:400) and anti-avidin-FITC (dilution 1:200) (Vector Laboratories, USA). Slides were mounted in Vectashield medium with DAPI (Vector Laboratories, USA, cat. No. H-1200-10).

Images of SCs after immunolocalization and FISH were captured using a CCD camera mounted on an Axioplan 2 microscope (Carl Zeiss, Germany) with filters No. 49 (DAPI), 10 (FITC), and 15 (TRITC) (ZEISS, Germany) and ISIS4 software (METASystems GmbH, Germany). Brightness and contrast of the images were adjusted using Corel PaintShop Photo Pro X6 (Alludo, Canada).

For the construction of the SC karyotype idiogram for the pine grosbeak, 28 cells were measured. The length of SCs and the position of centromeres were determined in micrometers using the MicroMeasure 3.3 program (Reeves, 2001). All raw data are provided in the Supplementary Material[1]. In each cell, SCs were ranked by relative length and centromeric index, and average values across all cells were calculated.

Descriptive statistics were obtained using Statistica 6.0 (StatSoft Inc., Tulsa, OK, USA, 2001). The text presents values as mean ± standard deviation.

Results

Neither the somatic nor the synaptonemal complex (SC) karyotype of the pine grosbeak (*Pinicola enucleator* L.) had been previously described. We found that the SC karyotype of this species comprises 41 pairs of bivalents from the standard set and a GRC ($2n = 82 + GRC$; fundamental number (FN) = 50) (Fig. 2). The total SC length was 334 ± 47 µm. Macrobivalents 1–7, as well as microbivalents 9 and 16, are submetacentric; microbivalents 8 and all other microbivalents are acrocentric (Fig. 3). All acrocentric microbivalents of the standard set contain centromeres composed of multiple centromeric domains. Such centromeres are conventionally referred to as metapolycentromeres (Grishko, Borodin, 2024). The GRC of the pine grosbeak is a large acrocentric macrochromosome. At the pachytene stage, it forms an acrocentric univalent that is recognized by antibodies against the SYCP3 protein, which forms the lateral element of the SC (Fig. 2).

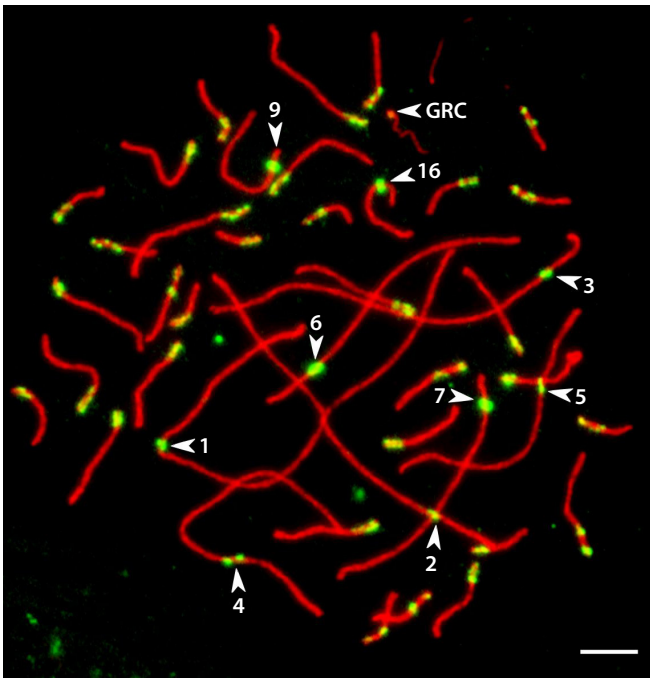


Fig. 2. Photograph of an SC spread from a pine grosbeak pachytene spermatocyte after immunolocalization of the SYCP3 protein (red) and centromeric proteins (green).

Arrowheads indicate submetacentric bivalents 1–7, 9, 16, and the GRC. Scale bar = 5 µm.

To assess the similarity between the GRC sequences of the bullfinch and the GRCs of other finch species, cross-species FISH was performed using a DNA probe for the bullfinch GRC obtained previously (Grishko et al., 2025). The probe hybridized weakly to the macro-GRC of the pine grosbeak and common redpoll, and did not hybridize to the micro-GRC of the European greenfinch and European goldfinch (see the Table).

In all four species, the probe labeled several regions on the standard bivalents. In SC spreads from the pine grosbeak, hybridization signals were observed on the short arm of macrobivalent 4 and in the long arm region of macrobivalent 5 (Fig. 4*a*). The probe also produced weak signals on many other bivalents. In SC spreads from the European greenfinch and common redpoll, the probe hybridized to the pericentromeric regions of several microbivalents and all macrobivalents, with the exception of a single metacentric macrobivalent (likely ZZ) (Fig. 4*b*, *d*). In SC spreads from the European goldfinch, the hybridization signal was detected in the pericentromeric region of macrobivalent 1 and on the short arm of macrobivalent 6 (Fig. 4*c*).

[1] Supplementary Material is available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Mal_Engl_30_3.xlsx

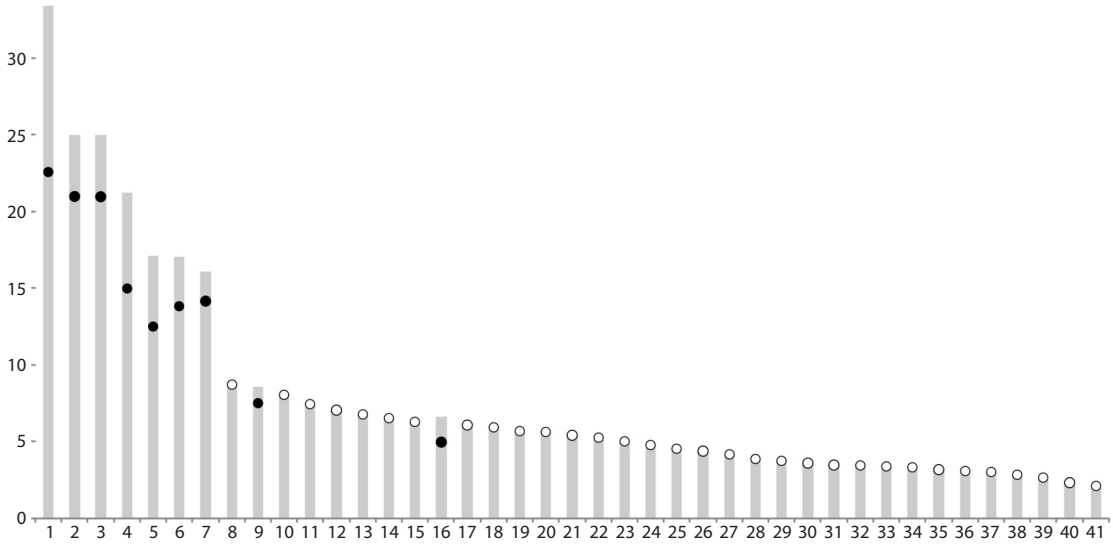


Fig. 3. SC idiogram of the pine grosbeak karyotype, excluding the GRC.

The Y axis shows the average SC length in μm. The X axis shows bivalents ordered by decreasing size. Black circles indicate the location of monocentromeres, white circles indicate metapolycentromeres.

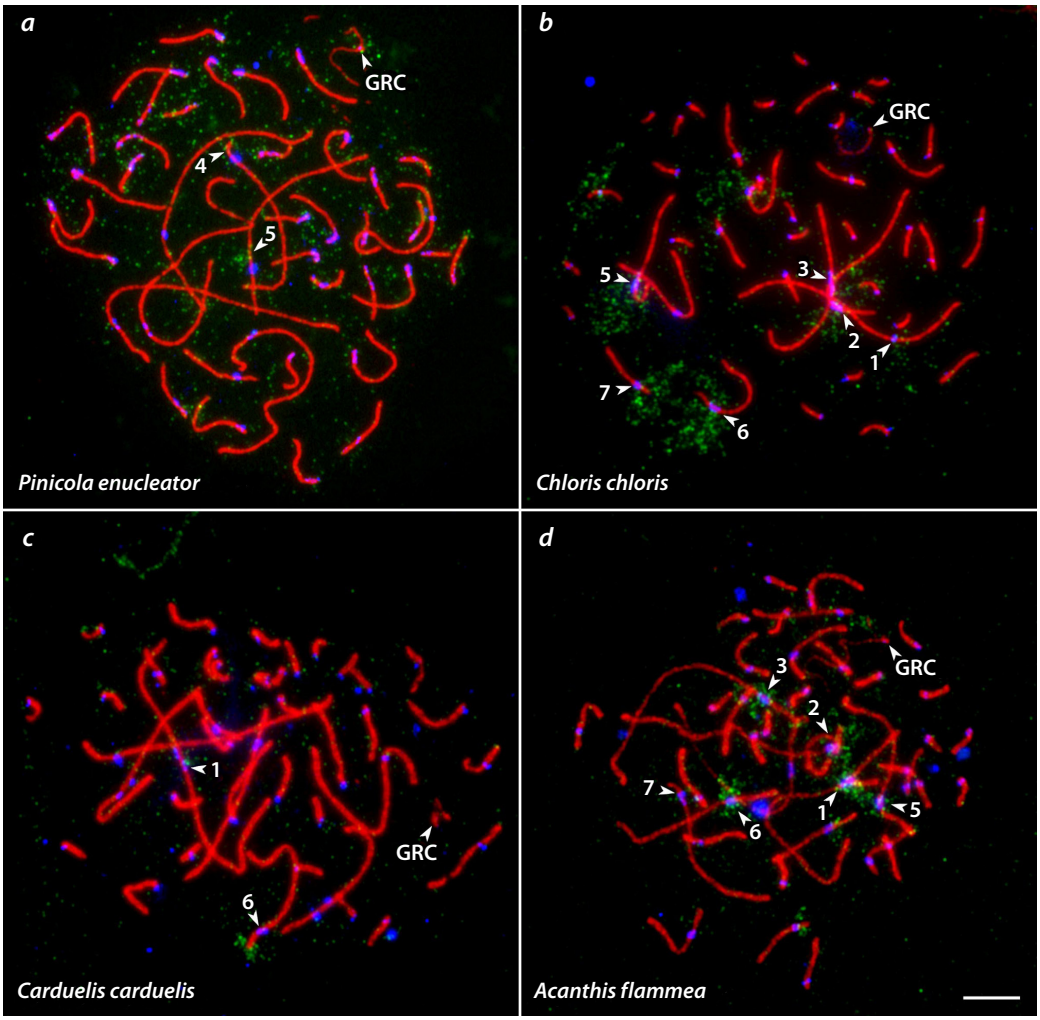


Fig. 4. Photographs of SC spreads from pachytene spermatocytes of the pine grosbeak (*a*), European greenfinch (*b*), European goldfinch (*c*), and common redpoll (*d*) after FISH with the bullfinch GRC DNA probe (green), immunolocalization of the SYCP3 protein (red), and centromeric proteins (blue).

Arrowheads indicate the GRC and standard bivalents labeled by the bullfinch GRC DNA probe. Scale bar = 5 μm.

Detection of hybridization signal after FISH with the bullfinch GRC DNA probe on SC spreads of finch species

Species	Signal	
	on GRC	on macrobivalent[α]
Bullfinch *Pyrrhula pyrrhula*[β]	Strong	c1, 6
Pine grosbeak *Pinicola enucleator*	Weak	p4, q5
European greenfinch *Chloris chloris*	Absent	c1-3, 5-7
European goldfinch *Carduelis carduelis*	Absent	c1, p6
Common redpoll *Acanthis flammea*	Weak	c1-3, 5-7

[α] c – centromere, p – short arm of the bivalent, q – long arm of the bivalent.
[β] (Grishko et al., 2025).

Discussion

In the present study, we have described, for the first time, the synaptonemal complex (SC) karyotype of the pine grosbeak, demonstrating that it, like the karyotypes of half of the studied finch species (Borodin et al., 2022; Malinovskaya et al., 2022), contains a macro-GRC. We established that many bivalents in this species possess metapolycentromeres. The presence of metapolycentromeres was previously demonstrated in the Eurasian bullfinch and the common linnet (*Linaria cannabina*) from the same family, Fringillidae (Grishko et al., 2023).

The primary result of this investigation is the detection of high interspecific variability in the genetic composition of the GRC among representatives of the family Fringillidae. Despite the relatively close phylogenetic relatedness among these finches (with divergence time between the bullfinch and the other studied species ranging from approximately 11 to 13 million years (Price et al., 2014; Hooper, Price, 2017)), the DNA probe for the bullfinch macro-GRC yielded a weak hybridization signal on the macro-GRC of the pine grosbeak and common redpoll and produced no detectable signal on the micro-GRC of the European greenfinch and European goldfinch.

The absence of a detectable signal on the micro-GRC of the greenfinch and goldfinch, in contrast to the weak signal on the macro-GRC of the pine grosbeak and redpoll, suggests that GRC size may be one of the factors influencing the degree of detectable sequence similarity, at least within a single family. This pattern aligns with data from the literature: for instance, with a shorter divergence time (~9 million years), a DNA probe for the micro-GRC of the Eurasian siskin produced a weak hybridization signal on the macro-GRC of the European goldfinch (Torgasheva et al., 2019). It is likely that over ~9–13 million years of divergence, micro-GRCs lost most of their shared sequences, resulting in the absence of a hybridization signal, whereas macro-GRCs retained a sufficient number of homologous sequences to produce a weak signal.

In our cross-species FISH experiments, we identified differences in the hybridization patterns of the bullfinch GRC DNA probe with the standard bivalents among different finch species. It was previously shown that this probe hybridized to the pericentromeric regions of specific macrobivalents and a number of microbivalents in the bullfinch (Grishko et al., 2025). In our cross-species FISH experiments, this probe demonstrated a different hybridization pattern. The observed interspecific differences in hybridization patterns indicate rapid evolution of the repetitive sequences located on the standard chromosomes, occurring against the backdrop of the overall conservation of avian genomes.

The hybridization of the bullfinch GRC DNA probe to the pericentromeric regions of all macrobivalents, except macrobivalent 4, in the greenfinch and redpoll may indicate the conservation of these sequences in these two species. This pattern contrasts with that observed in the pine grosbeak and goldfinch, where the probe hybridized specifically only to distinct regions of certain macrobivalents. This points to a more limited and species-specific distribution of the detectable sequences across the standard bivalents in these species. The weak diffuse signals detected on most bivalents in the pine grosbeak may represent "ghost" sequences that were once more widely distributed but have subsequently degraded or been replaced in most genomic regions.

Thus, the GRC is a rapidly evolving genomic element in passerine birds, showing a low degree of similarity even among species within the single family Fringillidae. The differences in the hybridization pattern of the bullfinch GRC DNA probe on spermatocyte nuclear spreads from different finch species indicate a species-specific dynamic of amplification and reduction in the copy number of the detectable sequences on the standard bivalents over the course of evolution.

Conclusion

Our study has demonstrated that the genetic composition of the germline-restricted chromosome (GRC) in individual representatives of the family Fringillidae is highly species-specific, indicating rapid evolution of the GRC. The presence of shared sequences between the bullfinch GRC and the standard bivalents of different finch species, coupled with the species-specific patterns of their localization, aligns with similar results from cross-species FISH experiments utilizing DNA probes derived from the GRC of other bird species.

References

Anderson L.K., Reeves A., Webb L.M., Ashley T. Distribution of crossing over on mouse synaptonemal complexes using immunofluorescent localization of MLH1 protein. *Genetics*. 1999;151(4):1569-1579. doi 10.1093/genetics/151.4.1569

Biederman M.K., Nelson M.M., Asalone K.C., Pedersen A.L., Saldanha C.J., Bracht J.R. Discovery of the first germline-restricted gene by subtractive transcriptomic analysis in the Zebra Finch, *Taeniopygia guttata*. *Curr Biol*. 2018;28(10):1620-1627. doi 10.1016/j.cub.2018.03.067

Borodin P., Chen A., Forstmeier W., Fouché S., Malinovskaya L., Pei Y., Reifová R., Ruiz-Ruano F.J., Schlebusch S.A., Sotelo-Muñoz M., Torgasheva A., Vontzou N., Suh A. Mendelian nightmares: the germline-restricted chromosome of songbirds. *Chromosome Res*. 2022;30(2-3):255-272. doi 10.1007/s10577-022-09688-3

Borodin P.M. Germline-restricted chromosomes of songbirds. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2023;27(6):641-650. doi 10.18699/VJGB-23-75

Grishko E.O., Borodin P.M. Structure and evolution of metapolycentromeres. *Vavilov J Genet Breed*. 2024;28(6):592-601. doi 10.18699/vjgb-24-66

Grishko E., Malinovskaya L., Slobodchikova A., Kotelnikov A., Torgasheva A., Borodin P. Cytological analysis of crossover frequency and distribution in male meiosis of cardueline finches (Fringillidae, Aves). *Animals*. 2023;13(23):3624. doi 10.3390/ani13233624

Grishko E., Malinovskaya L., Tishakova K., Borodin P. Germline restricted chromosome (GRC) in diploid and polyploid spermatocytes of the Eurasian bullfinch, *Pyrrhula pyrrhula* (Fringillidae, Passeriformes, Aves). *Animals*. 2025;15(23):3394. doi 10.3390/ani15233394

Hooper D.M., Price T.D. Chromosomal inversion differences correlate with range overlap in passerine birds. *Nat Ecol Evol*. 2017;1(10):1526-1534. doi 10.1038/s41559-017-0284-6

Kinsella C.M., Ruiz-Ruano F.J., Dion-Côté A.-M., Charles A.J., Gossmann T.I., Cabrero J., Kappei D., Hemmings N., Simons M.J.P., Camacho J.P.M., Forstmeier W., Suh A. Programmed DNA elimination of germline development genes in songbirds. *Nat Commun*. 2019;10(1):5468. doi 10.1038/s41467-019-13427-4

Kumar S., Stecher G., Suleski M., Hedges S.B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 2017;34(7):1812-1819. doi 10.1093/molbev/msx116

Liehr T., Kreskowski K., Ziegler M., Piaszinski K., Rittscher K. The standard FISH procedure. In: Liehr T. (Ed.) Fluorescence In Situ Hybridization (FISH). Springer Protocols Handbooks. Springer, 2017;109-118. doi 10.1007/978-3-662-52959-1_9

Malinovskaya L.P., Slobodchikova A.Y., Grishko E.O., Pristyazhnyuk I.E., Torgasheva A.A., Borodin P.M. Germline-restricted chromosomes and autosomal variants revealed by pachytene karyotyping of 17 avian species. *Cytogenet Genome Res*. 2022;162(3):148-160. doi 10.1159/000524681

Pei Y., Forstmeier W., Ruiz-Ruano F.J., Mueller J.C., Cabrero J., Camacho J.P.M., Alché J.D., … Hertel M., Teltscher K., Knief U., Suh A., Kempenaers B. Occasional paternal inheritance of the germline-restricted chromosome in songbirds. *Proc Natl Acad Sci USA*. 2022;119(4):e2103960119. doi 10.1073/pnas.2103960119

Peters A.H., Plug A.W., van Vugt M.J., de Boer P. A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. *Chromosome Res*. 1997;5(1):66-68. doi 10.1023/a:1018445520117

Pigozzi M.I., Solari A.J. The germ-line-restricted chromosome in the zebra finch: recombination in females and elimination in males. *Chromosoma*. 2005;114(6):403-409. doi 10.1007/s00412-005-0025-5

Price T.D., Hooper D.M., Buchanan C.D., Johansson U.S., Tietze D.T., Alström P., Olsson U., Ghosh-Harihar M., Ishtiaq F., Gupta S.K., Martens J., Harr B., Singh P., Mohan D. Niche filling slows the diversification of Himalayan songbirds. *Nature*. 2014;509(7499):222-225. doi 10.1038/nature13272

Reeves A. MicroMeasure: a new computer program for the collection and analysis of cytogenetic data. *Genome*. 2001;44(3):439-443. doi 10.1139/g01-037

Ruiz-Ruano F.J., Schlebusch S.A., Vontzou N., Moreno H., Biegler M.T., Kutschera V.E., Ekman D., … Segami J.C., Tan D.J.X., Torgasheva A., Whibley A., Suh A. Programmed DNA elimination drives rapid genomic innovation in two thirds of all bird species. *bioRxiv*. 2025. doi 10.1101/2025.07.16.664580

Schlebusch S.A., Rídl J., Poignet M., Ruiz-Ruano F.J., Reif J., Pajer P., Pačes J., Albrecht T., Suh A., Reifová R. Rapid gene content turnover on the germline-restricted chromosome in songbirds. *Nat Commun*. 2023;14(1):4579. doi 10.1038/s41467-023-40308-8

Torgasheva A.A., Malinovskaya L.P., Zadesenets K.S., Karamysheva T.V., Kizilova E.A., Akberdina E.A., Pristyazhnyuk I.E., Shnaider E.P., Volodkina V.A., Saifitdinova A.F., Galkina S.A., Larkin D.M., Rubtsov N.B., Borodin P.M. Germline-restricted chromosome (GRC) is widespread among songbirds. *Proc Natl Acad Sci USA*. 2019;116(24):11845-11850. doi 10.1073/pnas.1817373116

Torgasheva A., Malinovskaya L., Zadesenets K., Shnaider E., Rubtsov N., Borodin P. Germline-restricted chromosome (GRC) in female and male meiosis of the great tit (Parus major, Linnaeus, 1758). *Front Genet*. 2021;12:768056. doi 10.3389/fgene.2021.768056

Vontzou N., Pei Y., Mueller J.C., Reifová R., Ruiz-Ruano F.J., Schlebusch S.A., Suh A. Songbird germline-restricted chromosome as a potential arena of genetic conflicts. *Curr Opin Genet Dev*. 2023;83:102113. doi 10.1016/j.gde.2023.102113

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

Received November 11, 2025. Revised December 23, 2025. Accepted December 24, 2025.

ANIMAL GENETICS
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):451-460

doi 10.18699/vjgb-26-49

Genetic diversity of horses of the Sargarinsko-Alexeevskaya and Irmen cultures of the Ob-Irtysh region of Western Siberia and their genetic proximity to modern horses of indigenous breeds

M.A. Kusliy [1], A.A. Yurlova [1] ✉, N.V. Vorobyeva [1], A.A. Proskuryakova [1], M.A. Demin [2], S.M. Sitnikov [3], V.N. Zharonkin [4], S.S. Onishchenko [5], A.K. Kasparov [6], A.E. Tupikin [7], A.S. Graphodatsky [1], A.S. Molodtseva [1], A.A. Tishkin [8]

[1] Institute of Molecular and Cellular Biology of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Altai State Pedagogical University, Barnaul, Russia
[3] Autonomous non-profit organization "Altai Archaeological Society", Barnaul, Russia
[4] Department of Roskomnadzor for the Kemerovo Oblast, Kemerovo, Russia
[5] Committee for the Protection of Cultural Heritage of Kuzbass, Kemerovo, Russia
[6] Institute for the History of Material Culture of the Russian Academy of Sciences, St. Petersburg, Russia
[7] Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[8] Altai State University, Barnaul, Russia
✉ annush@mcb.nsc.ru

Abstract. The multidisciplinary approach is in increasing use in modern science for solving complicated problems. Molecular genetics not only helps us understand biological processes, such as evolution and speciation, but also sheds light on numerous historical questions as to the directions of peoples' migrations, the degree of interpenetration of contemporaneous cultures, and their continuity. In particular, investigation of phylogenetic relationships of domestic animals allows us to detail the interactions between bearers of different archaeological cultures. In the Bronze Age, Inner Asia and adjacent territories were characterized by intense human migrations and rapid spread of productive economies, including livestock farming. Here we examine the phylogenetic patterns of horses from two important Bronze Age Ob-Irtysh cultures in Western Siberia, the Sargarinsko-Alexeevskaya and Irmen ones, and the degree of their genetic proximity to horses from earlier (Andronovo and Eluninskaya) and later (Khereksur and "Deer Stone", Biykenskaya, Bystryanskaya and Pazyryk) cultures in the region and adjacent territories. Data obtained from sequencing and analysis of mitochondrial genomes reveal differences in the mitochondrial gene pool of horses from the Sargarinsko-Alexeevskaya and Irmen cultures of the south of Western Siberia, highlighting the unique mitochondrial genetic diversity of the original horse herds of these cultures and the lack of close breeding contacts between them. We demonstrate an overlap between the mitochondrial gene pools of horses from the Khereksur and "Deer Stone" cultures of Mongolia and the Andronovo culture. We also established continuity between many of the obtained haplotypes of horses from the Early, Developed, Late Bronze Age, and Early Iron Age in southern Western Siberia, indicating the preservation of a significant part of the maternal gene pool diversity of domestic horses in the region across several historical and cultural periods. The similarity of mitochondrial haplotypes among horses of the Sargarinsko-Alexeevskaya culture, modern horses of the Akhal-Teke breed of Central Asia, and indigenous breeds of East Asia and Southern Europe, as well as between horses of the Irmen culture and modern horses of local breeds in Northern Europe, may reflect the migration routes of bearers of these cultures after their disintegration in the region under study. It also characterizes features of the formation of these breeds in ancient times. However, nuclear genetic markers should also be investigated to corroborate these hypotheses.

Key words: ancient DNA; mitochondrial genome; phylogenetics; domestic horse; Bronze Age; Sargarinsko-Alexeevskaya culture; Irmen culture

For citation: Kusliy M.A., Yurlova A.A., Vorobyeva N.V., Proskuryakova A.A., Demin M.A., Sitnikov S.M., Zharonkin V.N., Onishchenko S.S., Kasparov A.K., Tupikin A.E., Graphodatsky A.S., Molodtseva A.S., Tishkin A.A. Genetic diversity of horses of the Sargarinsko-Alexeevskaya and Irmen cultures of the Ob-Irtysh region of Western Siberia and their genetic proximity to modern horses of indigenous breeds. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):451-460. doi 10.18699/vjgb-26-49

Funding. The research was supported by the Russian Science Foundation, project 22–18–00470-П "The World of Ancient Nomads of Inner Asia: Interdisciplinary Research on Material Culture, Sculptures and Economy", https://rscf.ru/project/22–18–00470/.

Генетическое разнообразие лошадей саргаринско-алексеевской и ирменской культур Западной Сибири и их генетическое родство с современными лошадьми аборигенных пород

М.А. Куслий [1], А.А. Юрлова [1] ✉, Н.В. Воробьева [1], А.А. Проскурякова [1], М.А. Демин [2], С.М. Ситников [3], В.Н. Жаронкин [4], С.С. Онищенко [5], А.К. Каспаров [6], А.Е. Тупикин [7], А.С. Графодатский [1], А.С. Молодцева [1], А.А. Тишкин [8]

[1] Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Алтайский государственный педагогический университет, Барнаул, Россия
[3] Автономная некоммерческая организация «Алтайское археологическое общество», Барнаул, Россия
[4] Управление Роскомнадзора по Кемеровской области Кемерово, Россия
[5] Комитет по охране объектов культурного наследия Кузбасса, Кемерово, Россия
[6] Институт истории материальной культуры Российской академии наук, Санкт-Петербург, Россия
[7] Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия
[8] Алтайский государственный университет, Барнаул, Россия
✉ annush@mcb.nsc.ru

Аннотация. В современной науке для решения сложных задач всё чаще используется мультидисциплинарный подход. Так, молекулярная генетика помогает нам понять не только такие биологические процессы, как эволюция и видообразование, но и проясняет многочисленные исторические вопросы о направлениях миграций народов, степени взаимопроникновения синхронных культур и их преемственности. В частности, изучение филогенетических взаимоотношений домашних животных позволяет детализировать картину взаимодействия носителей разных археологических культур. Внутренняя Азия и сопредельные территории в бронзовом веке характеризовались активными миграциями людей и быстрым распространением производящего хозяйства, в том числе животноводства. В нашей работе были изучены филогенетические паттерны лошадей двух важных культур Обь-Иртышья Западной Сибири эпохи бронзы: саргаринско-алексеевской и ирменской, а также степень их генетической близости с лошадьми более ранних (андроновской, елунинской) и более поздних культур (херексуров и "оленных" камней, бийкенской, быстрянской и пазырыкской) региона и сопредельных территорий. Данные, полученные после секвенирования и анализа митохондриальных геномов, свидетельствуют о различиях в митохондриальном генофонде лошадей саргаринско-алексеевской и ирменской культур юга Западной Сибири, что подчеркивает своеобразие митохондриального генетического разнообразия исходных табунов лошадей вышеперечисленных культур и отсутствие между ними тесных контактов в сфере разведения лошадей. Было показано пересечение митохондриальных генофондов лошадей культуры херексуров и "оленных" камней Монголии и андроновской культуры. Также нами установлена преемственность между многими полученными гаплотипами лошадей раннего, развитого, позднего бронзового века, раннего железного века с территории юга Западной Сибири, что свидетельствует о сохранении большой части разнообразия материнского генофонда домашних лошадей в регионе на протяжении нескольких историко-культурных периодов. Близость митохондриальных гаплотипов между лошадьми саргаринско-алексеевской культуры и современными лошадьми ахалтекинской породы Центральной Азии и аборигенных пород Восточной Азии и Южной Европы, а также между лошадьми ирменской культуры и современными лошадьми местных пород Северной Европы может указывать на пути миграции носителей этих культур после их распада на территории рассматриваемого региона, а также характеризовать особенности формирования этих пород в древности. Однако для подтверждения этих гипотез необходимо исследовать ядерные генетические маркеры.

Ключевые слова: древняя ДНК; митохондриальный геном; филогенетика; домашняя лошадь; бронзовый век; саргаринско-алексеевская культура; ирменская культура

Introduction

In the context of equine genetic research, the Bronze Age (BA) has been less studied than the Iron Age (Keyser-Tracqui et al., 2005; Dawei et al., 2007; Cai et al., 2009; Lei et al., 2009; Cieslak et al., 2010; Benecke et al., 2017; Fages et al., 2019; Vorobieva et al., 2020; Kusliy et al., 2021; Librado et al., 2021; Kusliy, 2023). However, some important data have already been obtained. Regarding the territory of Inner Asia, it has been found that horses of the Sintashta culture, which is associated with the first ritual burials of horses together with parts of chariots with spokes, made a significant contribution to the equine gene pool of many subsequent Central Asian cultures, including the Khereksur and "Deer Stone" culture of Mongolia (Fages et al., 2019) and the Altai Biykenskaya culture (Librado et al., 2021). It should be noted that the Sintashta culture was the earliest stage in the development of the Andronovo cultural and historical community, which spread across the territory of the Southern Urals, the south of Western Siberia, Kazakhstan, and the western part of Central Asia (Zubova et al., 2014).

In this study, we examined the mitogenome diversity in horses of the Irmen and Sargarinsko-Alexeevskaya cultures, widespread in the Ob-Irtysh region of Western Siberia in the Late Bronze Age (Grushin, 2020). Bone materials for genetic analysis had been obtained from the sites of Chekanovsky Log-I (Demin, Sitnikov, 1998), Barsuchikha-IV of the

Sargarinsko-Alexeevskaya culture, Kaltyshino V (Kovtun, 2022), Barsuchikha-IV, Gusinaya Lyaga-1 (Demin, 2015) of the Irmen culture. Researchers of some of these sites had noted the important role of horses in various spheres of life of the inhabitants of these sites; for example, the majority of bone remains of domestic animals found at the Gusinaya Lyaga-1 site belonged to horses (Demin, 2015). Materials from both cultures studied were identified at the archaeological sites of Barsuchikha-IV and Gusinaya Lyaga-1, so the researchers classified these complexes as mixed-type sites (Sitnikov, Gel'mel', 2017). The results of our study highlight differences in the horse maternal genetic lineages of the Sargarinsko-Alexeevskaya and Irmen cultures, which allows us to hypothesize different maternal origins of horses of these cultures and the absence of intensive exchange of horses between them. It is also shown that the horses of mixed sites on the border between these cultures are closer in mitogenomes to the Sargarinsko-Alexeevskaya culture horses. However, this hypothesis requires confirmation based on more sample sets from each culture.

Thus, according to archaeological data (Grigor'ev, 2018; Kovalevsky, 2020), the Andronovo culture made a great contribution to the formation of the Irmen and Sargarinsko-Alexeevskaya cultures. We included samples from Altai and Kazakhstan sites of this culture in our study in order to trace possible continuity. Archaeological data provide no comprehensive answer to questions about other ancestral cultures or the continuity of the cultures under consideration in relation to later cultures of the region (Sitnikov, 2013; Papin et al., 2018; Popova, 2019). In order to strengthen the evidentiary base for resolving these issues, we supplemented our materials with samples of the Early BA Eluninskaya culture, formed in the Forest-Steppe Altai as a result of the migrations of Indo-European tribes to the eastern and southern regions of Eurasia (Kiryushin, 2002; Grushin, 2019), and from synchronous and later sites of the Altai and Mongolia (Tishkin, 2007).

Materials and methods

Materials studied. Our material included two bone samples of the Eluninskaya culture from Altai (Berezovaya Luka site), two bone samples of the Andronovo culture from Altai and Kazakhstan (Chekanovsky Log-2 site (Altai), Tasty-Bulak site (Kazakhstan)), six bone samples from Altai mixed sites of the Irmen and Sargarinsko-Alexeevskaya cultures (Barsuchikha-IV and Gusinaya Lyaga-1 sites), two bone samples of the Irmen culture from southern Western Siberia (Kaltyshino V sites), four bone samples of the Sargarinsko-Alexeevskaya culture of Altai (Chekanovsky Log-I site), and one bone sample of the Bystryanskaya culture (Manzhikha-2 site (Altai)). Detailed information about the samples studied is given in Supplement 1[1]. Supplement also provides information on radiocarbon dates of samples whose archaeological dating was ambiguous. Radiocarbon dating was performed at the AMS Golden Valley (Novosibirsk, Russia). Accurate dating of a sample is necessary to prevent errors in the cultural assignment of samples.

[1] Supplements 1–3 are available at:
https://vavilovj-icg.ru/download/pict-2026-30/appx25.zip

Ancient mitogenome sequencing. All experiments were performed at the Institute of Molecular and Cellular Biology (IMCB SB RAS), Siberian Branch of the Russian Academy of Sciences, in a special ancient DNA laboratory, in accordance with the basic authenticity criteria for ancient DNA research (Willerslev, Cooper, 2005), which presently remain relevant: (1) The work areas for ancient DNA experiments before the PCR stage and after it were separated. (2) Laboratory rooms for experiments with ancient DNA before the PCR stage were equipped with a ventilation system that created elevated pressure; experimenters wore special laboratory suits and two pairs of gloves, the outer one of which was constantly changed; the gloves and all work surfaces were constantly wiped with decontaminants (DNArid (Biomedical Innovations), Dezomax (Maxima LLC)) before, during, and after work; ultraviolet lamps (30 W) and an air sterilizer (recirculator ORB-2N (POZIS)) were also used to decontaminate work surfaces; tubes with samples and reagents were opened only inside laminar-flow cabinets (BAVnp-01-"Laminar-S"-1.5 LORICA (LANSYSTEMS)) with the ventilation turned on. (3) We added negative controls at the stages of isolating ancient DNA and preparing genomic libraries, which we conducted through all processes of the experiment to verify the absence of cross-contamination and foreign contamination in the samples. (4) At the stage of sequencing data analysis, the sizes and base deamination profiles of ancient DNA fragments were evaluated (subsection Statistics and authenticity of sequencing data of the Results and Discussion section). (5) The ancient origin of the samples was proven based on the archaeological context of related materials and direct and indirect radiocarbon dating (Supplement 1).

The preparation of bone samples for DNA extraction and the DNA extraction process itself, as well as the method used to prepare double-stranded, dual-index libraries for sequencing, are described in detail in (Kusliy et al., 2021).

Two rounds of library enrichment were performed using hybridization with biotinylated modern mtDNA of *Equus caballus*, immobilized on Dynabeads M-280 Streptavidin magnetic particles (Life Technologies, USA). The method reported in (Maricic et al., 2010) was modified as follows: (1) When preparing biotinylated samples, we ligated a double-stranded adapter with a 3′ single 'T' nucleotide overhang (5′-CCTGCCTCGGATGTCCTTGAT-3′) to modern horse DNA fragments using a TruSeq Nano DNA Sample Preparation Kit (Illumina) according to manufacturer's recommendations. (2) Modern horse mitogenome fragment libraries were amplified using biotinylated primers (Biotin-5′-CCTGCC TCGGATGTCCTTGAT-3′). (3) Biotinylated probes were immobilized onto magnetic beads with streptavidin according to the Dynabeads™ Streptavidin Trial protocol using sodium citrate saline (SSC) as binding and washing buffer. (4) The final purification of magnetic particles after the enrichment procedure was carried out in SSC (3 times with 100 μL of 2× SSC at 65 °C for 5 min, 2 times with 100 μL of 0.2× SSC at room temperature). (5) At the final stage of the procedure, 20 cycles of amplification of enriched libraries were conducted. Amplification of the enriched libraries was performed in a volume of 50 μL containing 1x Phusion HF Buffer, 0.2 mM of each

The GenBank accession numbers of the obtained nucleotide sequences

Sample name	GenBank accession numbers
Bar4-1	PV926031
BerL-3-2	PV926032
BerL-4	PV926033
Chek1-1	PV926035
Chek1-2	PV926036
Chek1-3	PV926037
Chek1-4	PV926038
Chek1-6	PV926039
Gus1-2	PV926040
Gus1-3	PV926041
Gus1-4	PV926042
Gus1-5	PV926043
Gus1-6	PV926044
Kalt5-2	PV926045
Kalt5-3	PV926046
Man2-1	PV926047
Tas-1	PV926048

dNTP, 1 μM of each primer for library fragment adapters (SuD Nano DNA Library Prep Kit for Illumina), and 1 U of Phusion DNA polymerase. The PCR program was as follows: 30 s initial denaturation at 95 °C; 20 cycles, each of which included 20 s denaturation at 98 °C, 20 s primer annealing at 65 °C, 20 s elongation at 72°C; and final 5 min elongation at 72 °C.

Quantification of the obtained libraries was performed using a Qubit 4 Fluorometer (Invitrogen) and a Qubit dsDNA HS Assay Kit, according to manufacturer's recommendations. To check for contamination at the stages of DNA isolation, library preparation, and amplification, blanks were used that showed the absence of target DNA during final data analysis. Paired-end sequencing of the enriched libraries was carried out on the MiSeq platform (Illumina, USA) using a MiSeq v2 Reagent Kit (300-cycles, 2x150bp). Library sequencing was performed at Genomics Core Facility (ICBFM SB RAS, Novosibirsk).

Sequence data analysis. Secondary data analysis performed in PALEOMIX BAM Pipeline v1.3.2 (Schubert et al., 2014) is described in detail in (Kusliy et al., 2021).

Phylogenetic analysis. Multiple alignment of the mitogenome consensus sequences was conducted using the MAFFT v1.4.0 multiple sequence alignment program (Katoh et al., 2002) (plugin MAFFT in Geneious Prime v2024.0.4). PartitionFinder v2.1.1 (Lanfear et al., 2012) was used to select the best-fit partitioning schemes and models of molecular evolution for phylogenetic analyses. In the alignment of horse mitochondrial genome sequences, five partitions were identified with the following evolutionary models of nucleotide substitutions: HKY+I for the second codons of protein-coding genes; HKY+G+I for RNA coding genes, first and third codons of protein-coding genes; GTR+G+I for hypervariable regions. The Bayesian phylogenetic tree was constructed with the MrBayes v.3.2.6 program (Ronquist, Huelsenbeck, 2003) using the above partitioning schemes and evolutionary models and the following parameter values: 8 million generations of Markov chain Monte Carlo, sampling frequency 800; the first 25% of trees discarded. The tree visualization was performed with the FigTree v1.4.4. program ("http://tree.bio.ed.ac.uk/software/figtree/).

Haplotype diversity analysis. The haplotype median joining network and haplogroup map were generated in the PopART population genetics software version 1.7 (Bandelt et al., 1999; Leigh, Bryant, 2015) with standard parameter values.

Analysis of genetic differentiation of populations. The F_{ST} values, reflecting the measure of population differentiation (10,000 permutations), as well as nucleotide diversity values, were obtained with the Arlequin v3.5.2.2 integrated software package (Excoffier, Lischer, 2010).

Data availability. All bone samples were obtained from the collection in the repository of the Institute of History and International Relations of Altai State University (Barnaul, Russia) (collection manager Dr. Alexey A. Tishkin). The described study complies with all relevant regulations. The GenBank accession numbers of the mitogenome consensus sequences are given in Table.

Results and discussion

Statistics and authenticity of sequencing data. The ancient origin of the obtained sequences is proven by the profiles of nucleotide misincorporation in DNA fragments of the sequencing libraries obtained in the MapDamage v2.2.0 computational framework (Jónsson et al., 2013). The percentage of nucleotide substitutions therein increases towards the ends of the fragments due to the presence of protruding ends in the fragments of ancient DNA (Green et al., 2006; Sawyer et al., 2012). The distribution graphs of the sizes of the sequenced libraries also confirm the data authenticity, since the average fragment size of each library is extremely small (less than 100 bp), which is typical for degraded ancient DNA (Sawyer et al., 2012). The graphs of postmortem damage and fragmentation patterns are shown in Supplement 2. The statistical parameter values of the obtained sequencing data are given in Supplement 3. The frequency of deamination at the ends of DNA fragments, which is on the average 2–3 times higher than in the middle of the fragment (Supplement 3), also points to the ancient origin of the DNA under study.

Haplotyping and haplogroup diversity assessment among the ancient horses studied. We compared the consensus sequences of the mitogenomes of the ancient horses studied here and in previous works (Vorobieva et al., 2020; Kusliy et al., 2021) with each other, determined their characteristic nucleotide variants relative to the reference sequence and assigned them to previously determined haplogroups (Achilli et al., 2012). Based on this data, a map of the geographical distribution of these mitochondrial haplogroups in Inner and Central Asia in ancient times was constructed. The map is shown in Figure 1.

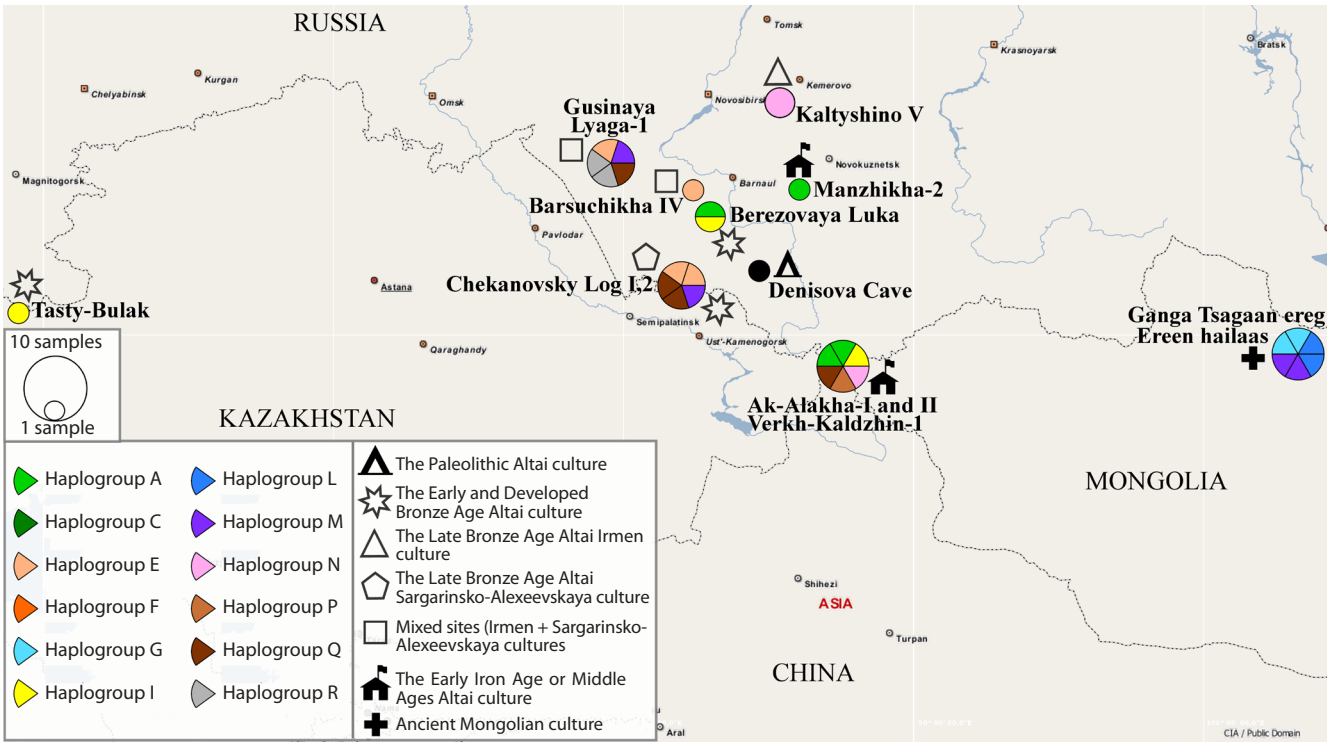


Fig. 1. Map of the geographical distribution of haplogroups of ancient and medieval horses in Inner Asia and adjacent territories studied here and earlier. The circle size is proportional to the number of haplotypes obtained. The sector colors highlight different haplogroups. The symbols show the belonging of the sites to a time range or archaeological culture. The legend is shown in the inset.

As can be seen from the figure, the haplotype of the Paleolithic Altai horse is not found among the studied ancient or medieval horses in Inner Asia and adjacent territories.

The most ancient haplotypes of domestic horses, namely those of the Early and Developed Bronze Age (Andronovo culture (Chekanovsky Log-2, purple sector; and Tasty-Bulak sites) and Eluninskaya culture (Berezovaya Luka site)), belong to haplogroups A, I, and M. These ancient haplogroups were also found in the Altai Krai (Irmen + Sargarinsko-Alexeevskaya cultures) and Mongolia horse groups in the Late Bronze Age, and in the mountain and steppe Altai in the Early Iron Age, which may point to a certain genetic continuity of horses of the corresponding cultures.

The figure also shows significant differences in the composition of haplogroups between the ancient horses of Altai and Mongolia, which share only haplogroup M, present in the Andronovo group. In contrast to the small overlap in mitochondrial genetic diversity shown between ancient horses of the Mongolia and Altai regions, modern horses of these regions show a much greater degree of genetic closeness (Kusliy et al., 2023). This observation is most likely indicative of a common origin and the absence of close contacts between horses associated with the cultures of these territories after the Developed Bronze Age.

Turning attention to the Irmen and Sargarinsko-Alexeevskaya cultures of the Late Bronze Age, which are in the focus of this study, we note that the mixed-type sites in the northern Altai Krai are close in haplogroup diversity to the sites of the Sargarinsko-Alexeevskaya culture of the southern Altai Krai, and they both differ greatly in these indicators from the sites of the Irmen culture, located in the north-eastern direction. As can be seen from the map, the mitochondrial haplogroups of horses from the Irmen sites coincide only with the haplogroups of horses from the Pazyryk (Verkh-Kaldzhin-1) sites of Early Iron Age, located in the south of the Altai Republic. Since, according to archaeological data, the Irmen culture did not contribute to the formation of the Pazyryk culture, this closeness of haplogroups most likely traces the common origin of the horses of these cultures but not continuity between them. In order to determine the differences between the mitogenome haplotypes of the studied ancient and medieval horses of Inner Asia and the adjacent territories from each other, we built a median joining haplotype network (Fig. 2).

As can be seen from the network, the most ancient haplotypes of domestic horses in the sample (Early and Developed Bronze Age, brown) are closest to the haplotypes of Mongolian horses of the Late Bronze Age (blue) and haplotypes of Altai horses of Early Iron Age (green). The constructed network shows the proximity of horse haplotypes of the Sargarinsko-Alexeevskaya culture and mixed sites of this and Irmen cultures (pink and red, respectively) and their remoteness from horse haplotypes of the Irmen culture, which cluster together with haplotypes of Early Iron Age Pazyryk horses from the territory of the Altai Mountains. The resulting haplotype network also visualizes the differences between the haplotypes of the ancient Mongolian horses under study and most of the

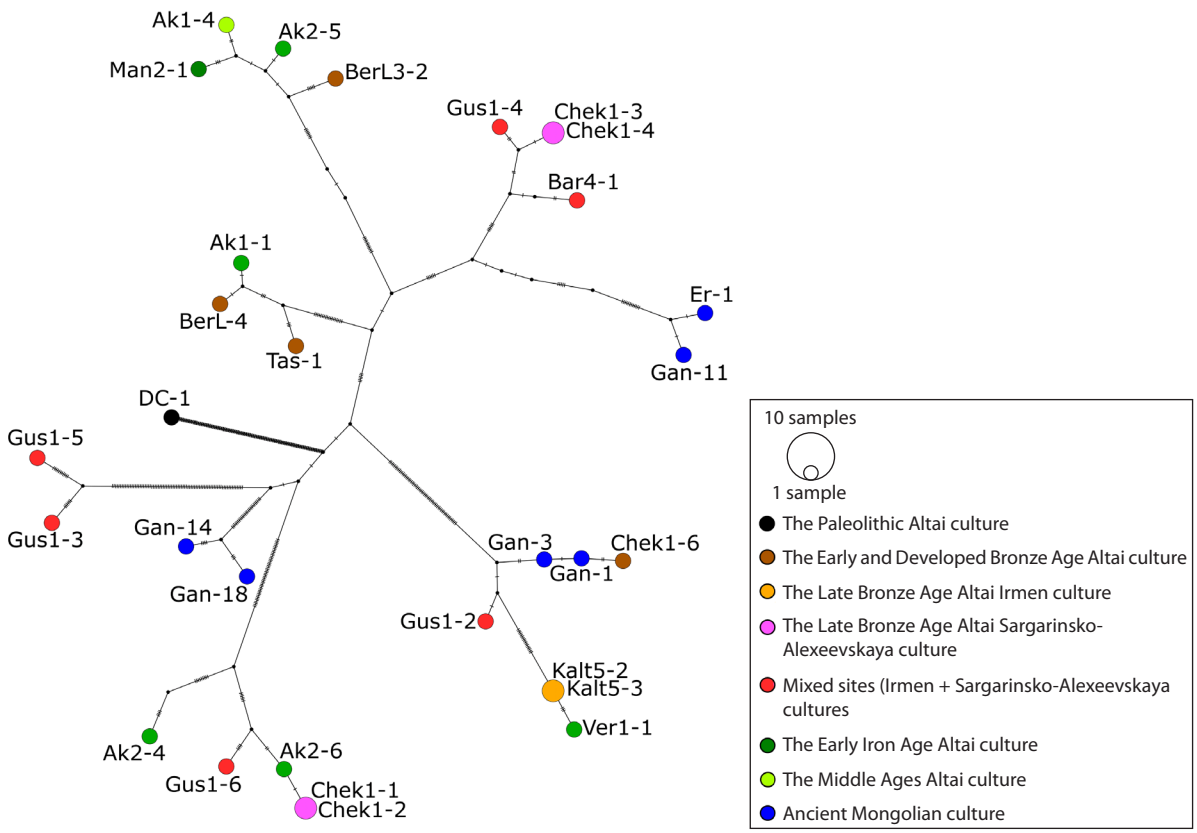


Fig. 2. Median joining haplotype network of studied ancient and medieval horses in Inner Asia and adjacent territories. The colors of the circles are associated with different semantic groups of horses (according to historical and cultural periods, belonging to a specific culture or region). The legend is shown in the inset. Lines on the branches of the network represent genetic variants that distinguish the ancestral haplotype from younger ones.

other haplotypes of the sample and the similarity of the Bronze Age Mongolian horses to the Andronovo horse, which is most likely related to their origin. The described figure shows a more detailed history of the studied ancient horse haplotypes, whose key features are outlined above.

Our analysis of population differentiation in the culture groups of Sargarinsko-Alexeevskaya (samples Chek1-1, Chek1-2, Chek1-3, Chek1-4) and Irmen (samples Kalt5-2, Kalt5-3) and in the group of mixed sites of these two cultures (samples Bar4-1, Gus1-2, Gus1-3, Gus1-4, Gus1-5, Gus1-6) revealed a high degree of differentiation between the Mixed and Irmen groups (F_{ST} = 0.24, p-value > 0.05) and a very high degree of differentiation (F_{ST} = 0.52, p-value > 0.05) between the Sargarinsko-Alexeevskaya and Irmen groups. Their F_{ST} analysis also showed an insignificant degree of differentiation between the Mixed and Sargarinsko-Alexeevskaya groups. However, since the sample sizes were small, and the p-values were unreliable (p-value > 0.05), this can only be considered an additional evidence.

Phylogeographic reconstructions. In order to see a more detailed picture of the relationships between haplotypes within haplogroups and to trace phylogenetic relationships to the present, we constructed a Bayesian phylogenetic tree based on mitogenome sequences of ancient, medieval, and modern horses from different regions of the world, including our samples and previously published ones (GenBank sequence database). The constructed tree is shown in Figure 3.

In the above discussion of phylogeographic reconstructions, only clades with high Bayesian posterior node probabilities are considered in the analysis of reliable phylogenetic relationships.

The constructed phylogenetic tree shows that haplogroup I represents the intersection of haplogroups of the Eluninskaya and Andronovo cultures of the Bronze Age. One of the Eluninskaya horses we studied is located basal to the clade of all other domestic horses of haplogroup A, which supports the origin of horses of this haplogroup from horses of this culture. However, since the Andronovo horses not included into our study could also occupy a basal position within this haplogroup, we cannot rule out the Andronovo origin. The Eluninskaya culture is known to form earlier (Early Bronze Age) than the Andronovo one (Developed Bronze Age). Most of the animal bones found in the Berezovaya Luka settlement (samples BerL-3.2, BerL-4) belong to domestic species (Kiryushin et al., 2011), but it has not been shown whether the horses of Berezovaya Luka site were domesticated or wild (hunted). Our data (Fig. 3, haplogroup I) clearly indicate that some horses of this settlement were domesticated, because the corresponding haplotypes are located within the clade of modern domestic horses. Since the development of animal

Fig. 3. Bayesian phylogenetic tree of ancient, medieval and modern horses from different regions of the world, constructed on the grounds of mitogenome sequences. The names of the samples consist of three parts, separated by an asterisk: (1) the geographical origin of the sample (the An prefix at the beginning denotes ancient and medieval samples, and the His prefix indicates historical horses); (2) registration number of the GenBank database or name of the sample and culture; (3) age of the sample. Letters A–R to the right of the clades correspond to the names of equine mitochondrial genome haplogroups according to the classification by Achilli et al. (2012). The colors of the sample names are associated with different semantic groups of horses (according to historical and cultural periods, belonging to a specific culture and/or region); the correspondence is described at the upper left of the figure. For better clarity, some clades of the tree have been collapsed. The Bayesian posterior probability of the tree topology is shown as numbers next to the tree nodes (probability less than 0.7 (confidence level) is highlighted in red). The numeral under the horizontal scale bar at the bottom of the figure indicates the number of nucleotide substitutions related to a segment of equal length on a tree branch.

husbandry in the Altai steppe zone is primarily associated with the bearers of the Eluninskaya culture, who migrated there at the beginning of the Early Bronze Age (Kiryushin et al., 2012), the South Siberian origin of some ancient domestic horses of Asia remains questionable. However, since the Eluninskaya and Andronovo cultures overlapped in time, and horses of the Andronovo culture spread very quickly to adjacent territories (Koryakova, Epimakhov, 2007; Lindner, 2020; Epimakhov, 2020), the question of local domestication or Andronovo origin of the Eluninskaya culture horses will remain relevant until whole genome data on horses of these cultures are obtained. Based on our data, we conclude that some domestic horses from both cultures are genetically very close. It should also be noted that, according to our results, the haplogroups of the Eluninskaya culture horses were found only among horses of Early Iron Age from the south of the Altai Republic (Biykenskaya, Pazyryk, Turkic cultures), and the haplogroups of the Andronovo culture were widespread not only among the Pazyryk horses of Altai, but also among the horses of the Late Bronze Age from the Ob-Irtysh region of Western Siberia (Irmen + Sargarinsko-Alexeevskaya cultures) and Mongolia (the Khereksur and "Deer Stone" culture). The closeness of the haplotypes of horses of the Khereksur and "Deer Stone" culture of Mongolia and the Andronovo culture, detected by us for the first time on the basis of mitogenome data, is most likely associated with continuity between the horses of these cultures. A horse of the mixed site of the Irmen and Sargarinsko-Alexeevskaya cultures, which occupied the basal position in clade M, is also quite close to the mitochondrial haplotypes under consideration (samples Gan-1, Gan-3, Chek1-6). The closeness between the haplotypes of domestic horses of the Andronovo culture, from which most domestic horses of Asia originated, and horses of later cultures of Mongolia and Siberia more likely points to their common origin from horses of the Andronovo cultural–historical community. The intersection of the mitogenome gene pools of horses of cultures later than Andronovo in Siberia and Mongolia is more likely to point to close contacts between archaeological cultures of these regions.

Within haplogroup I, the BerL-4 horse of the Eluninskaya culture, one of the most ancient in our sample occupies the closest position to horses of the Kustanai local breed of Kazakhstan and the indigenous Iranian breed. This observation looks consistent with the migration of part of the population of the Eluninskaya culture to the west after the arrival of the Andronovo population to their territory (Grushin, 2018).

Within haplogroup E (Fig. 3), which turned out to be one of the key haplogroups of the Sargarinsko-Alexeevskaya culture and mixed sites (Irmen+Sargarinsko-Alexeevskaya), the domestic horses of the latter-listed cultures we studied are located next to modern horses of the indigenous breeds of Italy (Maremmano), China (Jinjiang), and historical Yakutian horses. The last of these horse breeds migrated to Yakutia from southward regions together with the Yakut people several centuries ago (Librado et al., 2015). The genetic closeness we identified between the ancient domestic horses of the mixed sites we studied and the Yakut horses is also noticeable within haplogroup Q. There is evidence that the Maremmano horse breed was formed in the 8th century BCE by Etruscans (Giontella et al., 2020; Vernesi et al., 2004), an ancient people primarily of South European origin (Ghirotto et al., 2013; Vernesi et al., 2004). Previous studies have shown maternal genetic closeness between modern Maremmano horses and domestic horses of the Altai Biykenskaya culture of Early Iron Age (Vorobieva et al., 2020). The Jinjiang horse is a breed indigenous to Central and Southeast China; it formed at the end of the 1st millennium CE and developed with a small genetic infusion from foreign breeds (Ma et al., 2019). The similarity of mitochondrial haplotypes of horses of the Sargarinsko-Alexeevskaya culture, mixed sites, and modern domestic horses of indigenous breeds from different regions of China (Yanqi, Yimen, Yili, Sanhe) can also be traced within the Q and R haplogroups. The basal position of one of the horses from the mixed Irmen and Sargarinsko-Alexeevskaya culture site Barsuchikha-IV in relation to the other horses of clade E and clade M indicates that some horses of the Maremmano, Jinjiang, and Yakutian breeds might derive from ancient domestic horses of these cultures, with migration occurring in both the eastern and western directions.

As can be seen from the phylogenetic tree we constructed, the horses of the Sargarinsko-Alexeevskaya culture and the mixed sites of the Irmen and Sargarinsko-Alexeevskaya cultures are also very close in mitochondrial DNA to modern Akhal-Teke horses (haplogroup Q), one of the most ancient riding breeds of Central Asia (Szontagh et al., 2005).

Horses of the Irmen culture were assigned to clade N, which is mainly represented by modern horses of local breeds of Southern and Northern Europe and breeds formed on their basis. Within this haplogroup, horses of the Irmen culture are located in the subclade that also contains horses of Early Iron Age Pazyryk culture in Altai and modern horses of the English Shire breed and the American Saddlebred breeds. The English Shire breed is native to England (Stephens, Splan, 2013), and the American Saddlebred breed was developed on the basis of many riding and draft horse breeds of English and Spanish origin in North America in the late 18th to early 19th centuries (Regatieri et al., 2016). The described geographical distribution of haplotypes of clade N may reflect the predominant distribution of ancient haplotypes of this haplogroup in the western direction. It should be noted that clade N is not included in the haplogroup diversity of horses of the Sargarinsko-Alexeevskaya culture or mixed sites of both of the above-mentioned cultures under study.

The obtained data reflect the proximity of the mixed sites to the Sargarinsko-Alexeevskaya culture at the level of the gene pools of domestic horses, as well as differences in the mitogenome genetic diversity of domestic horses of the Sargarinsko-Alexeevskaya and Irmen cultures, but how strong they were can only be understood by expanding the sample of Irmen culture horses. The phylogenetic closeness of the mitogenome haplotypes of the studied ancient and modern horses from different regions of the world most likely points to features in the migration directions of bearers of the Irmen and Sargarinsko-Alexeevskaya cultures after their collapse.

Conclusions

The study of mitochondrial genomes of ancient horses of the Sargarinsko-Alexeevskaya and Irmen cultures in the south of Western Siberia, mixed sites of these cultures located on the border between their areas, older cultures (Eluninskaya and Andronovo) in Western Siberia and Kazakhstan, and later cultures of Western Siberia and Mongolia revealed the presence of horse haplogroups of the most ancient of the studied cultures among horses of the Earle Iron Age Mountains and Forest-steppe Altai cultures. Our phylogenetic reconstructions are first to show the closeness of haplotypes of Mongolian horses of the Khereksur and "Deer Stone" and Andronovo cultures on the grounds of mitogenome data. They have also identified differences between the horses of the Sargarinsko-Alexeevskaya and Irmen cultures. The horses of the mixed sites with ceramics of both of the aforementioned cultures are closer in mitochondrial DNA to the horses of the Sargarinsko-Alexeevskaya culture, some haplotypes of which, in turn, fell into the haplogroup with the highest content of Akhal-Teke horses. Other haplotypes of this culture turn out to be genetically close to horse haplotypes of ancient indigenous breeds of Southern Europe and China. In contrast, the horses of the Irmen culture show a closer relationship to the horses of the native breeds of Northern Europe. The differences between the mitochondrial gene pool of horses of the Sargarinsko-Alexeevskaya and Irmen cultures may be associated with the origin of the studied horses of these cultures from different stock populations, as well as with the absence of a strong exchange of horses between them, which most likely correlates with the different histories of the cultures under consideration. The close mitochondrial relationship between Bronze Age Mongolian and Siberian Andronovo horses is most likely connected with the origin of the former from the latter. Further study of horses of the cultures under study is required to understand how strong the differences in their mitochondrial gene pools were.

References

Achilli A., Olivieri A., Soares P., Lancioni H., Kashani B.H., Perego U.A., Nergadze S.G., … Silvestrelli M., Giulotto E., Pereira L., Bandelt H.-J., Torroni A. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc Natl Acad Sci USA*. 2012;109(7):2449-2454. doi 10.1073/pnas.1111637109

Bandelt H.J., Forster P., Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16(1):37-48. doi 10.1093/oxfordjournals.molbev.a026036

Benecke N., Pruvost M., Weber C. Horse skeletons: archaeozoological and genetic research. In: The Royal Mound of the Scythian Time Arzhan-2 in Tuva. Novosibirsk, 2017;250-256 (in Russian)

Cai D., Tang Z., Han L., Speller C.F., Yang D.Y., Ma X., Cao J., Zhu H., Zhou H. Ancient DNA provides new insights into the origin of the Chinese domestic horse. *J Archaeol Sci*. 2009;36(3):835-842. doi 10.1016/j.jas.2008.11.006

Cieslak M., Pruvost M., Benecke N., Hofreiter M., Morales A., Reissmann M., Ludwig A. Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS One*. 2010;5(12):e15311. doi 10.1371/journal.pone.0015311

Dawei C., Lu H., Chengzhi X., Shengnan L., Hui Z., Hong Z. Mitochondrial DNA analysis of Bronze Age horses recovered from Chifeng region, Inner Mongolia, China. *Prog Nat Sci*. 2007;17(5):544-550. doi 10.1080/10020070708541034

Demin M.A. Some results of the study and residential complexes on the territory of the Northern Kulunda. *Vestnik Altaiskoy Nauki = Vestnik of Altai Science*. 2015;(1):67-70 (in Russian)

Demin M.A., Sitnikov S.M. The Chekanovsky Log-I settlement is a new site of the Late Bronze Age in the Southwestern Altai. *Drevnosti Altaya. Izvestiya Laboratorii Arkheologii = Altai Antiquities. News of the Archaeology Laboratory*. 1998;6(3):43-54 (in Russian)

Epimakhov A.V. Radiocarbon arguments for the Abashevo origin of the Sintashta traditions in the Bronze Age. *Ural Hist J*. 2020;69(4):51-60. doi 10.30759/1728-9718-2020-4(69)-51-60 (in Russian)

Excoffier L., Lischer H.E.L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10(3):564-567. doi 10.1111/j.1755-0998.2010.02847.x

Fages A., Hanghøj K., Khan N., Gaunitz C., Seguin-Orlando A., Leonardi M., McCrory Constantz C., … Barrey E., Willerslev E., Outram A.K., Librado P., Orlando L. Tracking five millennia of horse management with extensive ancient genome time series. *Cell*. 2019;177(6):1419-1435.e31. doi 10.1016/j.cell.2019.03.049

Ghirotto S., Tassi F., Fumagalli E., Colonna V., Sandionigi A., Lari M., Vai S., Petiti E., Corti G., Rizzi E., De Bellis G., Caramelli D., Barbujani G. Origins and evolution of the Etruscans' mtDNA. *PLoS One*. 2013;8(2):e55519. doi 10.1371/journal.pone.0055519

Giontella A., Cardinali I., Lancioni H., Giovannini S., Pieramati C., Silvestrelli M., Sarti F.M. Mitochondrial DNA survey reveals the lack of accuracy in Maremmano horse studbook records. *Animals*. 2020;10(5):839. doi 10.3390/ani10050839

Green R.E., Krause J., Ptak S.E., Briggs A.W., Ronan M.T., Simons J.F., Du L., Egholm M., Rothberg J.M., Paunovic M., Pääbo S. Analysis of one million base pairs of Neanderthal DNA. *Nature*. 2006;444(7117):330-336. doi 10.1038/nature05336

Grigor'ev S.A. The settlement of Mochishche in the Southern Trans-Urals and the problms of periodization of the cultures of the Andronovo cultural-historical community. In: Vybornov A.A. (Ed.). XXI Ural Archaeological Meeting. Samara, 2018;100-102 (in Russian)

Grushin S.P. Cultural Interaction and Migration Processes in the Forest-Steppe Altai Region during the Early Bronze Age (22nd-18th Centuries BC). In: Eurasianism: Theoretical Potential and Practical Applications. Barnaul, 2018;53-57 (in Russian)

Grushin S.P. Eluninskaya culture and Utkul group of sites. In: Tishkin A.A. (Ed.). Histore of Altai. V 1. The most ancient era, antiquity and the Middle Ages. Barnaul, 2019;124-141 (in Russian)

Grushin S.P. The main stages of metallurgy development in Altai during the Bronze Age. In: Proceedings of the VI (XXII) All-Russian Archaeological Congress in Samara. V. I. Samara State Social and Pedagogical University. 2020;266-267 (in Russian)

Katoh K., Misawa K., Kuma K., Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059-3066. doi 10.1093/nar/gkf436

Keyser-Tracqui C., Blandin-Frappin P., Francfort H.-P., Ricaut F.-X., Lepetz S., Crubézy E., Samashev Z., Ludes B. Mitochondrial DNA analysis of horses recovered from a frozen tomb (Berel site, Kazakhstan, 3rd Century BC). *Anim Genet*. 2005;36(3):203-209. doi 10.1111/j.1365-2052.2005.01316.x

Kiryushin Yu.F. The Eneolithic and Early Bronze Age of the South of Western Siberia. Barnaul: Altai State University, 2002 (in Russian)

Kiryushin Yu.F., Grushin S.P., Tishkin A.A. Berezovaya Luka – The Settlement of Bronze Age Epoch in the Aleiskaya Steppe. V. II. Barnaul: Altai State University, 2011 (in Russian)

Kiryushin Yu.F., Kosincev P.A., Grushin S.P., Papin D.V. Cattle breeding in Altai during the Bronze Age. *Problems Archaeology, Ethnography, Anthropology Siberia Neighboring Territories*. 2012;18:180-183 (in Russian)

Koryakova L., Epimakhov A.V. The Urals and Western Siberia in the Bronze and Iron Ages. Cambridge University Press, 2007. doi 10.1017/CBO9780511618451

Kovalevsky S.A. The study of Late Bronze Age sites in the south of Western Siberia within the unified Irmen culture (the 1970s – the first half of the 1980s). *Theory Pract Archaeol Res*. 2020;(3):18-32. doi 10.14258/tpai(2020)3(31).-02 (in Russian)

Kovtun I.V. The period of transition from the Developed to the Late Bronze Age in the Lower Tom Region. *Uchenyye zapiski muzeya-zapovednika "Tomskaya Pisanitsa" = Research Notes of the Tomsk Pisanitsa Museum-Reserve*. 2022;(15):5-41. doi 10.24412/2411-7838-2022-15-5-41 (in Russian)

Kusliy M.A. Genetic diversity of ancient and modern horses of Altai and adjacent territories. Biodiversity and Ecology. Université Paul Sabatier – Toulouse III; Académie des sciences de Russie, 2023. Available: https://theses.hal.science/tel-04391016v1/file/2023TOU30188b.pdf

Kusliy M.A., Vorobieva N.V., Tishkin A.A., Makunin A.I., Druzhkova A.S., Trifonov V.A., Iderkhangai T.-O., Graphodatsky A.S. Traces of Late Bronze and Early Iron Age Mongolian horse mitochondrial lineages in modern populations. *Genes* (*Basel*). 2021;12(3):412. doi 10.3390/genes12030412

Kusliy M.A., Yurlova A.A., Neumestova A.I., Vorobieva N.V., Gutorova N.V., Molodtseva A.S., Trifonov V.A., … Iderkhangai T.-O., Kovalev A.A., Erdenebaatar D., Graphodatsky A.S., Tishkin A.A. Genetic history of the Altai breed horses: From ancient times to modernity. *Genes* (*Basel*). 2023;14(8):1523. doi 10.3390/genes14081523

Lanfear R., Calcott B., Ho S.Y.W., Guindon S. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol*. 2012;29(6):1695-1701. doi 10.1093/molbev/mss020

Lei C.Z., Su R., Bower M.A., Edwards C.J., Wang X.B., Weining S., Liu L., Xie W.M., Li F., Liu R.Y., Zhang Y.S., Zhang C.M., Chen H. Multiple maternal origins of native modern and ancient horse populations in China. *Anim Genet*. 2009;40(6):933-944. doi 10.1111/j.1365-2052.2009.01950.x

Leigh J.W., Bryant D. POPART: full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6(9):1110-1116. doi 10.1111/2041-210X.12410

Librado P., Der Sarkissian C., Ermini L., Schubert M., Jónsson H., Albrechtsen A., Fumagalli M., … Nielsen R., Willerslev E., Kantanen J., Prokhortchouk E., Orlando L. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc Natl Acad Sci USA*. 2015;112(50):E6889-E6897. doi 10.1073/pnas.1513696112

Librado P., Khan N., Fages A., Kusliy M.A., Suchan T., Tonasso-Calvière L., Schiavinato S., … Kroonen G.J., Kristiansen K., Wincker P., Outram A., Orlando L. The origins and spread of domestic horses from the Western Eurasian steppes. *Nature*. 2021;598(7882):634-640. doi 10.1038/s41586-021-04018-9

Lindner S. Chariots in the Eurasian Steppe: a Bayesian approach to the emergence of horse-drawn transport in the early second millennium BC. *Antiquity*. 2020;94(374):361-380. doi 10.15184/aqy.2020.37

Ma H., Wang S., Zeng G., Guo J., Guo M., Dong X., Hua G., Liu Y., Wang M., Ling Y., Ding X., Zhao C., Wu C. The origin of a coastal indigenous horse breed in China revealed by genome-wide SNP data. *Genes* (*Basel*). 2019;10(3):241. doi 10.3390/genes10030241

Maricic T., Whitten M., Pääbo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*. 2010;5(11):e14004. doi 10.1371/journal.pone.0014004

Papin D.V., Stepanova N.F., Fedoruk A.S. Late Bronze Age ceramics from a steppe region between the Ob and Irtysh rivers as a source for reconstructing ethnocultural interaction processes. *Vestnik Arheologii, Antropologii i Etnografii*. 2018;(3):19-31. doi 10.20874/2071-0437-2018-42-3-019-031 (in Russian)

Popova B.S. The monuments of the Irmen culture of the Lower Tom river area. The historiographical review. *Tomsk State Univ J Hist*. 2019;57:163-169. doi 10.17223/19988613/57/27 (in Russian)

Regatieri I.C., Eberth J.E., Sarver F., Lear T.L., Bailey E. Comparison of DMRT3 genotypes among American Saddlebred horses with reference to gait. *Anim Genet*. 2016;47(5):603-605. doi 10.1111/age.12458

Ronquist F., Huelsenbeck J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19(12):1572-1574. doi 10.1093/bioinformatics/btg180

Sawyer S., Krause J., Guschanski K., Savolainen V., Pääbo S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One*. 2012;7(3):e34131. doi 10.1371/journal.pone.0034131

Schubert M., Jónsson H., Chang D., Der Sarkissian C., Ermini L., Ginolhac A., Albrechtsen A., … Nielsen R., Excoffier L., Willerslev E., Shapiro B., Orlando L. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci USA*. 2014;111(52):E5661-E5669. doi 10.1073/pnas.1416991111

Sitnikov S.M. On the origins and the cultural-historical contacts of Sargara-Alexeevo population. In: Beisenov A.Z. (Ed.). Begazy-Dandybai culture of steppe Eurasia. Almaty, 2013;417-425 (in Russian)

Sitnikov S.M., Gel'mel' Yu.I. Some results of the research into 'Gusinaya Lyaga-1' settlement. *Theory Pract Archaeol Res*. 2017;(1):49-61. doi 10.14258/tpai(2017)1(17).-04 (in Russian)

Stephens T.D., Splan R.K. Population history and genetic variability of the American Shire horse. *Anim Genet Resour*. 2013;52:31-38. doi 10.1017/S2078633613000052

Szontagh A., Bán B., Bodó I., Cothran E.G., Hecker W., Józsa C., Major Á. Genetic diversity of the Akhal-Teke horse breed in Turkmenistan based on microsatellite analysis. In: Bodo I., Alderson L., Langlois B. (Eds). Conservation Genetics of Endangered Horse Breeds. Wageningen Academic Publishers, 2005;123-128

Tishkin A.A. Making of periodizational and cultural and chronological schemes: historical experience and modern conception of study of ancient and medieval nations of Altay. Barnaul: Altai State University, 2007 (in Russian)

Vernesi C., Caramelli D., Dupanloup I., Bertorelle G., Lari M., Cappellini E., Moggi-Cecchi J., Chiarelli B., Castrì L., Casoli A., Mallegni F., Lalueza-Fox C., Barbujani G. The Etruscans: A population-genetic study. *Am J Hum Genet*. 2004;74(4):694-704. doi 10.1086/383284

Vorobieva N.V., Makunin A.I., Druzhkova A.S., Kusliy M.A., Trifonov V.A., Popova K.O., Polosmak N.V., Molodin V.I., Vasiliev S.K., Shunkov M.V., Graphodatsky A.S. High genetic diversity of ancient horses from the Ukok Plateau. *PLoS One*. 2020;15(11):e0241997. doi 10.1371/journal.pone.0241997

Willerslev E., Cooper A. Ancient DNA. *Proc Biol Sci*. 2005;272(1558):3-16. doi 10.1098/rspb.2004.2813

Zubova A.V., Chikisheva T.A., Pozdnyakov D.V. Anthropological aspects of the genesis of representatives of the Andronovo cultural and historical community. In: Molodin V.I., Epimakhov A.V. (Eds). The Aryans in the Eurasian Steppes: the Bronze and Early Iron Ages in the Steppes of Eurasia and Contiguous Territories. Barnaul: Altai State University, 2014;541-554 (in Russian)

Conflict of interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received October 7, 2025. Revised December 28, 2025. Accepted January 1, 2026.

HUMAN GENETICS
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):461-469

doi 10.18699/vjgb-26-50

Toxic metals and genetic polymorphism in indigenous populations of northern Asia and America

B.A. Malyarchuk, N.V. Pokhilyuk

Institute of Biological Problems of the North of the Far Eastern Branch of the Russian Academy of Sciences, Magadan, Russia
malbor@mail.ru

Abstract. The polymorphism of genes encoding enzymes involved in heavy metal metabolism was analyzed in indigenous Siberian populations based on the mercury, lead, and cadmium contents in the blood of Canadian Inuit carrying different genotypes. Additionally, we examined the polymorphism of genetic loci associated with sensitivity to arsenic exposure in indigenous Siberian populations using data on inorganic arsenic content in the urine of indigenous Andean populations who had consumed drinking water with elevated arsenic levels for thousands of years. A population genetic approach was used to seek genetic markers of toxic metal exposure in humans by analyzing genetic differences between populations living in different natural environments and under different conditions of toxic element contamination. Statistically significant differences were primarily observed between indigenous populations in Northeast Siberia (Siberian Eskimo (Yupik), Chukchi, and Koryaks) and samples from the central (Evens, Evenki, and Yakuts) and southern (Altaians, Shors, and Buryats) regions of Siberia. The maximum population branch statistics (*PBS*) values, which indicate the probable effect of selection on genetic loci sensitive to mercury exposure, were identified in seven gene loci: *MTHFR* (rs2274976 and rs1801131), *GPX4* (rs713041), *ABCB1* (rs1128503), *AHR* (rs2066853), *TXNRD2* (rs5748469), and *SEPHS2* (rs1133238). Loci rs713041 (*GPX4*), rs7483 (*GSTM3*), and rs2282143 (*SLC22A1*) can be considered genetic markers of lead exposure. Loci rs2274976 (*MTHFR*) and rs1056836 (*CYP1B1*) provide information about cadmium distribution in blood. It was found that protective variants of the *AS3MT* gene polymorphism are widespread (65.8 %) in the indigenous populations of Northeast Siberia. This is despite the lack of information regarding the long-term consumption of arsenic-contaminated drinking water by indigenous peoples along the Chukotka and Priokhotye coasts. It is hypothesized that seafood, which constitutes the core of the traditional "Arctic" diet of the indigenous populations inhabiting the coastal regions of the northern seas, may potentially be a significant source of arsenic and other toxic elements in Northeast Siberia. Further molecular, biochemical, and toxicological studies are necessary to elucidate the mechanisms by which toxic metals impact the genetic structure of indigenous populations in the Far North over long periods of time.

Key words: genetic markers; genetic differentiation; human populations; Siberia; mercury; lead; cadmium; arsenic

For citation: Malyarchuk B.A., Pokhilyuk N.V. Toxic metals and genetic polymorphism in indigenous populations of northern Asia and America. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):461-469. doi 10.18699/vjgb-26-50

Acknowledgements. The authors are grateful to A.N. Litvinov (IBPN FEB RAS) for assistance in this study.

Токсичные металлы и генетический полиморфизм у коренного населения Севера Азии и Америки

Б.А. Малярчук, Н.В. Похилюк

Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, Россия
malbor@mail.ru

Аннотация. В настоящей работе у коренного населения Сибири проанализирован полиморфизм генов, кодирующих ферменты метаболизма тяжелых металлов, на основе данных о содержании ртути, свинца и кадмия в крови у канадских эскимосов – носителей различных генотипов; изучен также полиморфизм генетических локусов, связанных с чувствительностью к воздействию мышьяка по данным о содержании неорганического мышьяка в моче у коренного населения Анд, длительное время (в течение тысячелетий) потребляющего питьевую воду с повышенным содержанием мышьяка. Для поиска генетических маркеров воздействия токсичных металлов на организм человека использован популяционно-генетический подход, направленный на анализ генетических различий между популяциями, проживающими в разных условиях природной среды (и в разных условиях загрязнения токсичными элементами). Статистически значимые различия главным образом обнаружены между коренным населением Северо-Восточной Сибири (эскимосы, чукчи и коряки) и выборками из центральной (эвены, эвенки, якуты) и южной (алтайцы, шорцы, буряты) частей

Сибири. Максимальные значения статистики *PBS* (population branch statistics), свидетельствующие о вероятном действии отбора на генетические локусы чувствительности к воздействию ртути, выявлены в семи локусах генов *MTHFR* (rs2274976 и rs1801131), *GPX4* (rs713041), *ABCB1* (rs1128503), *AHR* (rs2066853), *TXNRD2* (rs5748469) и *SEPHS2* (rs1133238). К генетическим маркерам воздействия свинца могут быть отнесены локусы rs713041 (*GPX4*), rs7483 (*GSTM3*) и rs2282143 (*SLC22A1*). В отношении распределения кадмия в крови информативны локусы rs2274976 (*MTHFR*) и rs1056836 (*CYP1B1*). Обнаружено, что в популяциях коренного населения Северо-Востока Сибири с высокой частотой (65.8 %) распространены варианты полиморфизма гена *AS3MT*, защищающие от воздействия неорганического мышьяка, хотя сведения о долговременном употреблении коренными жителями побережий Чукотки и Приохотья питьевой воды, загрязненной мышьяком, отсутствуют. Предполагается, что вероятным источником мышьяка, а также других токсичных элементов на Северо-Востоке Сибири могут быть морепродукты, являющиеся основой традиционной «арктической» диеты коренного населения прибрежных районов северных морей. Для прояснения механизмов долговременного воздействия токсичных металлов на генетическую структуру популяций коренного населения Крайнего Севера необходимы дальнейшие молекулярно-генетические, биохимические и токсикологические исследования.
Ключевые слова: генетические маркеры; генетическая дифференциация; популяции человека; Сибирь; ртуть; свинец; кадмий; мышьяк

Introduction

Toxic metals such as mercury, lead, cadmium, and arsenic (a semimetal) are among the most widespread environmental pollutants. They are also toxic to humans and pose a health threat. The pollution of ecosystems in the Far North is of particular concern. As shown in a number of studies, various pollutants accumulate there and enter the human body through water, food, and air (Fernandez-Llamazares et al., 2020; Basu et al., 2022; Adlard et al., 2024). The traditional diets of indigenous peoples in the far northern regions of Asia and America, particularly in coastal areas, consist of foods (e. g., marine mammals, fish, and shellfish) that contain high levels of contaminants (Becker, 2000; Gamov et al., 2022). The industrial regions of Siberia also face the problem of heavy metal pollution, which endangers human health.

Despite the ongoing biomonitoring of pollutants in human biological tissues, primarily hair and blood, the interpretation of its data may be complicated by genetic differences between human populations living in different natural and climatic zones. Meanwhile, few population genetic studies are being conducted to identify genetic markers indicative of the impact of pollutants on the body. The Inuit of Canada and Greenland are the most studied populations in this regard (Ayotte et al., 2011; Ghisari et al., 2013; Parajuli et al., 2018, 2021).

The obtained data suggest that blood pollutant levels are associated with polymorphisms in genetic loci for several biological pathways, including folic acid, lipid, and xenobiotic metabolism. For instance, genetic variants linked to increased or decreased concentrations of heavy metals (mercury, cadmium, and lead) and organic pollutants (4,4′-DDE and PCB-153) in the blood have been identified in Canadian Inuit populations (Parajuli et al., 2018, 2021). It has been established that the indigenous populations of the Andes in Argentina, Chile, and Bolivia have genetically adapted to the toxic effects of drinking water with high concentrations of arsenic over thousands of years (Engström et al., 2011; Apata et al., 2017; De Loma et al., 2022). As a result, these populations have experienced a significant increase in the frequency of protective variants that encode enzymes metabolizing and eliminating arsenic from the body more efficiently.

In light of this, studies of the distribution of genetic variants related to the metabolism of toxic metals in other ethnic groups are relevant. Here we analyze the frequency distribution of polymorphic variants of genes involved in toxic element metabolism and associated with sensitivity to mercury, arsenic, cadmium, and lead in regional groups of indigenous Siberians.

Materials and methods

The study used the previously published results of the genotyping of 200 indigenous Siberian individuals, which had been obtained using the Illumina OmniExpress Bead Chips marker panel (Cardona et al., 2014). The northeast Siberian indigenous population sample (N = 58) included residents of the Bering Sea coast (Eskimo (Yupik) and Chukchi) and the Sea of Okhotsk coast (Koryaks). The Central Siberian sample included Evens from the Magadan Region, Evenks from the Krasnoyarsk Krai, and Yakuts from various regions of Yakutia (N = 70). The Southern Siberian sample included Shors from the Kemerovo Region, Southern Altaians from the Altai Republic and the Kemerovo Region, and Buryats from various regions of Buryatia (N = 72). Despite coexisting in the Severo-Evensk District of the Magadan Region since approximately the 16th century, the Koryaks and Evens are genetically distinct and classified into different regional groups (Cardona et al., 2014).

We examined indigenous Siberian populations, analyzing polymorphism in 78 loci presumably associated with interindividual differences in heavy metal (mercury, cadmium, and lead) concentrations in the blood of Canadian Inuit (Parajuli et al., 2018, 2021). The results are shown in Table S1[1]. These loci are located in genes that encode enzymes involved in key biological pathways, including antioxidant properties, folic acid metabolism, glutathione metabolism, hemoglobin, ion transporters, oxidative stress, selenoproteins, xenobiotic metabolism, metal-sensitive mechanisms (metallothioneins), lipid metabolism, and inflammatory processes.

Allele and genotype distribution, heterozygosity, and genetic differentiation of populations (Fst differences) were studied

[1] Supplementary Tables S1–S7 are available at:
https://vavilovj-icg.ru/download/pict-2026-30/appx26.xlsx

using the Arlequin 3.5 software package (Excoffier, Lischer, 2010). Deviations from the Hardy–Weinberg equilibrium in populations and differences in allele and genotype frequencies were assessed using Fisher's exact test. The degree of linkage disequilibrium between alleles at two different loci on a chromosome was determined using r^2 statistics. To assess the effect of selection on autosomal loci, the Population Branch Statistics (PBS) was used. This statistic is based on *F*st differences between three compared populations (Yi et al., 2010). High PBS values may point at strong selective pressure on the locus, causing allele frequencies to change faster than would be expected by drift alone.

Results and discussion

Genetic markers of mercury exposure

Mercury (Hg) is a toxic chemical element that threatens the health of people around the world. The high toxicity of inorganic and organic mercury compounds, especially methylmercury, is due to their interaction with thiol (-SH) groups of cysteine amino acid residues. This interaction causes disturbances in protein structure and enzyme active centers, as well as damage to cell membranes and organelles, resulting in increased oxidative stress (Arefieva et al., 2010; Basu et al., 2022).

Analysis of polymorphisms in loci associated with mercury metabolism in Canadian Inuit (according to data from R.P. Parajuli et al. (2018; 2021)) revealed statistically significant differences in allele distribution in 47 of 78 loci in pairwise comparisons of regional groups in Siberia (Table S1). Polymorphic variants of the seven loci with the highest PBS values, indicative of a probable selective effect on the locus (ranging from 0.266 to 0.043), had previously been demonstrated to be associated with blood mercury levels in Canadian Inuit carrying different genotypes (Table 1). These are loci rs2274976 (*MTHFR* gene), rs713041 (*GPX4*), rs1801131 (*MTHFR*), rs1128503 (*ABCB1*), rs2066853 (*AHR*), rs5748469 (*TXNRD2*), and rs1133238 (*SEPHS2*). The results of the analysis point to an increase in the frequency of alleles associated with elevated blood mercury levels from South to Northeast Siberia (Table S2). The highest frequencies of marker alleles (rs2274976-T, rs713041-C, rs1801131-G, rs1128503-A, rs2066853-G) were recorded among both Canadian Inuit and the indigenous populations of the coastal regions of Northeast Siberia (the Yupik, Chukchi, and Koryak peoples).

Additionally, the rs5748469-A variant of *TXNRD2* and the rs1133238-A variant of *SEPHS2* were only identified at maximum frequency in Northeast Siberia. Furthermore, the rs1133238-A allele primarily occurred in heterozygotes, who had the highest mercury concentrations in their blood. In terms of GG and GA+AA frequency, the Northeast Siberian sample significantly differs from other Siberian samples ($P < 10^{-4}$, Fisher's exact test).

The genetic loci identified as being associated with changes in blood mercury levels in Canadian Inuit are related to the following biological pathways: folate metabolism (*MTHFR* gene), selenoprotein metabolism (*GPX4* and *SEPHS2*), transport of various molecules (*ABCB1*), xenobiotic metabolism (*AHR*), and oxidative stress (*TXNRD2*). According to ClinVar Miner (https://clinvarminer.genetics.utah.edu/), the polymorphism of the *MTHFR* gene (rs2274976, Arg594Gln) is functionally neutral, although a slight increase in blood homocysteine concentration was observed in heterozygotes (Melo et al., 2006). The *MTHFR* gene encodes methylenetetrahydrofolate reductase, which converts homocysteine into the biologically active form of folic acid.

Similarly, the polymorphism of the rs1801131 locus (Glu429Ala) leads to a slight decrease in the activity of the encoded enzyme but does not affect homocysteine levels in blood plasma. It is also considered a neutral change (Weisberg et al., 2001). Conversely, the Ala222Val substitution at the rs1801133 locus is associated with a significant decrease in MTHFR activity (Frosst et al., 1995). This results in an increased risk of hyperhomocysteinemia due to methylenetetrahydrofolate reductase deficiency. It should be noted that the indigenous populations of Northeast and Central Siberia have a lower frequency of the rs1801133-A variant, associated with lower enzyme activity, than other regions of the world (Table S1). The *MTHFR* gene has a block organization (Trifonova et al., 2012). In particular, the polymorphism variants at loci rs2274976 and rs1801131, located at a distance of 3.5 thousand base pairs (kbp) from each other, are in a state of linkage.

Our analysis demonstrated that the indigenous population of Northeast Siberia exhibited significantly higher linkage disequilibrium between these loci (r^2 values ranging from 0.6 in the Chukchi to 1.0 in the Yupik) than more southerly populations (r^2 values ranging from 0.2 to 0.3). Therefore, the elevated frequencies of the rs2274976-T and rs1801131-G variants and pronounced linkage disequilibrium in the indigenous northeastern Siberians may be of functional significance, potentially linked to alterations in methylenetetrahydrofolate reductase activity.

The rs713041-C variant of the glutathione peroxidase *GPX4* gene, whose frequency increases to more than 80 % in northern Asia and America, appears to be important in protecting the body from the effects of peroxides, which are formed when exposed to various factors, including mercury. Studies have shown that GPX4 expression is higher for transgenes containing the rs713041-C variant when selenium levels are sufficient (Gautrey et al., 2011).

The *SEPHS2* gene encodes selenophosphate synthetase 2, responsible for selenium metabolism, which in turn plays an important role in protecting against the toxic effects of mercury and methylmercury (Jorge et al., 2024). Thus, the association of the rs1133238-A variant with elevated blood mercury concentrations observed in Canadian Inuit may be related to the protective significance of this allele. The *ABCB1* gene encodes P-glycoprotein, a transporter that removes various substances from cells, possibly including mercury (Sánchez Rodríguez et al., 2020). Therefore, the rs1128503-A variant, which is highly

Table 1. Distribution of polymorphism variants associated with sensitivity to heavy metal exposure among the indigenous populations of Siberia and Canadian Inuit

Metabolic pathway	Gene	Polymorphism variant	Frequency of polymorphism variant				*P* (*F*st)	*PBS*
			Canadian Inuit	Northeast Siberia (1)	Central Siberia (2)	South Siberia (3)		
				Hg				
Folate metabolism	*MTHFR*	rs2274976-C	0.488	0.595	0.871	0.944	(1–2) 0 (1–3) 0 (2–3) 0.039	0.266
Selenoprotein	*GPX4*	rs713041-T	0.113	0.152	0.321	0.410	(1–2) 0.0019 (1–3) 0 (2–3) 0.137	0.106
Folate metabolism	*MTHFR*	rs1801131-T	0.425	0.543	0.721	0.674	(1–2) 0.0054 (1–3) 0.04 (2–3) 0.44	0.045
Metabolism of xenobiotics	*AHR*	rs2066853-A	0.119	0.112	0.221	0.264	(1–2) 0.033 (1–3) 0.0028 (2–3) 0.41	0.05
Oxidative stress	*TXNRD2*	rs5748469-C	0.192	0.079	0.193	0.187	(1–2) 0.01 (1–3) 0.016 (2–3) 0.99	0.043
Selenoprotein	*SEPHS2*	rs1133238-A	0.143	0.32	0.13	0.125	(1–2) 0 (1–3) 0 (2–3) 0.99	0.102
Transporters	*ABCB1*	rs1128503-G	0.233	0.224	0.336	0.437	(1–2) 0.052 (1–3) 0.0006 (2–3) 0.084	0.051
				Pb				
Selenoprotein	*GPX4*	rs713041-T	0.113	0.152	0.321	0.410	(1–2) 0.0019 (1–3) 0 (2–3) 0.137	0.106
Glutathione metabolism	*GSTM3*	rs7483-C	0.32	0.198	0.421	0.41	(1–2) 0 (1–3) 0.0004 (2–3) 0.908	0.101
Transporters	*SLC22A1*	rs2282143-T	0.129	0.129	0.014	0.028	(1–2) 0.0003 (1–3) 0.0016 (2–3) 0.69	0.084
				Cd				
Folate metabolism	*MTHFR*	rs2274976-C	0.488	0.595	0.871	0.944	(1–2) 0 (1–3) 0 (2–3) 0.039	0.266
Metabolism of xenobiotics	*CYP1B1*	rs1056836-C	0.099	0.070	0.207	0.174	(1–2) 0.0021 (1–3) 0.016 (2–3) 0.547	0.054

Note. The data for the Canadian Inuit were obtained from R.P. Parajuli et al. (2018, 2021). *P* (*F*st) is the statistical significance of *F*st differences in allele frequency between Siberian populations in pairwise comparisons.

prevalent (approximately 80 %) in the Far North, may offer protection by reducing the toxic effects of mercury.

The *AHR* gene encodes the aromatic hydrocarbon receptor, which regulates the transcription of genes for enzymes that promote the metabolism of xenobiotics, including mercury, cadmium, arsenic, and nickel (Mohammadi-Bardbori et al., 2015; Alqahtani et al., 2024). The *TXNRD2* gene encodes mitochondrial thioredoxin reductase 2, an enzyme involved in the antioxidant defense system. Polymorphisms in these genes appear to influence susceptibility to mercury poisoning (Crespo-Lopez et al., 2023).

Genetic markers of lead exposure

The main mechanism of lead (Pb) toxicity involves the inhibition of enzymes by binding to thiol groups of cysteine residues (Gonzalez-Villalva et al., 2025). There is no defined safety threshold for lead in the blood because even low concentrations of this chemical element can cause serious problems, such as damage to cell membranes, impaired hemoglobin synthesis, and oxidative stress. Studies of Canadian Inuit populations have revealed that their susceptibility to lead exposure is linked to several genes involved in selenoprotein metabolism (*GPX4* gene), oxidative stress (*TXNRD2*), transport of various

molecules (*SLC22A1* and *SLCO1B1*), xenobiotic metabolism (*CYP2C19*), and other pathways (Parajuli et al., 2018, 2021).

Among indigenous Siberian populations, the greatest differentiation (*PBS* values ranging from 0.106 to 0.08) was found between the northeastern Siberian sample and other groups at loci rs713041 (*GPX4*), rs7483 (*GSTM3*), and rs2282143 (*SLC22A1*) (Tables S3 and S4). Thus, the rs713041 locus of the *GPX4* gene for glutathione peroxidase appears to protect the body not only from mercury exposure but also from lead toxicity (Table 1).

The *GSTM3* gene encodes glutathione S-transferase mu 3, which plays an important role in the detoxification of xenobiotics. The rs7483-T variant has been shown to significantly increase the activity of this enzyme by leading to the amino acid substitution Val224Ile, allowing for more efficient metabolism of drugs (Tetlow et al., 2004). This increased activity may also affect the metabolism of heavy metals, particularly lead. Therefore, the increased frequency of the rs7483-T variant of the gene, which encodes the more active form of GSTM3, among the indigenous northeastern Siberians, compared to other regional groups in Siberia and Canadian Inuit, is apparently associated with its protective role against the toxic effects of lead.

The *SLC22A1* gene encodes organic cation transporter 1 (OCT1), which plays an important role in transporting various substances, including drugs, across cell membranes. The rs2282143-T variant has been shown to reduce the ability of OCT1 to transport drugs into cells without significantly affecting *SLC22A1* gene expression (Chen et al., 2018). Therefore, the increased frequency of the rs2282143-T allele in indigenous populations of Northeast Siberia and Canadian Inuit may be associated with the protective role of this genetic polymorphism variant in response to heavy metal exposure.

The *ALAD* gene is the best known gene associated with increased sensitivity to lead poisoning (Stajnko et al., 2024). This gene encodes δ-aminolevulinic acid dehydratase (ALK dehydratase), which plays a role in the synthesis of porphobilinogen, the precursor of tetrapyrroles (heme and other compounds). The ALK dehydratase enzyme is highly sensitive to inhibition by heavy metals (Perini et al., 2024). Polymorphism at the rs1805313 locus has been shown to be closely associated with blood lead levels and to influence the expression of ALK dehydratase in blood cells (Warrington et al., 2015). Although no differences in blood lead and mercury concentrations were found among Canadian Inuit with different rs1805313 genotypes, comparative analysis revealed increased frequencies of the rs1805313-A variant in Canadian Inuit and indigenous peoples of Northeast Siberia (*PBS* value = 0.143) (Table S4). This may be due to the protective role of this *ALAD* gene polymorphism against lead.

Genetic markers of cadmium exposure

Cadmium (Cd) is one of the most common environmental pollutants, and it exerts several toxic effects. These include changes in gene expression, inhibition of DNA repair processes, interference with apoptosis and autophagy, oxidative stress, and interaction with trace elements necessary for the catalytic activity of enzymes (Qu, Zheng, 2024; Wang et al., 2025). Cadmium can compete with trace elements (e. g., Zn, Mn, Fe, Se, and Mg) for binding sites in the active centers of antioxidant enzymes, damaging them in the process (Đukić-Ćosić et al., 2020).

For people living in Arctic regions, the main sources of cadmium exposure are not only the consumption of traditional foods but also smoking (Becker, 2000; Adlard et al., 2024). For instance, Canadian Inuit individuals have been found to have blood cadmium concentrations eight times higher than normal (Parajuli et al., 2021). They also identified associations between cadmium levels and polymorphisms in genes associated with antioxidants (*TXNRD2* and *SELS*), oxidative stress (*NOS1*), transport (*ABCC1*), cytochrome P450 (*CYP2D6* and *CYP24A1*), and folate metabolism (*MTHFR*). We analyzed the distribution of polymorphism variants at several loci associated with cadmium metabolism in indigenous Siberians. However, the analysis revealed only one variant of the *MTHFR* gene, rs2274976-T. A statistically significant increase in the frequency of this variant is observed in both Northeast Siberia and among Canadian Inuit (Table 1 and Table S5). Among the latter group, the increased frequency of the rs2274976-T allele is accompanied by increased blood concentrations of cadmium and mercury, as noted above.

Statistically significant differences between the sample from Northeast Siberia and other Siberian groups were also found in the distribution of alleles of the rs1056836 locus of the *CYP1B1* gene. This locus also proved to be associated with differences in blood cadmium distribution among Canadian Inuit (Table 1 and Table S2). The rs1056836-G variant frequency is highest (over 90 %) among indigenous peoples in Northeast Siberia and the Canadian Inuit. This variant is associated with lower blood cadmium levels.

Also, our research identified genetic loci potentially associated with heavy metal exposure. These loci are characterized by relatively high *PBS* (ranging from 0.087 to 0.043). Statistically significant increases in allele frequencies were found in Canadian Inuit and indigenous peoples of Northeast Siberia for cadmium (rs688 of the *LDLR* gene) and mercury (rs628031 of the *SLC22A1* gene and rs732774 of the *ATP7B* gene), and only in northeastern Siberians for mercury (rs4895441 of the *HBS1L* gene) (Table S6). However, no differences in heavy metal content were found among Canadian Inuit carriers of different genotypes at these loci. Nevertheless, these loci are promising for further research in the field of environmental genetics of indigenous Siberian populations.

Genetic markers of arsenic exposure

Arsenic (As) is an element that can pose a threat to human health. Most seafood has elevated levels of arsenic. However, marine animals such as whales and seals mainly contain arsenic in the form of arsenobetaine, which appears to be nontoxic and harmless to humans (Becker, 2000). Nevertheless,

small amounts of other organic forms of arsenic are present in seafood and may be more toxic than inorganic arsenic (Bagryantseva, Khotimchenko, 2021). It has been shown that organic forms of arsenic can be converted into methylated and inorganic forms during metabolism. These forms can damage biological molecules and induce negative effects in metabolic processes (Bagryantseva, Khotimchenko, 2021).

Previous studies indicated that long-term consumption of water with high arsenic concentrations led to the selection of genetic variants that promote arsenic detoxification in indigenous Andean populations (Schläwicke Engström et al., 2009; Eichstaedt et al., 2015; Apata et al., 2017; De Loma et al., 2022). First, these changes affect the *AS3MT* gene for arsenic methyltransferase, which converts inorganic arsenic (iAs) into monomethylarsonic acid (MMA) and less toxic dimethylarsinic acid (DMA).

The activity level of AS3MT can be assessed from the ratio between DMA and MMA in urine: the more active the enzyme, the lower the percentage of MMA and the higher the percentage of DMA. Studies of female residents of San Antonio de los Cobres (SAC) in northern Argentina who consume water with a high concentration of arsenic (approximately 0.2 mg/L) have shown that other enzymes besides arsenic methyltransferase are involved in arsenic detoxification. These enzymes include those involved in one-carbon metabolism (*MTRR* and *CHDH* genes) and antioxidant protection (*PRDX2* and *GLRX* genes) (Schläwicke Engström et al., 2009).

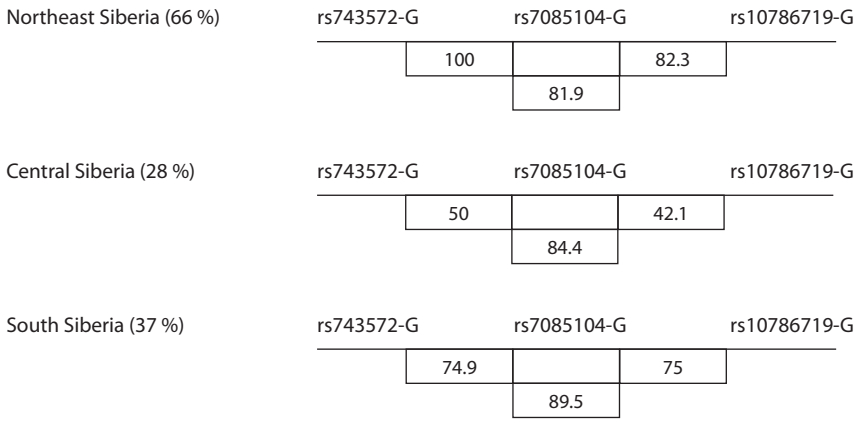
To date, a number of mutations in the *AS3MT* gene have been identified that affect the enzymatic efficiency of different forms of arsenic methyltransferase. We examined the population distribution of polymorphism frequencies at seven *AS3MT* gene loci and nine loci in the *CYP17A1*, *PRDX2*, *DNMT1a*, *INMT*, *MTRR*, *CNNM2*, *WBP1L*, and *N6AMT1* genes (Table S7). Statistically significant differences and maximum *PBS* values ranging from 0.217 to 0.083 were identified between indigenous populations of Northeast Siberia and other regions at the *AS3MT* gene loci rs7085104 and rs10786719 and the *CYP17A1* gene loci rs743572 and rs17115100.

It has been shown that a widespread protective haplotype is prevalent among Argentinians and Bolivians as a result of their adaptation to elevated levels of arsenic in their drinking water. Allelic variants of this haplotype are associated with the lowest MMA levels in urine (De Loma et al., 2022). One of the marker variants of this haplotype is rs7085104-G, which occurs at a frequency of 72.9 % in Argentinians (Engström et al., 2011). Our analysis of *AS3MT* and *CYP17A1* gene polymorphisms in the indigenous Siberian population also demonstrated close linkage between polymorphism variants at the rs743572, rs7085104, and rs10786719 loci, which are separated by 40.8 kbp (see the Figure).

The rs7085104-G variant had the highest frequency (65.8 %) and the GGG haplotype was most prevalent in northeastern Siberians (Table 2). Thus, the results show that the frequency of the polymorphic *AS3MT* and *CYP17A1* variants sensitive to the effects of arsenic increases in the northeastern direction in indigenous Siberian populations.

The main cause of this trend seems to be the traditional diets of indigenous peoples in the coastal regions of Northeast Siberia, such as the Eskimo, Chukchi, and Koryaks. Their ancestors have long consumed seafood (fish, shellfish, algae, and marine mammals), and the indigenous peoples of the northern seas still focus their diets on seafood. However, it is known that seafood contains high concentrations of heavy metals and arsenic (Becker, 2000; Gamov et al., 2022).

Arsenic in such products is primarily found in an organic form that is less toxic (mainly as arsenobetaine and arsenocholine), though it is also present in small amounts in an inorganic form (e. g., in animal liver and kidneys) (Bagryantseva, Khotimchenko, 2021). Constant seafood consumption over generations could have led to the selection of *AS3MT* variants that allow for more efficient elimination of iAs. Interestingly, our results showed that the optimal variants of genetic polymorphism in Northeast Siberia were the same as those found in indigenous Andean populations, who had consumed water with elevated concentrations of inorganic arsenic for a long time.



Linkage disequilibrium (r^2) between polymorphism variants at loci rs743572 of the *CYP17A1* gene and rs7085104 and rs10786719 of the *AS3MT* gene in the indigenous population of Siberia.

Table 2. Frequencies of genotypes and alleles of locus rs7085104 of the *AS3MT* gene in indigenous populations of Siberia and Argentina

Population (*N*)	Genotypes			Alleles		H_e	*P*
	GG	*GA*	*AA*	*G*	*A*		
Argentina (170)	0.547	0.365	0.088	0.729	0.271	0.396	0.33
Northeast Siberia (58)	0.456	0.404	0.14	0.658	0.342	0.45	0.39
Central Siberia (70)	0.071	0.414	0.515	0.279	0.721	0.405	1.0
South Siberia (72)	0.181	0.375	0.444	0.368	0.632	0.468	0.13
Siberia, in total (200)	0.221	0.397	0.382	0.42	0.58	0.489	0.009

Note. *N* – sample size. H_e – expected heterozygosity; *P* – statistical significance of deviation from Hardy–Weinberg equilibrium (significant at $P < 0.05$). Data for the Argentine sample are taken from (Engström et al., 2011).

Conclusion

This study analyzes genetic polymorphisms associated with sensitivity to toxic metals (mercury, lead, cadmium, and arsenic) in indigenous Siberian populations. Statistically significant differences were primarily observed between populations in Northeast Siberia (Eskimo (Yupiks), Chukchi, and Koryaks) and those in Central and Southern Siberia. The maximum *PBS* values indicating the probable effect of selection on genetic loci sensitive to mercury exposure were identified in seven gene loci: *MTHFR* (rs2274976 and rs1801131), *GPX4* (rs713041), *ABCB1* (rs1128503), *AHR* (rs2066853), *TXNRD2* (rs5748469), and *SEPHS2* (rs1133238). Loci rs713041 (*GPX4*), rs7483 (*GSTM3*), and rs2282143 (*SLC22A1*) can be considered genetic markers of lead exposure. Loci rs2274976-T (*MTHFR*) and rs1056836 (*CYP1B1*) may provide insight into cadmium distribution in the blood.

It should be noted that, in most cases, indigenous Siberian populations show an increase in genetic variants associated with elevated levels of mercury, lead, and cadmium in the northeastern direction. Furthermore, this trend aligns with the findings regarding Canadian Inuit populations (Parajuli et al., 2018, 2021). In our opinion, the most likely reason for this is the long-term adaptation of indigenous peoples along the Bering, Chukchi, and Okhotsk Sea coasts to their specific natural environment. This adaptation occurred over several millennia, and it is mainly due to their high seafood consumption. Seafood is rich in beneficial micronutrients, omega-3 polyunsaturated fatty acids, and selenium. However, it also contains toxic substances.

The widespread occurrence of genetic variants associated with elevated levels of heavy metals in the blood among Eskimo, Chukchi, and Koryaks is likely due to the formation of relatively harmless complexes of heavy metals with selenium and other detoxifying agents, such as glutathione and thioredoxin. These complexes reduce the bioavailability of metals and minimize their harmful effects. However, additional molecular-genetic and biochemical studies of indigenous populations are needed to determine the true causes. The available literature on the mechanisms of heavy metal detoxification and the risks of diseases caused by exposure in indigenous peoples of the Far North is insufficient (Ayotte et al., 2011; Ghisari et al., 2013; Parajuli et al., 2018, 2021).

The genetic consequences of human adaptation to environmental exposure to inorganic arsenic appear to be a better-studied issue (González-Martínez et al., 2024). Studies of indigenous populations in mountainous regions of Argentina, Chile, and Bolivia have shown that long-term consumption of drinking water with elevated arsenic levels has resulted in the selection of *AS3MT* gene polymorphism variants that enable more efficient arsenic methylation and elimination (Schlebusch et al., 2015).

Our study shows that the prevalence of *AS3MT* gene polymorphisms that protect against exposure to inorganic arsenic is significantly higher in indigenous northeastern Siberians. However, we have no information about long-term consumption of arsenic-contaminated drinking water by indigenous inhabitants of the Chukotka and Priokhotye coasts. Seafood is likely a source of arsenic in Northeast Siberia because it contains organic and inorganic arsenic. Further molecular-genetic and biochemical studies of indigenous Far North populations, as well as the microelement composition of groundwater and traditional food sources, are necessary in this regard.

References

Adlard B., Bonefeld-Jørgensen E.C., Dudarev A.A., Olafsdottir K., Abass K., Ayotte P., Caron-Beaudoin É., … Ratelle M., Rautio A., Timmerman A., Weihe P., Wennberg M. Levels and trends of metals in human populations living in the Arctic. *Int J Circumpolar Health.* 2024;83(1):2386140. doi 10.1080/22423982.2024.2386140

Alqahtani M.A., El-Ghiaty M.A., El-Mahrouk S.R., El-Kadi A.O.S. Differential modulatory effects of methylmercury (MeHg) on *Ahr*-regulated genes in extrahepatic tissues of C57BL/6 mice. *Biol Trace Elem Res.* 2024;202(11):5071-5080. doi 10.1007/s12011-023-04050-y

Apata M., Arriaza B., Llop E., Moraga M. Human adaptation to arsenic in Andean populations of the Atacama Desert. *Am J Phys Anthropol.* 2017;163(1):192-199. doi 10.1002/ajpa.23193

Arefieva A.S., Barigina V.V., Zatsepina O.V. The present-day ideas about impact of mercuric compounds at cell and system levels (re-

view). *Ekologiya Cheloveka = Human Ecology*. 2010;8:35-41 (in Russian)

Ayotte P., Carrier A., Ouellet N., Boiteau V., Abdous B., Sidi E.A., Château-Degat M.L., Dewailly É. Relation between methylmercury exposure and plasma paraoxonase activity in inuit adults from Nunavik. *Environ Health Perspect.* 2011;119(8):1077-1083. doi 10.1289/ehp.1003296

Bagryantseva O.V., Khotimchenko S.A. Risks associated with the consumption of inorganic and organic arsenic. *Voprosy Pitaniya = Problems of Nutrition.* 2021;90(6):6-17. doi 10.33029/0042-8833-2021-90-6-6-17 (in Russian)

Basu N., Abass K., Dietzd R., Krümmel E., Rautio A., Weihe P. The impact of mercury contamination on human health in the Arctic: a state of the science review. *Sci Total Environ.* 2022;831:154793. doi 10.1016/j.scitotenv.2022.154793

Becker P.R. Concentration of chlorinated hydrocarbons and heavy metals in Alaska Arctic marine mammals. *Mar Pollut Bull.* 2000; 40(10):819-829. doi 10.1016/S0025-326X(00)00076-X

Cardona A., Pagani L., Antao T., Lawson D.J., Eichstaedt C.A., Yngvadottir B., Shwe M.T., … Willerslev E., Tyler-Smith C., Malyarchuk B.A., Derenko M.V., Kivisild T. Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PloS One*. 2014;9(5):e98076. doi 10.1371/journal.pone.0098076

Chen W., Zhang X., Zhang W., Peng C., Zhu W., Chen X. Polymorphisms of *SLCO1B1* rs4149056 and *SLC22A1* rs2282143 are associated with responsiveness to acitretin in psoriasis patients. *Sci Rep.* 2018;8:13182. doi 10.1038/s41598-018-31352-2

Crespo-Lopez M.E., Barthelemy J.L., Lopes-Araújo A., Santos-Sacramento L., Leal-Nazaré C.G., Soares-Silva I., Macchi B.M., do Nascimento J.L.M., Arrifano G.P., Augusto-Oliveira M. Revisiting genetic influence on mercury exposure and intoxication in humans: a scoping review. *Toxics.* 2023;11(12):967. doi 10.3390/toxics11120967

De Loma J., Vicente M., Tirado N., Ascui F., Vahter M., Gardon J., Schlebusch C.M., Broberg K. Human adaptation to arsenic in Bolivians living in the Andes. *Chemosphere.* 2022;301:134764. doi 10.1016/j.chemosphere.2022.134764

Đukić-Ćosić D., Baralić K., Javorac D., Buha Djordjevic A., Bulat Z. An overview of molecular mechanisms in cadmium toxicity. *Curr Opin Toxicol.* 2020;19:56-62. doi 10.1016/j.cotox.2019.12.002

Eichstaedt C., Antao T., Cardona A., Pagani L., Kivisild T., Mormina M. Positive selection of *AS3MT* to arsenic water in Andean populations. *Mutat Res*. 2015;780:97-102. doi 10.1016/j.mrfmmm.2015.07.007

Engström K., Vahter M., Mlakar S.J., Concha G., Nermell B., Raqib R., Cardozo A., Broberg K. Polymorphisms in arsenic(+III oxidation state) methyltransferase (AS3MT) predict gene expression of AS3MT as well as arsenic metabolism. *Environ Health Perspect*. 2011;119: 182-188. doi 10.1289/ehp.1002471

Excoffier L., Lischer H.E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10(3):564-567. doi 10.1111/j.1755-0998.2010.02847.x

Fernandez-Llamazares A., Garteizgogeascoa M., Basu N., Brondizio E.S., Cabeza M., Martínez-Alier J., McElwee P., Reyes-García V. A state-of-the-art review of indigenous peoples and environmental pollution. *Integr Environ Assess Manag.* 2020;16(3):324-341. doi 10.1002/ieam.4239

Frosst P., Blom H.J., Milos R., Goyette P., Sheppard C.A., Matthews R.G., Boers G.J., den Heijer M., Kluijtmans L.A., van den Heuvel L.P. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet.* 1995;10(1):111-113. doi 10.1038/ng0595-111

Gamov M.K., Ivanova A.E., Mironova E.K., Tsygankov V.Yu. Heavy metals and arsenic in commercial fish of the Sea of Japan, Sea of Okhotsk, and Bering Sea: current status (literature review). *Morskoy Biologicheskii Zhurnal = Marine Biological Journal.* 2022;7(4): 14-30. doi 10.21072/mbj.2022.07.4.02 (in Russian)

Gautrey H., Nicol F., Sneddon A.A., Hall J., Hesketh J. A T/C polymorphism in the *GPX4* 3′UTR affects the selenoprotein expression pattern and cell viability in transfected Caco-2 cells. *Biochim Biophys Acta.* 2011;1810(6):584-591. doi 10.1016/j.bbagen.2011.03.016

Ghisari M., Long M., Bonefeld-Jørgensen E.C. Genetic polymorphisms in *CYP1A1*, *CYP1B1* and *COMT* genes in Greenlandic Inuit and Europeans. *Int J Circumpolar Health.* 2013;72:21113. doi 10.3402/ijch.v72i0.21113

González-Martínez F., Johnson-Restrepo B., Quiñones L.A. Arsenic inorganic exposure, metabolism, genetic biomarkers and its impact on human health: a mini-review. *Toxicol Lett.* 2024;398:105-117. doi 10.1016/j.toxlet.2024.06.008

Gonzalez-Villalva A., Marcela R.L., Nelly L.V., Patricia B.N., Guadalupe M.R., Brenda C.T., Maria Eugenia C.V., Martha U.C., Isabel G.P., Fortoul T.I. Lead systemic toxicity: a persistent problem for health. *Toxicology.* 2025;515:154163. doi 10.1016/j.tox.2025.154163

Jorge A.O.S., Chamorro F., Carpena M., Echave J., Pereira A.G., Oliveira M.B.P.P., Prieto M.A. Protection of selenium against methylmercury in the human body: a comprehensive review of biomolecular interactions. *Biol Life Sci Forum*. 2024;35(1):8. doi 10.3390/blsf2024035008

Melo S.S., Persuhn D.C., Meirelles M.S., Jordao A.A., Vannucchi H. G1793A polymorphisms in the methylenetetrahydrofolate gene: effect of folic acid on homocysteine levels. *Mol Nutr Food Res.* 2006; 50(8):769-774. doi 10.1002/mnfr.200600020

Mohammadi-Bardbori A., Vikstrom Bergander L., Rannug U., Rannug A. NADPH oxidase-dependent mechanism explains how arsenic and other oxidants can activate aryl hydrocarbon receptor signaling. *Chem Res Toxicol.* 2015;28(12):2278-2286. doi 10.1021/acs.chemrestox.5b00415

Parajuli R.P., Goodrich J.M., Chan L.H.M., Ayotte P., Lemire M., Hegele R.A., Basu N. Genetic polymorphisms are associated with exposure biomarkers for metals and persistent organic pollutants among Inuit from the inuvialuit settlement region, Canada. *Sci Total Environ.* 2018;634:569-578.

Parajuli R.P., Goodrich J.M., Chan H.M., Lemire M., Ayotte P., Hegele R.A., Basu N. Variation in biomarker levels of metals, persistent organic pollutants, and omega-3 fatty acids in association with genetic polymorphisms among Inuit in Nunavik, Canada. *Environ Res.* 2021;200:111393. doi 10.1016/j.envres.2021.111393

Perini J.A., Cardoso J.V., Knesse A.O., Pessoa-Silva F.O., Vasconcellos A.C.S., Machado D.E., Basta P.C. Single-nucleotide polymorphisms associated with mercury levels and neurological symptoms: an overview. *Toxics.* 2024;12(3):226. doi 10.3390/toxics12030226

Qu F., Zheng W. Cadmium exposure: mechanisms and pathways of toxicity and implications for human health. *Toxics.* 2024;12:388. doi 10.3390/toxics12060388

Sánchez Rodríguez L.H., Medina Pérez O.M., Rondón González F., Rincón Cruz G., Rocha Muñoz L., Flórez-Vargas O. Genetic polymorphisms in multispecific transporters mitigate mercury nephrotoxicity in an artisanal and small-scale gold mining community in Colombia. *Toxicol Sci.* 2020;178(2):338-346. doi 10.1093/toxsci/kfaa142

Schläwicke Engström K., Nermell B., Concha G., Strömberg U., Vahter M., Broberg K. Arsenic metabolism is influenced by polymorphisms in genes involved in one-carbon metabolism and reduction reactions. *Mutat Res.* 2009;667(1-2):4-14. doi 10.1016/j.mrfmmm.2008.07.003

Schlebusch C.M., Gattepaille L.M., Engström K., Vahter M., Jakobsson M., Broberg K. Human adaptation to arsenic-rich environments. *Mol Biol Evol.* 2015;32(6):1544-1555. doi 10.1093/molbev/msv046

Stajnko A., Palir N., Snoj Tratnik J., Mazej D., Sešek Briški A., Runkel A.A., Horvat M., Falnoga I. Genetic susceptibility to low-level lead exposure in men: insights from ALAD polymorphisms. *Int J Hyg Environ Health*. 2024;256:114315. doi 10.1016/j.ijheh.2023.114315

Tetlow N., Robinson A., Mantle T., Board P. Polymorphism of human mu class glutathione transferases. *Pharmacogenetics*. 2004;14(6):359-368. doi 10.1097/00008571-200406000-00005

Trifonova E.A., Eremina E.P., Urnov F.D., Stepanov V.A. The genetic diversity and structure of linkage disequilibrium of the *MTHFR* gene in populations of Northern Eurasia. *Acta Naturae.* 2012;4(1):53-69

Wang H., Gan X., Tang Y. Mechanisms of heavy metal cadmium (Cd)-induced malignancy. *Biol Trace Elem Res.* 2025;203(2):608-623. doi 10.1007/s12011-024-04189-2

Warrington N.M., Zhu G., Dy V., Heath A.C., Madden P.A., Hemani G., Kemp J.P., … Montgomery G.W., Martin N.G., Davey Smith G., Evans D.M., Whitfield J.B. Genome-wide association study of blood lead shows multiple associations near ALAD. *Hum Mol Genet.* 2015;24(13):3871-3879. doi 10.1093/hmg/ddv112

Weisberg I.S., Jacques P.F., Selhub J., Bostom A.G., Chen Z., Curtis Ellison R., Eckfeldt J.H., Rozen R. The 1298A→C polymorphism in methylenetetrahydrofolate reductase (MTHFR): in vitro expression and association with homocysteine. *Atherosclerosis*. 2001;156(2):409-415. doi 10.1016/s0021-9150(00)00671-7

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

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

Received August 16, 2025. Revised November 27, 2026. Accepted January 27, 2026.

HUMAN GENETICS
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):470-481

doi 10.18699/vjgb-26-51

The association study of genetic variants with developing musical aptitude in humans

A.V. Kazantseva [1, 2], A.V. Toropova [3], E.K. Khusnutdinova [1], S.B. Malykh [4, 5]

[1] Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
[2] Ufa State Petroleum Technological University, Russia
[3] Moscow Pedagogical State University, Moscow, Russia
[4] Psychological Institute of the Russian Academy of Education, Moscow, Russia
[5] M.V. Lomonosov Moscow State University, Department of Psychology, Moscow, Russia
Kazantsa@mail.ru

Abstract. The development of musical abilities, including absolute pitch, musical memory, rhythm sense, and musicality, at a high degree is determined by a hereditary component (up to 68 %). The studies implementing a genome-wide linkage and association approach to musical aptitude have revealed more than 100 genetic loci. This spectrum is comprised of the genes encoding for transcription factors and those responsible for neurogenesis and synaptic plasticity, genes fixed as a result of positive selection of musicality, and those related to inner ear formation. Since no studies linking musical aptitude with genes have been previously conducted in Russia, the present study aimed at replicating the association of 17 previously identified genetic variants with developing musical abilities in Russians. Genotyping of SNPs in the *GATA2*, *PCDH7*, *UNC5C*, *ASAP1*, *SBSPON*, *DCBLD2*, *KALRN*, *VLDLR*, *OTOF*, *GRIN2B*, *FoxP1*, *FoxP2*, *BDNF*, *EGR1*, and *SNCA* genes was performed using competitive allele-specific PCR in a sample of students who underwent rigorous contest selection at admission to the conservatory and in the corresponding control group. A series of logistic regression analyses were used both to evaluate the main effect of SNP and to identify the best prognostic model based on various loci. The mathematical model obtained by including only statistically significant SNPs consisted of *GATA2* rs9854612, *SNCA* rs356168, rs3910105, *ASAP1* rs3057, and *VLDLR* rs1454626 ($p = 0.0018$, pseudo $r^2 = 0.188$, AUC = 0.791). The addition of all examined SNPs as predictors enabled the construction of a statistically significant model with a higher predictive ability ($p = 0.012$, pseudo $r^2 = 0.380$, AUC = 0.889). The results revealed indicate a potential cumulative gene effect, confirming the involvement of dopaminergic and GABAergic neurotransmission, the reelin pathway and the role of alpha-synuclein in musicality formation.

Key words: music; logistic regression; mathematical model; cognitive abilities; α-synuclein; dopamine; thickness of Heschl's gyrus

For citation: Kazantseva A.V., Toropova A.V., Khusnutdinova E.K., Malykh S.B. The association study of genetic variants with developing musical aptitude in humans. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3): 470-481. doi 10.18699/vjgb-26-51

Исследование ассоциации генетических вариантов с развитием музыкальных способностей человека

А.В. Казанцева [1, 2], А.В. Торопова [3], Э.К. Хуснутдинова [1], С.Б. Малых [4, 5]

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

Аннотация. Формирование музыкальных способностей, включающих абсолютный слух, музыкальную память, чувство ритма, музыкальность, в значительной степени определяется наследственной составляющей (до 68 %). Проведенные к настоящему времени работы с использованием полногеномного анализа сцепления и ассоциаций с музыкальной одаренностью позволили выявить более 100 генетических локусов. В этот спектр входят гены транскрипционных факторов, регуляции нейрогенеза и синаптической пластичности; гены, закрепленные в ходе позитивной селекции музыкальности, а также связанные с особенностями формирования внутреннего уха. Поскольку ранее в Российской Федерации исследований по изучению связи музыкального таланта с генетической компонентой не проводилось, настоящая работа направлена на репликацию ассоциации ранее идентифицированных 17 однонуклеотидных вариантов (SNP) с формированием музыкальных способностей у

русских. Генотипирование полиморфных локусов в генах *GATA2*, *PCDH7*, *UNC5C*, *ASAP1*, *SBSPON*, *DCBLD2*, *KALRN*, *VLDLR*, *OTOF*, *GRIN2B*, *FoxP1*, *FoxP2*, *BDNF*, *EGR1*, *SNCA* проводилось с помощью конкурентной аллель-специфичной ПЦР в выборке студентов, прошедших строгий конкурсный отбор при поступлении в консерватории, и в соответствующей контрольной группе. Метод логистической регрессии применялся как для оценки основного эффекта отдельных полиморфных вариантов, так и для выявления наилучшей прогностической модели, содержащей различные генетические локусы. Математическая модель, полученная в результате включения только ассоциированных SNP, состояла из генетических локусов *GATA2* rs9854612, *SNCA* rs356168 и rs3910105, *ASAP1* rs3057 и *VLDLR* rs1454626 (p = 0.0018, псевдо-r^2 = 0.188, AUC = 0.791). Добавление всех изученных генетических локусов в качестве предикторов в регрессионный анализ позволило создать статистически значимую модель, обладающую более высокой прогностической способностью (p = 0.012, псевдо-r^2 = 0.380, AUC = 0.889). Полученные результаты указывают на потенциальный кумулятивный эффект белковых продуктов изученных генов, подтверждая вовлеченность дофаминергической и ГАМКергической нейротрансмиссии, рилинового пути и роль альфа-синуклеина в формировании музыкальности.

Ключевые слова: музыка; логистическая регрессия; математическая модель; когнитивные способности; α-синуклеин; дофамин; толщина извилины Хешля

Introduction

Music is an integral part of cognitive and social interaction between humans, which was used as an important component in the information transfer even before the speech formation. It is assumed that music existed more than 35,000 years ago, evidence of which is found in archaeological excavations of caves in Germany, Austria, and France (Conard et al., 2009). Despite the fact that currently this function mainly belongs to verbal communication, musical abilities are closely related to cognitive and mental activity. Musical abilities cover a significant range of various individual psychological characteristics, including absolute hearing, musical memory, rhythm sense, and musicality (Aikina, 2017), and represent quantitative parameters, the distribution of which in a population corresponds to the law of normal distribution.

The ability to perceive and reproduce music is assumed to be caused by a hereditary component, which initiated research seeking to identify the genes responsible for the development of musical abilities (musical talent). The results from twin studies as the primary stage of examining the role of genes and the environment in the formation of behavioral and cognitive abilities (Kazantseva, 2008) indicate a significant contribution of the hereditary component: genes determine up to 68 % of variance in musical aptitude (Oikkonen et al., 2015).

Some of the pioneer studies in the field of genetics of high-level musical aptitude were based on the examination of large-sized pedigrees of Finnish origin, which were characterized by accumulated skills of musical aptitude in families (Pulli et al., 2008). Using the genome-wide linkage approach (GWLS), the abovementioned and other research teams succeeded in identifying chromosomal regions 3q21.3 (including the *GATA2* gene), 4p15.1 (including the *PCDH7* gene), 4q22 (including the *UNC5C* gene), 8q24.21 (including the *ASAP1* gene), and 8q21.11 (including the *SBSPON* gene) as linked with specificity of music perception (Pulli et al., 2008; Oikkonen et al., 2015) and absolute pitch (Theusch et al., 2009).

Published genome-wide association studies (GWAS) identified genetic loci, which determine individual variance in thickness of Heschl's gyrus, i. e., the central region of the auditory cortex, responsible for speech and music perception, including the *DCBLD2* rs72932726 and the *KALRN* rs333332 (Cai et al., 2014). On the other hand, GWAS enabled the detection of SNPs related to positive evolutionary selection of musical abilities (Liu et al., 2016). Namely, the study of Finnish respondents with high-degree musicality resulted in the detection of several genomic regions, including the *VLDLR*, *FOXP1*, *OTOF*, and *GRIN2B* genes, which formed the haplotypes being statistically more frequent among musically talented persons. The involvement of the abovementioned genes is unsurprising, since specific changes in their nucleotide sequences have been previously linked with speech impairment in humans (Rappold et al., 2023), vocalization and changes in gene expression in songbirds (Adam et al., 2016; So et al., 2019; Heim et al., 2023).

A development of any quantitative trait is known to be caused by the effects of multiple genetic loci, which requires research devoted to musical abilities to be conducted using data on a large number of genes and their structural variants. Therefore, together with GWAS and GWLS data, the results of functional studies, which were carried out with the use of model animals and enabled the estimation of the expression levels of annotated genes, are of great interest. In particular, a differential expression of the *FoxP2*, *BDNF*, and *EGR1* genes was observed in songbirds during learning and vocalization and in model mice via listening to music (Li et al., 2010; Drnevich et al., 2012; Shi et al., 2013). Moreover, changes in the expression patterns of the alpha-synuclein gene (*SNCA*), the mutations of which are linked to predisposition to Parkinson's disease and other synucleinopathies (Järvelä et al., 2018), were determined in humans as a result of listening to classical music (Kanduri et al., 2015).

Within the present study we have selected SNPs in the genes related to the development of musical abilities and characterized by changes in gene expression observed via listening to music. Gene polymorphisms were obtained either directly from the GWAS and GWLS or – in case information on certain

SNPs being linked with musical abilities was absent – from the studies reporting the impact of genetic loci on the development of related behavioral traits (Crawford et al., 2008; Hu et al., 2011; Koks et al., 2021). The final information on the genes and their genetic variants selected as the objects of the present study is provided in Table 1.

The present study represents the extension of our previous research, which estimated a probability to develop high-level musical aptitude based on 10 genetic variants (Kazantseva et al., 2023). Accordingly, the present research is aimed at replicating the involvement of the extended panel of genetic variants in developing musical talent in ethnic Russians from the Russian Federation.

Materials and methods

The experimental part of the study included a selection of individuals characterized by musical talent who were students at conservatories (Moscow) and passed a severe competitive contest upon admission, during which they demonstrated outstanding abilities in entrance exams on solfeggio, harmony, polyphony, reading scores, etc., depending on the chosen specialty.

The sample consisted of ethnical Russians aged 18–22 years (N = 100, 66 % women; mean age 19.36 ± 1.44 years). The control group was comprised of Russian students at the Universities of Moscow and Ufa corresponding by age and sex to the sample of musicians (N = 200, 67 % women; mean age 19.84 ± 1.85 years). All the examined subjects were not observed in neuropsychiatric institutions and denied a family history of mental disorders. Voluntary consent to participate in the study was obtained from all the respondents. The study was approved by the Bioethical Committee at the Institute of Biochemistry and Genetics of UFRC of RAS.

A collection of biological material (5 ml saliva) was carried out in the tubes containing a preserving agent according to the manufacturer's recommendations (Oragene DNA, DNA Genotek, Canada). Subsequent DNA isolation was performed using specific kits (PrepIT, DNA Genotek, Canada), followed by the assessment of DNA quality via spectrophotometer. Genotyping of examined SNPs was conducted via endpoint real-time PCR on the CFX96 DNA Analyzer (BioRad, USA) using KASP chemistry (LGC Genomics, UK).

A correspondence of observed genotype frequencies distribution to the theoretically expected one based on the Hardy–Weinberg law was carried out in the control group and revealed a deviation of the *PCDH7* rs13109270; therefore, this SNP was excluded from subsequent steps of the analysis. Statistical analysis was based on testing of various regression models of analyzed genetic loci, including additive, dominant, recessive, overdominant, and codominant (PLINK v.1.09).

The use of the additive model demonstrates whether the effect of the alternative allele is accumulative; the dominant model shows whether the effect of the tested allele is observed in the presence of at least one copy (i. e. in the heterozygous state and in the homozygote of the effect allele); the recessive model shows whether the effect of the examined allele is prominent with the presence of both copies. The overdominant model describes a state when the highest effect is observed when both alleles are present (i. e. in the case of a heterozygous genotype) compared with both reference or alternative alleles (i. e. in the case of any homozygous genotype). The codominant model assumes that any genotype demonstrates a significant effect on developing the examined trait in a non-additive manner, i. e., independently of other genotypes (Kutikhin et al., 2017).

Multiple logistic regression analysis was used to test for different models (including all the SNPs) and to design the final mathematical model, which included statistically significant predictors and explained the highest proportion of variance in developing high musical aptitude (R v.4.4.2). To estimate the models' quality, the Akaike information criterion (AIC) was used: the lower its value, the higher the model's quality. The best model was visualized as the ROC (Receiver Operating Characteristic) curve reporting a quantitative measure of the AUC (area under curve) parameter.

Results

At the first stage of association analysis between genetic variants and high-level musical aptitude, we confirmed the involvement of the *ASAP1* rs3057 (p = 0.032) in developing musical abilities at a statistically significant level (Table 2). In particular, enhanced chances of demonstrating musical abilities were characteristic of heterozygous carriers of the *ASAP1* rs3057 C/T genotype compared with homozygotes (T/T and C/C genotypes) (OR = 1.94, 95 % CI 1.05–3.56), which stands in evidence of a higher effect of the heterozygous genotype of this SNP exactly on manifesting musical talent. Other genetic variants failed to demonstrate an association with musicality at a single-locus level in any analyzed model: the additive, dominant, recessive, and deviation from dominance models.

Since the development of high-level musical aptitude, like any other complex trait, represents a result of the interaction of proteins encoded by various genes, the subsequent stage was based on a series of multiple regression analyses, which consisted of 16 SNPs as predictors at the initial level (model 1 (Table 3) $\chi^2 = 46.24$, $p = 0.012$). After the exclusion of statistically insignificant SNPs, we obtained a model that included the *GATA2* rs9854612, *SNCA* rs356168 and rs3910105, *ASAP1* rs3057, and *VLDLR* rs1454626 gene polymorphisms (model 2 (Table 3) $\chi^2 = 24.61$, $p = 0.0018$).

According to the final model, the highest chances of manifesting musical talent will be realized in the case of simultaneous presence of the following genotypes: the *GATA2* rs9854612 A/G (p = 0.026), *SNCA* rs356168 T/T (p = 0.007) and rs3910105 T/T (p = 0.043), *ASAP1* rs3057 C/T (p = 0.027), and *VLDLR* rs1454626 A/A (p = 0.027).

Therefore, the highest genetic effect linked to developing musical abilities will be observed in the case of the presence of a heterozygous variant compared with the carriers of any

Table 1. Polymorphic variants associated with developing musical abilities

No.	Region	Gene (protein)	SNP	Related trait	Protein function	Reference
Genome-wide linkage analyses (GWLA)						
1	3q21.3	*GATA2* (GATA-binding protein 2)	rs9854612	Music perception (the ability to distinguish pitch, duration, and sound patterns)	Involved in the development of inner ear cells	Oikkonen et al., 2015
2	4p15.1	*PCDH7* (protocadherin 7)	rs13146789			
3			rs13109270			
4	8q24.21	*ASAP1* (ankyrin repeat and PH domain 1)	rs3057	Absolute pitch	Regulates membrane transport and cytoskeleton remodeling	Theusch et al., 2009
5	8q21.11	*SBSPON* (somatomedin B and thrombospondin type 1 domain containing)	rs1007750		Extracellular matrix component	
6	4q22.3	*UNC5C* (unc-5 netrin receptor)	rs9307160	Musical aptitude	Participates in axons elongation and cell migration; interacts with ROBO family receptors associated with dyslexia	Pulli et al., 2008
Genome-wide association studies (GWAS)						
7	3q12.1	*DCBLD2* (discoidin, CUB and LCCL domain containing 2)	rs72932726	Thickness of Heschl's gyrus, i. e., the central region of the auditory cortex, responsible for sounds perception	Regulates cell growth; is associated with developmental delay	Cai et al., 2014
8	3q21.1-q21.2	*KALRN* (kalirin RhoGEF kinase)	rs333332		Interacts with the HTT protein, the expansion of which is the cause of Parkinson's disease; mutations are linked with changes in the neurons' morphology	
9	9p24.2	*VLDLR* (very low-density lipoprotein receptor)	rs1454626[a]	Positive evolutionary selection of musical aptitude	A target for the FoxP2 involved in the formation of singing and speech	Liu et al., 2016; ASD Working group et al., 2017
10	3p13	*FOXP1* (transcription factor forkhead box P1)	rs6803008		Associated with speech impairments, ASD; with music and singing in songbirds	
11	2p23	*OTOF* (otoferlin)	rs4416176[b]		Involved in exocytosis of inner ear cells; mutations in the gene are associated with hearing loss	
12	12p13.1	*GRIN2B* (glutamate ionotropic receptor NMDA type subunit 2B)	rs3764030		Responsible for regulation of synaptic plasticity and learning; linked with music perception and singing in songbirds	
Functional studies (estimation of expression level)						
13	7q31.1	*FoxP2* (transcription factor forkhead box P)	rs2396753	Decreased expression disrupts learning in songbirds; mutations in the gene cause autosomal-dominant speech disorder	Transcription factor, which regulates the expression of 300–400 genes; is highly expressed in the striatum and involved in the development of brain regions related to speech development during embryogenesis	Shi et al., 2013

Table 1 (end)

No.	Region	Gene (protein)	SNP	Related trait	Protein function	Reference
14	11p14.1	*BDNF* (brain-derived neurotrophic factor)	rs6265	Increased expression of the *BDNF* (Met/Met) gene in the prefrontal cortex, amygdala, and hippocampus in transgenic mice during listening to music	Regulates neurogenesis and synaptic plasticity	Li et al., 2010
15	4q22.1	*SNCA* (alpha-synuclein)	rs356168	Increased gene expression in humans via listening to music	Regulated by the GATA2	Kanduri et al., 2015
16			rs3910105[c]			
17	5q31.2	*EGR1* (early growth response 1)	rs7729723	Increased gene expression in songbirds during song listening and vocalization	Transcriptional regulator of FoxP2	Drnevich et al., 2012

Note. We have selected polymorphisms in genes with a previously reported link with: [a] lipoprotein levels (Crawford et al., 2008); [b] social abilities in patients with autism spectrum disorder (ASD) (Hu et al., 2011); [c] differential gene expression caused by various allelic variants (Koks et al., 2021).

homozygote (*ASAP1* rs3057, overdominant model) or the carriers of both copies of the alternative allele (*GATA2* rs9854612, codominant model).

The results of logistic regression also point to the potential positive accumulated effect of the T allele (in both SNPs of the *SNCA* gene: rs356168 and rs3910105, recessive model, the effect will be observed only in the presence of both alleles of this gene) and the C allele (*VLDLR* rs1454626, dominant model, it will be evident in the presence of at least one copy of the effect allele) on musical talent.

Construction of ROC curves and the calculated area under curve (AUC) testify that model 1, which consists of all 16 SNPs, has the highest prognostic ability (AUC = 0.889) (see the Figure), although not all included predictors demonstrate a statistically significant effect on developing musical abilities. Despite the lower predictive ability of the final model (model 2) (AUC = 0.791) (see the Figure), it is characterized by better quality parameters (AIC = 131.31 for model 1, AIC = 123.94 for model 2).

Discussion

Within the framework of the present study, we carried out a replication association analysis of gene polymorphisms, which had previously demonstrated their relation to musical aptitude and music perception. The inclusion of all examined genetic variants as predictors in the regression analysis enabled the construction of the statistically significant model with high predictability (area under the ROC curve = 0.889). However, the final model consisting of the data from five gene polymorphisms, i. e. the *GATA2* rs9854612, *SNCA* rs356168 and rs3910105, *ASAP1* rs3057, and *VLDLR* rs1454626, demonstrated more statistical significance and better quality (area under the ROC curve = 0.791).

Genetic loci identified in GWLA

In the present study we have analyzed genetic loci rs9854612, rs13146789, and rs13109270 located in proximity to the *GATA2* and *PCDH7* genes, which demonstrated a linkage with the specificity of music perception at a genome-wide level (Oikkonen et al., 2015). It is known that GATA binding protein 2, encoded by the same gene, is involved in the development of inner ear hair cells, which are important for tonotopic mapping (Oikkonen et al., 2015). Moreover, GATA2 participates in dopaminergic neurotransmission and regulates the expression of the alpha-synuclein gene (*SNCA*), the enhanced transcription of which is observed together with dopamine release during listening to music (Järvelä, 2018), pointing to the involvement of the brain reward system in this process.

The present study demonstrated a positive effect of the simultaneous presence of single copies of reference and alternative alleles in the *GATA2* rs9854612 gene polymorphism on increased chances of developing musical talent. In turn, protocadherins regulate neuronal migration, differentiation, and synaptogenesis and participate in the formation of cochlear-nuclear and amygdaloid complexes (Oikkonen et al., 2015), promoting their impact on musical abilities. The opposite effect was mentioned during a segregation analysis of a deletion of a cluster of protocadherin genes in the 5q31.1 genomic region in large pedigrees, which resulted in reduced musical creativity (Ukkola-Vuoti et al., 2013).

Interesting data were obtained in a recent study indicating a link between plasma hypermethylation of the gene cluster, including *PCDH7*, and lead concentration in the blood of pregnant women whose previous pregnancy resulted in the birth of an ASD child (Aung et al., 2022). Based on gene ontology analysis provided by the authors, the level of this toxic metal is related to such biological processes as neuronal development and changes in the immune system regulation, which may cause abnormal development of the fetal nervous system. Since no findings on a functional role of allelic variants of the examined SNPs in the *GATA2* and *PCDH7* genes were published to date, it cannot be unambiguously concluded

Table 2. Allele and genotype frequencies of examined polymorphisms in the group of musicians and in the control group and the results of association analysis

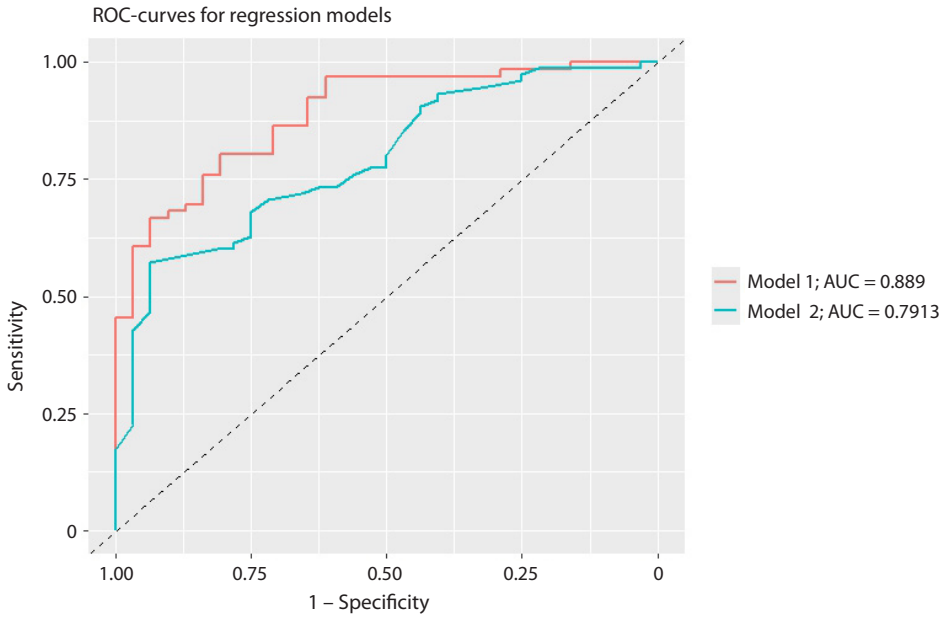
Gene	SNP	Frequencies					*p*	P_{HWE}
		genotype[a]			allele[b]			
OTOF	rs4416176	GG	AG	AA	G	A	0.843	0.977
		0.013	0.253	0.734	0.139	0.861		
		0.010	0.200	0.790	0.110	0.890		
FOXP1	rs6803008	AA	AG	GG	A	G	0.616	0.679
		0.115	0.436	0.449	0.333	0.667		
		0.150	0.510	0.340	0.405	0.595		
DCBLD2	rs72932726	AA	AG	GG	A	G	0.983	0.761
		0.038	0.288	0.674	0.181	0.819		
		0.050	0.317	0.633	0.208	0.792		
KALRN	rs333332	TT	CT	CC	T	C	0.951	0.651
		0.138	0.424	0.438	0.350	0.650		
		0.119	0.416	0.465	0.327	0.673		
GATA2	rs9854612	AA	AG	GG	A	G	0.357	0.631
		0.051	0.449	0.500	0.276	0.724		
		0.057	0.314	0.629	0.214	0.786		
PCDH7	rs13109270	CC	CT	TT	C	T	0.427	< **0.001**
		0.324	0.188	0.488	0.419	0.581		
		0.317	0.238	0.445	0.436	0.564		
	rs13146789	TT	TG	GG	T	G	0.816	0.835
		0.202	0.494	0.304	0.449	0.551		
		0.152	0.494	0.354	0.399	0.601		
SNCA	rs356168	CC	CT	TT	C	T	0.897	0.681
		0.175	0.512	0.313	0.431	0.569		
		0.260	0.532	0.208	0.526	0.474		
	rs3910105	CC	CT	TT	C	T	0.110	0.655
		0.291	0.405	0.304	0.494	0.506		
		0.215	0.532	0.253	0.481	0.519		
UNC5C	rs9307160	CC	CT	TT	C	T	0.958	1.000
		0.118	0.448	0.434	0.342	0.658		
		0.079	0.406	0.515	0.282	0.718		
EGR1	rs7729723	GG	AG	AA	G	A	0.441	0.229
		0.227	0.506	0.267	0.480	0.520		
		0.170	0.560	0.270	0.450	0.550		
FoxP2	rs2396753	CC	AC	AA	C	A	0.724	0.410
		0.128	0.487	0.385	0.372	0.628		
		0.152	0.535	0.313	0.419	0.581		
SBSPON	rs1007750	CC	AC	AA	C	A	0.247	0.997
		0.182	0.376	0.442	0.370	0.630		
		0.099	0.426	0.475	0.312	0.688		
ASAP1	rs3057	TT	CT	CC	T	C	**0.032**	0.842
		0.152	0.645	0.203	0.475	0.525		
		0.228	0.485	0.287	0.470	0.530		
VLDLR	rs1454626	CC	AC	AA	C	A	0.702	0.995
		0.039	0.368	0.593	0.224	0.776		
		0.089	0.426	0.485	0.302	0.698		
BDNF	rs6265	AA	AG	GG	A	G	0.711	0.995
		0.013	0.237	0.750	0.131	0.869		
		0.040	0.313	0.647	0.197	0.803		
GRIN2B	rs3764030	AA	AG	GG	A	G	0.753	0.755
		0.038	0.418	0.544	0.247	0.753		
		0.030	0.343	0.627	0.202	0.798		

Note. P_{HWE} – a level of significance for the Hardy–Weinberg equilibrium test in the control group; *p* – a level of significance based on regression analysis is given for comparative models between a heterozygous genotype with a combined sample of homozygotes (overdominant model); genotype (a) and allele (b) frequencies are reported for the group of musicians (upper lane) and the control group (bottom lane), respectively. Statistically significant differences are marked in bold.

Table 3. Predictive models of liability to developing high-level musical abilities detected as a result of multiple logistic regression analyses

Predictor	Group	Reference group	Model 1		Model 2	
			β	*p*	β	*p*
GATA2 rs9854612	AG	GG	2.9	**0.002**	1.2	**0.026**
SNCA rs356168	TT	CC	3.5	**0.004**	2.3	**0.007**
SNCA rs3910105	TT	TC	2.6	**0.043**	1.5	**0.043**
ASAP1 rs3057	CT	TT+CC	2.2	**0.007**	1.1	**0.027**
VLDLR rs1454626	CC+AC	AA	–1.4	0.079	–1.2	**0.020**
DCBLD2 rs72932726	AA+AG	GG	–1.1	0.153	–	–
KALRN rs333332	TT	CC	–1.2	0.238	–-	–
FoxP2 rs2396753	AC	AA	–1.2	0.149	–	–
PCDH7 rs13146789	TG+TT	GG	–0.03	0.970	–	–
UNC5C rs9307160	CT	CC	0.4	0.574	–	–
SBSPON rs1007750	CC	AA	0.9	0.441	–	–
BDNF rs6265	AG	GG	–1.3	0.084	–	–
GRIN2B rs3764030	AG	GG	1.1	0.166	–	
EGR1 rs7729723	AG	AA	0.01	0.983	–	
FoxP1 rs6803008	AG	GG	–1.1	0.110	–	
OTOF rs4416176	AG	AA	1.5	0.158	–	–
Pseudo-r^2			0.380		0.188	
χ^2 Model			46.24		24.61	
AIC			131.31		123.94	
Model *p*-value			0.012		0.0018	

Note. r^2 – proportion of variance explaining a predictive ability of the model; β – regression coefficient; AIC – Akaike information criterion. Statistically significant differences are marked in bold.



The results of modeling ROC curves for logistic regression models. Including various combinations of the examined genetic loci as predictors (model 1 contains all 16 SNPs; model 2 contains five most significant SNPs in the *GATA2*, *SNCA*, *ASAP*, and *VLDLR* genes).

A detailed description of variables included in the models is given in Table 3. The values of areas under the curves (AUC) for each model are reported in the legend.

that a differential expression of these genes accompanies the formation of musical talent in the examined sample of Russians.

The assessment of a single-locus effect resulted in the detection of a statistically significant increase in the *ASAP1* rs3057 C/T genotype frequency in students with high-degree musical abilities compared with the control group, which assumes the existence of an overdominant model of inheritance characteristic for this SNP, since the highest positive effect was observed in the case of the simultaneous presence of both reference and alternative alleles. The examined gene polymorphism is known for its regulatory effect (Bryzgalov et al., 2013) and was previously detected as linked with absolute pitch (LOD = 3.46 in Europeans) (Theusch et al., 2009).

Existing data signify that the *ASAP1* gene is one of 30 genes differentially expressed in the prefrontal cortex as a response to ADHD treatment (dela Peña et al., 2014), indicating its relevance to the regulation of cognitive processes. In addition, published findings reported a link between other polymorphisms in the *ASAP1* gene with behavioral characteristics in normal and pathological states, including schizophrenia (Goes et al., 2015), cognitive decline (Mega Vascular Cognitive Impairment and Dementia (MEGAVCID) consortium, 2024), and education level (Okbay et al., 2022).

The examined *SBSPON* rs1007750 was initially identified as linked with absolute pitch segregation (Theusch et al., 2009) and, in addition, was present in a set of differentially expressed genes in postmortem tissues of the prefrontal cortex of schizophrenia patients (Maycox et al., 2009).

Another genetic variant, which was included in a prognostic model of musical talent, is rs9307160 located in the netrin receptor gene (*UNC5C*), which was linked with accumulation of musicality in Finnish families (Pulli et al., 2008). The netrin receptor UNC5C is expressed in dopaminergic neurons and participates in netrin signal transfer within neuronal migration during development and axonal guidance (Treccarichi et al., 2024). Previous research evidences a relation of certain mutations in the *UNC5C* gene to developing familial forms of Alzheimer's disease, schizophrenia, bipolar and depressive disorders, and other mental disorders (Treccarichi et al., 2024) due to shifted neuronal sensitivity to apoptosis.

Preserved cognitive functioning at the elderly age is related to polymorphisms in the *UNC5C* gene as well as to changes in its methylation and diminished levels of its expression in the dorsolateral prefrontal cortex, which predominantly mediates the changes in the presynaptic terminal (White et al., 2017). Despite the unknown functional significance of the rs9307160 in the regulation of the *UNC5C* gene expression, the abovementioned findings indicate its potential involvement in the development of both cognitive and behavioral reactions mediated by dopaminergic neurotransmission.

Genetic loci identified in GWAS

Genetic variants in the *VLDLR*, *FOXP1*, *OTOF*, and *GRIN2B* genes, which were included in our regression model, have been initially identified as connected with positive evolutionary selection of musical abilities (Liu et al., 2016). The *VLDLR* gene encodes low-density lipoprotein receptor involved in the reelin signaling cascade and representing the FoxP2 target. The latter is a transcription factor playing a significant regulatory role in vocalization and speech formation. In particular, a knockout of the *FoxP2* gene promotes a decline in the level of target VLDLR protein in model animals, whereas an enhanced expression of the *VLDLR* gene was determined in specific brain region of songbirds during learning and vocalization (Adam et al., 2016), which points to a potential role of the reelin signaling cascade in cognitive processes related to musicality. In addition, increased expression of the genes controlled by FoxP2 was observed in human peripheral blood as a result of listening to music (Kanduri et al., 2015).

According to previous research, the selected *VLDLR* rs1454626 gene polymorphism (C allele) is associated with enhanced lipoprotein level and body mass index; interestingly, this effect was strengthened with a simultaneous presence of the "risky" ε4 variant in the apolipoprotein E (*APOE*) gene (Crawford et al., 2008). The data obtained by our groups on the association of the rs1454626 C allele with a lower probability to develop musical abilities coincide with representations of the link between cognitive functioning and high-density lipoprotein-to-low-density lipoprotein ratio (Poliakova, Wellington, 2023) and tend towards a dominant model of inheritance of a negative effect of the C allele on developing high-level musical aptitude.

Transcription factor genes examined within the present study, i. e. *FoxP1* (rs6803008) and *FoxP2* (rs2396753), regulate the expression of a high number of genes, including reelin pathway genes (*VLDLR*). *De novo* mutations in the *FoxP1* gene underlie the etiology of FoxP1 syndrome – an autosomal-dominant disease characterized by abnormalities in the cognitive and language abilities, communication, and behavior, including ASD, ADHD, anxiety disorders, and oromotor dysfunctions (Rappold et al., 2023). Moreover, FoxP1 mediates vocal preferences and motivation for active listening to bird singing in female zebra finches (Heim et al., 2023), which may indicate a potential role of this transcription factor in the formation of human musicality.

Another significant member of the processes of cognitive functioning and synaptic plasticity is the glutamatergic system, an important component of which is the *GRIN2B* gene encoding the glutamate receptor subunit, examined in the present study. An increase in the mRNA level of this gene was recorded as a result of songbird vocalization (in particular, zebra finches) (So et al., 2019). According to published studies, the examined allelic variant rs3764030 (A allele) in the *GRIN2B* gene was also associated with improved information processing speed due to enhanced gene expression (Jiang et al., 2017). Based on our findings, the rs3764030 A allele is linked with a positive prognostic effect toward musical abilities under the framework of the regression model, which is consistent with the abovementioned data.

The otoferlin gene was also identified as one with a selective advantage for evolutionary consolidation of musical abilities (Liu et al., 2016). This gene encodes a transmembrane protein involved in the exocytosis of the inner ear cells, indicating its potential involvement in musicality. Since no previous research demonstrated an association of a distinct polymorphism in the *OTOF* gene with musical traits, we have selected rs4416176, since its allelic variants had an earlier reported association with social abilities in ASD patients (Hu et al., 2011).

The role of the *OTOF* in hearing impairment gene has been repeatedly examined, whereas the introduction of the *OTOF* gene therapy in children with congenital deafness demonstrated a pronounced positive effect on hearing restoration, improvement of auditory and speech abilities, and music perception, even compared with patients with cochlear implantation (Cheng et al., 2025).

Genetic loci in the *DCBLD2* (rs72932726) and *KALRN* (rs333332) genes have been selected based on GWAS results as associated with the thickness of Heschl's gyrus (Cai et al., 2014) – an auditory cortex region involved in sounds and speech perception, linguistic tone, and regulating inner ear development. It should be noted that musicians are characterized by exaggerated gray matter volume in the right and left parts of the auditory cortex, while an increase in the right Heschl's gyrus is typical for musicians with absolute pitch (Wengenroth et al., 2014). rs72932726 in the *DCBLD2* gene, which was introduced in the final regression model in the present study, is also linked with the structural changes in brain regions related to speech and language processing (Cai et al., 2014).

Another examined gene is *KALRN*, encoding kalirin kinase – a synaptic regulator, which is responsible for dendrite morphogenesis, axonal growth, and brain connectivity (Parnell et al., 2021). Impaired regulation of the *KALRN* gene was observed in cognitive deficit, schizophrenia, and ASD (Parnell et al., 2021), while the genetic variant rs333332 in the *KALRN* gene demonstrated the association with the thickness of the left Heschl's gyrus (Cai et al., 2014).

Polymorphisms in differentially expressed genes

The final regression model of liability to musical aptitude consisted of gene variants previously demonstrating functional changes in the level of gene expression. Namely, lower expression of the transcription factor *FoxP2* gene, known for its important role in neuronal differentiation, brain and speech development, was observed in the case of impaired songbird vocalization (Shi et al., 2013), while structural changes in this gene were related to clinical symptoms of schizophrenia (Salmón-Gómez et al., 2025) and speech perception (Ocklenburg et al., 2013). In addition, based on a recent systematic review, the rs2396753 C allele was associated with a reduced gray matter density in the brain and was more frequent in individuals with verbal hallucinations accompanying schizophrenia (Salmón-Gómez et al., 2025). At the same time, our findings indicate the association of the rs2396753 C allele with absent musicality.

It is known that early growth response protein (EGR1) involved in neuronal plasticity and mediated by differential expression microRNAs miR-23a and miR-23b is a transcriptional regulator of the *FoxP2* gene (Nair et al., 2021). Multiple studies point to the involvement of EGR1 protein in memory and learning, while its expression in different brain regions demonstrates a stress-inducible pattern, thus accompanying such mental disorders as schizophrenia, depression, and Alzheimer's disease (Gallo et al., 2018).

Brain-derived neurotrophic factor (*BDNF*) is one of the most frequently analyzed genes in the context of cognitive abilities and disabilities involved in the regulation of neurogenesis and synaptic plasticity, enhanced expression of which is observed after listening to music (Li et al., 2010). According to model 1, the rs6265 (Val66Met) A/G (Met/Val) genotype is characterized by a lower odds ratio of developing musical talent compared with the G/G (Val/Val) genotype. Multiple studies evidence that the rs6265 Met variant (coded by the A alelle) is linked with reduced gene expression and diminished gray matter volume in brain regions participating in memory formation, self-control and emotional regulation (Kunikullaya et al., 2025).

Previously, a positive effect of musical listening on increased BDNF level was also determined, which promoted a decline in depressive symptoms (Yeh et al., 2015). Such dependence is suggested to be caused by the involvement of rs6265 in the regulation of neuroplasticity in the auditory cortex of the brain. The analysis of mismatch negativity (MMN) as a value reflecting the process of automatic error detection by the auditory cortex and preconception processes, which was conducted during electroencephalography, revealed that the most significant changes in MMN peaks were characteristic of musicians carrying the Val/Val genotype compared with musicians with the Met allele (Bonetti et al., 2023).

In turn, music perception affects changes in alpha-synuclein (*SNCA*) gene expression at the molecular level (Kanduri et al., 2015). Encoded protein is involved in regulation of dopamine biosynthesis via altering its reuptake from the synaptic cleft by transporter protein and demonstrates its link with the ethiopathogenesis of Parkinson's disease (PD). At the SNPs level, existing data evidence the association of the *SNCA* rs356168 C/C genotype with increased risk of PD and an enhanced alpha-synuclein level in CD45+ blood cells of patients (Emelyanov et al., 2018). It is assumed that rs356168 regulates a ratio between long (based on the 3′ region) and short transcripts of alpha-synuclein, therefore determining its ability to oligomerize and PD development (Rhinn et al., 2012).

In the present study we demonstrated the effect of alternative rs356168 and rs3910105 T/T genotypes on musical talent within a recessive model of inheritance, which suggests that a favorable effect can be observed only in the presence of both copies of the T allele. The data obtained are unsurprising, since they coincide with conclusions published in a recent systematic review (Navarro et al., 2023) pointing to a link between dopaminergic neurotransmission, alpha-synuclein

level, the positive effect of music listening and reproduction (Kanduri et al., 2015; Järvelä, 2018), and clinical manifestations of neurodegenerative diseases.

Moreover, the obtained evidence on a significant cumulative effect of the examined genes coincides with the data of a meta-analysis (Kunikullaya et al., 2025), which confirmed the role of brain-derived neurotrophic factor, alpha-synuclein, and GATA2 in modified neurogenesis and neuromediation caused by music listening and reproduction. The last one initiates exaggerated levels of neurotrophins, neuroprotectors, and synaptic plasticity regulators, improves immune functions, and reduces stress.

The conducted study has several limitations. Firstly, the analyzed sample has a small size; nevertheless, it is homogeneous by age, ethnic origin, and level of musical education. Secondly, the number of the studied genetic variants is limited and should be expanded for future research. In addition, several genetic loci were taken based on literature data on their association with similar phenotypes without information on their functional significance. Thirdly, despite the multifactorial nature of musical abilities, we had no data to conduct mathematical modeling accounting for potential environmental predictors such as the qualitative and quantitative specifics of musical education in childhood, the presence of a musical environment, socioeconomic status, etc. Fourthly, it is necessary to consider the pilot character of the present study and insufficient statistical power in the analysis of the predictive mode, including 16 SNPs, which may reflect the presence of type I and II errors.

Further research in this field should focus on expanding the panel of examined genetic loci to study the predisposition to high-level musical abilities and assessing their functional significance in the context of regulating gene expression.

Conclusion

Within the framework of replication association analysis, using a set of genetic loci previously identified in European samples, we constructed a prognostic model including five SNPs estimating a liability to musical aptitude in Russians at a significant level of sensitivity and specificity.

Despite polymorphic loci being selected based on various approaches (GWAS, GWLS, differential expression data), the examined genes frequently overlap, and encoded proteins are frequently included in the same molecular pathways. Moreover, a significant part of the examined genes encode transcription factors, pointing to a potential impact of various molecular pathways regulated by these factors in developing musical abilities.

The data obtained indirectly evidence the interaction of the examined genes, confirming that musical aptitude is related to the specificity of dopaminergic and GABAergic neurotransmission, synaptic plasticity, neurogenesis, and reelin signaling cascade. Moreover, a commonality of molecular genetic factors involved in the ethiopathogenesis of schizophrenia, cognitive impairments, ASD, affective disorders, and musical talent was determined.

References

Adam I., Mendoza E., Kobalz U., Wohlgemuth S., Scharff C. FoxP2 directly regulates the reelin receptor *VLDLR* developmentally and by singing. *Mol Cell Neurosci*. 2016;74:96-105. doi 10.1016/j.mcn.2016.04.002

Aikina L.P. On the development of children's musical abilities. *Mir Nauki, Kul'tury, Obrazovaniya = World Sci Cult Educ*. 2017;3(64):221-223 (in Russian)

Aung M.T., Bakulski K.M., Feinberg J.I., F. Dou J., D. Meeker J., Mukherjee B., Loch-Caruso R., Ladd-Acosta C., Volk H.E., Croen L.A., Hertz-Picciotto I., Newschaffer C.J., Fallin M.D. Maternal blood metal concentrations and whole blood DNA methylation during pregnancy in the Early Autism Risk Longitudinal Investigation (EARLI). *Epigenetics*. 2022;17(3):253-268. doi 10.1080/15592294.2021.1897059

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

Bonetti L., Bruzzone S.E.P., Paunio T., Kantojärvi K., Kliuchko M., Vuust P., Palva S., Brattico E. Moderate associations between *BDNF* Val66Met gene polymorphism, musical expertise, and mismatch negativity. *Heliyon*. 2023;9(5):e15600. doi 10.1016/j.heliyon.2023.e15600

Bryzgalov L.O., Antontseva E.V., Matveeva M.Y., Shilov A.G., Kashina E.V., Mordvinov V.A., Merkulova T.I. Detection of regulatory SNPs in human genome using ChIP-seq ENCODE data. *PLoS One*. 2013;8(10):e78833. doi 10.1371/journal.pone.0078833

Cai D.C., Fonteijn H., Guadalupe T., Zwiers M., Wittfeld K., Teumer A., Hoogman M., … Homuth G., Fisher S.E., Grabe H.J., Francks C., Hagoort P. A genome-wide search for quantitative trait loci affecting the cortical surface area and thickness of Heschl's gyrus. *Genes Brain Behav*. 2014;13(7):675-685. doi 10.1111/gbb.12157

Cheng X., Zhong J., Zhang J., Cui C., Jiang L., Liu Y.W., Chen Y., … Wang W., Chen B., Chen Z.Y., Li H., Shu Y. Gene therapy vs cochlear implantation in restoring hearing function and speech perception for individuals with congenital deafness. *JAMA Neurol*. 2025;82(9):941-951. doi 10.1001/jamaneurol.2025.2053

Conard N.J., Malina M., Münzel S.C. New flutes document the earliest musical tradition in southwestern Germany. *Nature*. 2009;460(7256):737-740. doi 10.1038/nature08169

Crawford D.C., Nord A.S., Badzioch M.D., Ranchalis J., McKinstry L.A., Ahearn M., Bertucci C., … Schellenberg G.D., Nickerson D.A., Heagerty P.J., Wijsman E.M., Jarvik G.P. A common *VLDLR* polymorphism interacts with *APOE* genotype in the prediction of carotid artery disease risk. *J Lipid Res*. 2008;49(3):588-596. doi 10.1194/jlr.M700409-JLR200

dela Peña I., Kim H.J., Sohn A., Kim B.N., Han D.H., Ryu J.H., Shin C.Y., Noh M., Cheong J.H. Prefrontal cortical and striatal transcriptional responses to the reinforcing effect of repeated methylphenidate treatment in the spontaneously hypertensive rat, animal model of attention-deficit/hyperactivity disorder (ADHD). *Behav Brain Funct*. 2014;10:17. doi 10.1186/1744-9081-10-17

Drnevich J., Replogle K.L., Lovell P., Hahn T.P., Johnson F., Mast T.G., Nordeen E., … Ball G.F., Brenowitz E.A., Wade J., Mello C.V., Clayton D.F. Impact of experience-dependent and -independent factors on gene expression in songbird brain. *Proc Natl Acad Sci USA*. 2012;109(Suppl. 2):17245-17252. doi 10.1073/pnas.1200655109

Emelyanov A., Kulabukhova D., Garaeva L., Senkevich K., Verbitskaya E., Nikolaev M., Andoskin P., … Timofeeva A., Prakhova L., Ilves A., Vlasova I., Pchelina S. SNCA variants and alpha-synuclein level in CD45+ blood cells in Parkinson's disease. *J Neurol Sci*. 2018;395:135-140. doi 10.1016/j.jns.2018.10.002

Gallo F.T., Katche C., Morici J.F., Medina J.H., Weisstaub N.V. Immediate early genes, memory and psychiatric disorders: focus on

c-fos, *Egr1* and *arc*. *Front Behav Neurosci*. 2018;12:79. doi 10.3389/fnbeh.2018.00079

Goes F.S., McGrath J., Avramopoulos D., Wolyniec P., Pirooznia M., Ruczinski I., Nestadt G., … Lencz T., Darvasi A., Mulle J.G., Warren S.T., Pulver A.E. Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am J Med Genet B Neuropsychiatr Genet*. 2015;168(8):649-659. doi 10.1002/ajmg.b.32349

Heim F., Fisher S.E., Scharff C., Ten Cate C., Riebel K. Effects of cortical FoxP1 knockdowns on learned song preference in female zebra finches. *eNeuro*. 2023;10(3):ENEURO.0328-22.2023. doi 10.1523/ENEURO.0328-22.2023

Hu V.W., Addington A., Hyman A. Novel autism subtype-dependent genetic variants are revealed by quantitative trait and subphenotype association analyses of published GWAS data. *PLoS One*. 2011;6(4):e19067. doi 10.1371/journal.pone.0019067

Järvelä I. Genomics studies on musical aptitude, music perception, and practice. *Ann N Y Acad Sci*. 2018;1423(1):82-91. doi 10.1111/nyas.13620

Jiang Y., Lin M.K., Jicha G.A., Ding X., McIlwrath S.L., Fardo D.W., Broster L.S., Schmitt F.A., Kryscio R., Lipsky R.H. Functional human *GRIN2B* promoter polymorphism and variation of mental processing speed in older adults. *Aging*. 2017;9(4):1293-1306. doi 10.18632/aging.101228

Kanduri C., Kuusi T., Ahvenainen M., Philips A.K., Lähdesmäki H., Järvelä I. The effect of music performance on the transcriptome of professional musicians. *Sci Rep*. 2015;5:9506. doi 10.1038/srep09506

Kazantseva A.V. Molecular grounds of temperament and personal features: Cand. Sci. (Biol.) Dissertation. Ufa, 2008 (in Russian)

Kazantseva A.V., Enikeeva R.F., Toropova A.V., Zakharov I.M., Ismatullina V.I., Khusnutdinova E.K., Malykh S.B. A replication study of genetic variants associated with high-level musical aptitude. *Res Results Biomed*. 2023;9(2):181-190. doi 10.18413/2658-6533-2023-9-2-0-3

Koks S., Pfaff A.L., Bubb V.J., Quinn J.P. Transcript variants of genes involved in neurodegeneration are differentially regulated by the APOE and MAPT haplotypes. *Genes* (*Basel*). 2021;12(3):423. doi 10.3390/genes12030423

Kunikullaya U.K., Pranjić M., Rigby A., Pallás-Ferrer I., Anand H., Kunnavil R., Jaschke A.C. The molecular basis of music-induced neuroplasticity in humans: a systematic review. *Neurosci Biobehav Rev*. 2025;175:106219. doi 10.1016/j.neubiorev.2025.106219

Kutikhin A.G., Yuzhalin A.E., Ponasenko A.V. How to analyze and present genetic epidemiology data in candidate studies. *Fundamental'naya i Clinicheskaya Meditsina = Fundamental and Clinical Medicine*. 2017;2(2):77-82. doi 10.23946/2500-0764-2017-2-2-77-82 (in Russian)

Li W.J., Yu H., Yang J.M., Gao J., Jiang H., Feng M., Zhao Y.X., Chen Z.Y. Anxiolytic effect of music exposure on $BDNF^{Met/Met}$ transgenic mice. *Brain Res*. 2010;1347:71-79. doi 10.1016/j.brainres.2010.05.080

Liu X., Kanduri C., Oikkonen J., Karma K., Raijas P., Ukkola-Vuoti L., Teo Y.Y., Järvelä I. Detecting signatures of positive selection associated with musical aptitude in the human genome. *Sci Rep*. 2016;6:21198. doi 10.1038/srep21198

Maycox P.R., Kelly F., Taylor A., Bates S., Reid J., Logendra R., Barnes M.R., … Akbar M.T., Hirsch S., Mortimer A.M., Barnes T.R., de Belleroche J. Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Mol Psychiatry*. 2009;14(12):1083-1094. doi 10.1038/mp.2009.18

Mega Vascular Cognitive Impairment and Dementia (MEGAVCID) consortium. A genome-wide association meta-analysis of all-cause and vascular dementia. *Alzheimers Dement.* 2024; 20(9):5973-5995. doi 10.1002/alz.14115

Nair P.S., Raijas P., Ahvenainen M., Philips A.K., Ukkola-Vuoti L., Järvelä I. Music-listening regulates human microRNA expression. *Epigenetics*. 2021;16(5):554-566. doi 10.1080/15592294.2020.1809853

Navarro L., Gómez-Carballa A., Pischedda S., Montoto-Louzao J., Viz-Lasheras S., Camino-Mera A., Hinault T., Martinón-Torres F., Salas A. Sensogenomics of music and Alzheimer's disease: an interdisciplinary view from neuroscience, transcriptomics, and epigenomics. *Front Aging Neurosci*. 2023;15:1063536. doi 10.3389/fnagi.2023.1063536

Ocklenburg S., Arning L., Gerding W.M., Epplen J.T., Güntürkün O., Beste C. *FOXP2* variation modulates functional hemispheric asymmetries for speech perception. *Brain Lang*. 2013;126(3):279-284. doi 10.1016/j.bandl.2013.07.001

Oikkonen J., Huang Y., Onkamo P., Ukkola-Vuoti L., Raijas P., Karma K., Vieland V.J., Järvelä I. A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. *Mol Psychiatry*. 2015;20(2):275-282. doi 10.1038/mp.2014.8

Okbay A., Wu Y., Wang N., Jayashankar H., Bennett M., Nehzati S.M., Sidorenko J., … Turley P., Visscher P.M., Beauchamp J.P., Benjamin D.J., Young A.I. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet*. 2022;54(4):437-449. doi 10.1038/s41588-022-01016-z

Parnell E., Shapiro L.P., Voorn R.A., Forrest M.P., Jalloul H.A., Loizzo D.D., Penzes P. KALRN: a central regulator of synaptic function and synaptopathies. *Gene*. 2021;768:145306. doi 10.1016/j.gene.2020.145306

Poliakova T., Wellington C.L. Roles of peripheral lipoproteins and cholesteryl ester transfer protein in the vascular contributions to cognitive impairment and dementia. *Mol Neurodegener*. 2023;18(1):86. doi 10.1186/s13024-023-00671-y

Pulli K., Karma K., Norio R., Sistonen P., Göring H.H., Järvelä I. Genome-wide linkage scan for loci of musical aptitude in Finnish families: evidence for a major locus at 4q22. *J Med Genet*. 2008;45(7):451-456. doi 10.1136/jmg.2007.056366.

Rappold G., Siper P., Kostic A., Braden R., Morgan A., Koene S., Kolevzon A. FOXP1 syndrome. In: GeneReviews® [Internet]. Seattle (WA): University of Washington, 2023

Rhinn H., Qiang L., Yamashita T., Rhee D., Zolin A., Vanti W., Abeliovich A. Alternative α-synuclein transcript usage as a convergent mechanism in Parkinson's disease pathology. *Nat Commun*. 2012;3:1084. doi 10.1038/ncomms2032

Salmón-Gómez G., Suárez-Pinilla P., Setién-Suero E., Martínez-Asensi C., Ayesa-Arriola R. FoxP2 and Schizophrenia: a systematic review. *J Psychiatr Res*. 2025;190:205-215. doi 10.1016/j.jpsychires.2025.07.016

Shi Z., Luo G., Fu L., Fang Z., Wang X., Li X. miR-9 and miR-140-5p target *FoxP2* and are regulated as a function of the social context of singing behavior in zebra finches. *J Neurosci*. 2013;33(42):16510-16521. doi 10.1523/JNEUROSCI.0838-13.2013

So L.Y., Munger S.J., Miller J.E. Social context-dependent singing alters molecular markers of dopaminergic and glutamatergic signaling in finch basal ganglia Area X. *Behav Brain Res*. 2019;360:103-112. doi 10.1016/j.bbr.2018.12.004

Theusch E., Basu A., Gitschier J. Genome-wide study of families with absolute pitch reveals linkage to 8q24.21 and locus heterogeneity. *Am J Hum Genet*. 2009;85(1):112-119. doi 10.1016/j.ajhg.2009.06.010

Treccarichi S., Failla P., Vinci M., Musumeci A., Gloria A., Vasta A., Calabrese G., Papa C., Federico C., Saccone S., Calì F. *UNC5C*: novel gene associated with psychiatric disorders impacts dysregulation of axon guidance pathways. *Genes*. 2024;15(3):306. doi 10.3390/genes15030306

Ukkola-Vuoti L., Kanduri C., Oikkonen J., Buck G., Blancher C., Raijas P., Karma K., Lähdesmäki H., Järvelä I. Genome-wide copy number variation analysis in extended families and unrelated individuals characterized for musical aptitude and creativity in music. *PloS One*. 2013;8(2):e56356. doi 10.1371/journal.pone.0056356

Wengenroth M., Blatow M., Heinecke A., Reinhardt J., Stippich C., Hofmann E., Schneider P. Increased volume and function of right auditory cortex as a marker for absolute pitch. *Cereb Cortex.* 2014; 24(5):1127-1137. doi 10.1093/cercor/bhs391

White C.C., Yang H.S., Yu L., Chibnik L.B., Dawe R.J., Yang J., Klein H.U., … Honer W.G., Sperling R.A., Schneider J.A., Bennett D.A., De Jager P.L. Identification of genes associated with dissociation of cognitive performance and neuropathological burden: multistep analysis of genetic, epigenetic, and transcriptional data. *PLoS Med*. 2017;14(4):e1002287. doi 10.1371/journal.pmed.1002287

Yeh S.H., Lin L.W., Chuang Y.K., Liu C.L., Tsai L.J., Tsuei F.S., Lee M.T., Hsiao C.Y., Yang K.D. Effects of music aerobic exercise on depression and brain-derived neurotrophic factor levels in community dwelling women. *Biomed Res Int*. 2015;2015:135893. doi 10.1155/2015/135893

Conflict of interest. The authors declare no conflict of interest.
Received September 19, 2025. Revised February 6, 2026. Accepted February 25, 2026.

MEDICAL GENETICS
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):482-489

doi 10.18699/vjgb-26-52

Diagnostic efficiency of whole exome sequencing in the search for genetic causes of hereditary diseases in Yugra (West Siberia, Russia)

M.Yu. Donnikov [1] ✉, P.A. Suchko [2, 3], A.V. Morozkina [1], L.N. Kolbasin [1, 4], E.A. Popova [1, 4], S.I. Papanov [1, 4], Yu.S. Koshevaya [5], L.G. Danilov [3], Yu.A. Eismont [6], O.S. Glotov [2, 6], L.V. Kovalenko [1]

[1] Medical Institute of Surgut State University, KHMAO-Yugra, Surgut, Russia
[2] Research Institute of Obstetrics, Gynecology, and Reproductology named after D.O. Ott, St. Petersburg, Russia
[3] Saint Petersburg State University, St. Petersburg, Russia
[4] KHMAO-Yugra Surgut Regional Clinical Center for Maternity and Childhood Protection, Medical Genetic Counseling Service, Surgut, KHMAO-Yugra, Russia
[5] Saint-Petersburg State Medical Diagnostic Center (Genetic Medical Center), St. Petersburg, Russia
[6] Federal Scientific and Clinical Center of Infectious Diseases of the Federal Medical and Biological Agency, St. Petersburg, Russia
✉ donnikov@gmail.com

Abstract. Whole-exome sequencing (WES) has revolutionized the diagnostics of hereditary diseases, yet its efficacy varies across populations. Data on the genetic architecture of rare hereditary disorders in many Russian regions, including the ethnically diverse Khanty-Mansi Autonomous Okrug (Yugra) are scarce. The aim of this study was to evaluate the diagnostic yield of WES for identifying genetic variants associated with hereditary disorders in this ethnically heterogeneous population. The study involved 286 probands with suspected hereditary disorders observed by regional geneticists in the years 2021–2024. WES was performed on the DNBSEQ-G50 platform (MGI, China). Bioinformatic analysis included variant calling and annotation using population databases and pathogenicity prediction tools. Identified variants were classified according to ACMG/Russian Medical Genetics Society guidelines and correlated with clinical phenotypes. Molecular genetic diagnoses were categorized as definitive, partial, potential (based on variants of unknown significance), or unknown. The examined cohort was predominantly pediatric, the most common clinical indications were neurological, dysmorphic, and metabolic disorders. Definitive molecular diagnoses were established in 24.8 % of patients. Inclusion of potential diagnoses increased the total yield to 48.6 %. Diagnostic efficacy varied significantly among disease categories ranging from 58.3 % for renal disorders to 0 % for neurodevelopmental disorders. A total of 420 unique variants were analyzed, and missense changes were the most frequent among clinically significant findings. The most commonly implicated genes were *ATP7B*, *GJB2*, *ABCA4*, and *GALT*. The study results indicate that WES is an effective first-tier molecular tool for a wide range of suspected hereditary diseases in the Yugra population, with a diagnostic yield comparable to similar studies abroad. The findings support the utility of WES in diverse populations and highlight the potential for increasing yield through trio-WES and periodic data reanalysis.

Key words: whole exome sequencing; hereditary diseases; new generation sequencing; genetic counseling; diagnostic effectiveness; molecular genetic diagnosis

For citation: Donnikov M.Yu., Suchko P.A., Morozkina A.V., Kolbasin L.N., Popova E.A., Papanov S.I., Koshevaya Yu.S., Danilov L.G., Eismont Yu.A., Glotov O.S., Kovalenko L.V. Diagnostic efficiency of whole exome sequencing in the search for genetic causes of hereditary diseases in Yugra (West Siberia, Russia). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):482-489. doi 10.18699/vjgb-26-52

Funding. The study was supported by the Foundation for Scientific and Technological Development of Yugra, project 2023-574-05.

Диагностическая эффективность полноэкзомного секвенирования в поиске генетических причин наследственных заболеваний в ХМАО-Югре (Западная Сибирь, Россия)

М.Ю. Донников [1] ✉, П.А. Сучко [2, 3], А.В. Морозкина [1], Л.Н. Колбасин [1, 4], Е.А. Попова [1, 4], С.И. Папанов [1, 4], Ю.С. Кошевая [5], Л.Г. Данилов [3], Ю.А. Эйсмонт [6], О.С. Глотов [2, 6], Л.В. Коваленко [1]

[1] Медицинский институт, Сургутский государственный университет, Сургут, ХМАО-Югра, Россия
[2] Научно-исследовательский институт акушерства, гинекологии и репродуктологии им. Д.О. Отта, Санкт-Петербург, Россия
[3] Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
[4] Сургутский окружной клинический центр охраны материнства и детства, Россия, Сургут, ХМАО-Югра, Россия
[5] Диагностический центр (медико-генетический), Санкт-Петербург, Россия
[6] Федеральный научно-клинический центр инфекционных болезней Федерального медико-биологического агентства, Санкт-Петербург, Россия
✉ donnikov@gmail.com

Аннотация. Полноэкзомное секвенирование произвело революцию в диагностике наследственных заболеваний, однако его эффективность варьирует в разных популяциях. Данные о генетической архитектуре редких наследственных заболеваний во многих регионах России, включая этнически разнообразный Ханты-Мансийский автономный округ – Югру, ограниченны. Настоящее исследование направлено на оценку диагностической ценности полноэкзомного секвенирования для выявления генетических вариантов, ассоциированных с наследственными заболеваниями в этой этнически неоднородной популяции. В исследование включили 286 пробандов с предполагаемыми наследственными заболеваниями, наблюдаемых генетиками региона в период с 2021 по 2024 г. Полноэкзомное секвенирование было выполнено на платформе DNBSEQ-G50 (MGI, Китай). Биоинформатический анализ включал выявление и аннотацию вариантов с использованием популяционных баз данных и инструментов прогнозирования патогенности. Найденные варианты были распределены по категориям в соответствии с рекомендациями ACMG/Российского общества медицинских генетиков и коррелировали с клиническими фенотипами. Молекулярно-генетические диагнозы были классифицированы как полные, частичные, потенциальные (на основании вариантов неизвестной значимости) или отсутствующие. Обследованная когорта преимущественно состояла из детей, наиболее частыми клиническими показаниями были неврологические, дисморфические и метаболические нарушения. Окончательный молекулярно-генетический диагноз был установлен у 24.8 % пациентов. С учетом потенциальных заключений общая значимость исследования увеличилась до 48.6 %. Диагностическая эффективность значительно варьировала в зависимости от категории заболеваний: от 58.3 % для заболеваний почек до 0 % для нарушений нейрогенеза. Проанализировано в общей сложности 420 уникальных вариантов генов, при этом миссенс-варианты были самыми встречающимися среди клинически значимых результатов. Наиболее часто выявлялись гены *ATP7B*, *GJB2*, *ABCA4*, *GALT*. Результаты исследования показали, что полноэкзомное секвенирование – эффективный молекулярный тест первой линии для широкого спектра предполагаемых наследственных заболеваний у жителей Югры с диагностической ценностью, сопоставимой с аналогичными исследованиями за рубежом. Полученные данные подтверждают результативность полноэкзомного секвенирования в гетерогенной популяции и подчеркивают потенциал повышения диагностической ценности за счет формата «трио» и периодического повторного анализа данных.

Ключевые слова: полноэкзомное секвенирование; наследственные заболевания; секвенирование нового поколения; генетическое консультирование; эффективность диагностики; молекулярная диагностика

Introduction

Due to the development of molecular genetic diagnostic methods, especially high-throughput sequencing, significant progress has been made in determining the molecular nature of hereditary diseases over the past 15 years. In particular, the introduction of whole-genome (WGS) and whole-exome (WES) sequencing has enabled both large-scale population projects to describe the frequencies of genetic variants and the analysis of complex clinical cases with unclear disease etiology.

Although technological advances allow for more precise results and cheaper research, the clinical interpretation of genomic data necessary for a specific patient has become a new challenge (Petersen et al., 2017). Given the enormous amount of data obtained from WGS, the main obstacles to implementing this method are the difficulty of interpretation in the context of a specific disease, the high cost, and the burden on laboratory infrastructure with the general diagnostic efficiency of about 40 % (Stranneheim et al., 2021).

In terms of diagnostic efficiency, WES allows focusing on the analysis of protein-coding regions of genes. It is hardly inferior to WGS, and its advantage is less laborious interpretation. However, its key disadvantages are the loss of information outside of exons, the uneven coverage of gene sequences by probes, and limitations in analyzing copy number variations (CNVs) (Wang et al., 2017; Ross et al., 2020).

Since the diagnostic efficiency of WES for hereditary diseases significantly exceeds that of targeted panels, its main niche includes cases with suspected rare genetic diseases, diseases with recently identified or extended genes, a suspected heterogeneous disease in a young child, and negative results from other diagnostic methods (Okuneva et al., 2020).

According to OMIM Morbid Map Scorecard (www.omim.org/statistics/geneMap, accessed September 25, 2025), there are currently 6,619 phenotypes with 4,661 involved genes corresponding to single gene disorders and traits with known molecular basis. Most common forms of monogenic diseases vary significantly not only among different countries but also within distinct regions of a country. Therefore, it is of utmost importance to study the spectrum of genetic variants and monogenic diseases in all regions of Russia (Zinchenko et al., 2019), taking into account the vast diversity of subpopulations that have been practically unexplored by large-scale studies (Barbitoff et al., 2024).

The aim of this study was to identify the genetic causes of the most common hereditary diseases in the West Siberian Russian Khanty-Mansi Autonomous Okrug (Yugra) and to assess the diagnostic efficiency of WES in this population.

Materials and methods

Patient enrollment. Patient selection was carried out from 2021 to 2024 inclusive, as they were admitted for consultation at the regional medical genetic service located in Surgut (Yugra). For the study, 286 probands of various ages and ethnicities were selected according to the following inclusion criteria: the suspected monogenic nature of the disease (early age of manifestation, indications in the family history, inherited nature of the disorder, rare and specific symptoms of multiple organ damage), as well as already diagnosed hereditary diseases with an unspecified molecular cause based on clinical picture and common biochemical tests. For molecular genetic testing, 5 mL of peripheral blood (with EDTA) were taken from patients and transferred for processing to the Yugra biobank laboratory, established at the Surgut State University.

All participants (or their official representatives) provided informed consent for participation in the study and personal data processing. The study was conducted in accordance with the Helsinki Declaration.

DNA extraction, library preparation, and sequencing. Genomic DNA was extracted from peripheral blood using a MagPure Blood DNA Kit (Magen, China). Whole-exome libraries were prepared using a KAPA HyperPlus Kit and KAPA HyperExome probes (Roche, United States). The libraries were converted with a MGIEasy Universal Library Conversion Kit (MGI, China) and sequenced on a DNBSEQ-G50 system (MGI) in the paired-end mode with the read length of 150 bp, following the manufacturer's recommendations.

Bioinformatics analysis. Samples with an average 70× coverage of target regions and at least 10× coverage width of 98 % were included in further analysis. Samples that did not pass quality control were sent for repeated library preparation and sequencing.

Mapping of obtained reads to the human reference genome (hg19 in 2021–2022 and hg38 in 2023–2024) was performed using BWA (0.7.16) (Li, 2011). Post-processing steps of alignments, variant calling, and filtering were carried out using the Genome Analysis Toolkit (Van der Auwera, O'Connor, 2020).

Variant annotation for all known transcripts of each gene from the RefSeq database was performed using snpEff (v.5.1) (Cingolani et al., 2012), with population frequencies of identified variants from The 1000 Genomes Project and gnomAD samples added to the annotation. Pathogenicity prediction of sequence variants was performed using DANN (Quang et al., 2015), GERP (Davydov et al., 2010), REVEL (Ioannidis et al., 2016), SIFT (Kumar et al., 2009), PolyPhen2 (Adzhubei et al., 2010), PrimateAI (Sundaram et al., 2018). Algorithms AdaBoost (Pashaei et al., 2016) and SpliceAI (Strauch et al., 2022) were also used for assessing the impact of variants on the splice site function.

Clinical data interpretation. To assess the clinical relevance of identified sequence variants the following resources were used: OMIM database, disease-specific databases (if available), and scientific literature data. Reports included only variants that had a possible relationship to the patient's clinical manifestations or met other criteria specified in the physician's referral. Polymorphisms classified as benign or likely benign were excluded from the report.

Identified variants were categorized according to criteria from ACMG guidelines (Richards et al., 2015) and the Russian Society of Medical Genetics (Ryzhkova et al., 2019) as pathogenic (P), likely pathogenic (LP), variants of unknown (clinical) significance (VUS), and asymptomatic carriers (AC). The last category included heterozygous pathogenic and likely pathogenic variants not causing the disease in the proband but strongly associated with other monogenic diseases.

Classification of molecular diagnoses. Based on the number of identified variants in each gene and inheritance type, molecular testing results were categorized as follows:

– complete molecular genetic diagnosis (MGD): at least one heterozygous or hemizygous P/LP variant with a dominant or X-linked type of inheritance, as well as one homozygous P/LP variant or two (potentially) compound heterozygous P/LP variants with a recessive type of inheritance;
– partial MGD: one heterozygous P/LP variant in a gene associated with an autosomal recessive disease;
– potential MGD: one VUS clearly associated with the phenotype with autosomal dominant or X-linked inheritance, or two VUSes clearly associated with the phenotype with autosomal recessive inheritance;
– no MGD: absence of identified variants, one P/LP variant or VUS in a gene with autosomal recessive inheritance, incomplete relation of VUS associated with autosomal dominant, or X-linked inheritance to clinical presentation;
– incidental findings: only AC variants.

Statistical analysis and graph plotting were performed using the RStudio programming environment. Patient data and genetic variants were imported from a proprietary database (Glotov et al., 2025). They are available upon request.

Results

Characteristics of the examined cohort of patients

From 2021 to 2024 inclusive, 286 probands with suspected genetic diseases were referred from the regional medical genetics counseling service for whole-exome sequencing. The gender, age, and ethnic compositions of the subjects are presented in Table 1. The majority of patients (86.7 %) belong to the child-adolescent age group, the mean age of the cohort being 10.9 years. The most represented ethnic groups are Russians (65.4 %) and Tatars (11.5 %).

Regarding the structure of symptom categories (Table 2), neurological disorders occupy the leading position in Yugra (41.3 %). They are followed by dysmorphic syndromes (12.2 %) and metabolic disorders (11.2 %). It should be noted that psychiatric disorders are the only category not represented in the study group.

Clinical and molecular characteristics of identified genetic variants related to the phenotype

A total of 420 genetic variants (SNVs, InDels) were identified in patients. These variants were associated either with the observed phenotype or with the asymptomatic heterozygous carrier status (AC) of various monogenic diseases. Although, as shown in Figure 1, most of the identified variants are not of confirmed clinical significance, their potential contribution to disorder manifestation cannot be ignored.

The clinically significant variants detected (194 out of 420) (Fig. 2) belong to the following classes: missense, 43.3 %; stop-gained, 21.6 %; frameshift deletions, 20.1 %; splice donor/acceptor variants, 8.2 %; frameshift insertions, 5.2 %; inframe deletions, 1.5 %. It should be noted that the class of repeat length variation is not represented in our study, as the bioinformatics algorithm used has not undergone strict validation for the analysis of this variation type.

The inheritance patterns of identified clinically significant variants are presented in Figure 3. Gene variants with the autosomal recessive inheritance pattern have a significant

Table 1. Demographic characteristics of the studied cohort (*N* = 286)

Characteristic	Patients, no. (%)	Characteristic	Patients, no. (%)
Gender		Ethnic structure	
Male	159 (55.6)	Russians	187 (65.4)
Female	127 (44.4)	Tatars	33 (11.5)
Onset age		Azerbaijanis	11 (3.8)
Neonatal	38 (13.3)	Tajiks	10 (3.5)
Infancy	85 (29.7)	Ukrainians	8 (2.8)
Childhood	128 (44.8)	Chechens	8 (2.8)
Adolescent	18 (6.3)	Dagestanis	7 (2.4)
Adult	17 (5.9)	Uzbeks	6 (2.1)
Median onset age [Q1; Q3]	2 [0.5; 4.5]	Kumyks	3 (1.1)
Age at sampling		Armenians	3 (1.1)
Neonatal	1 (0.4)	Belarusians	2 (0.7)
Infancy	4 (1.4)	Germans	2 (0.7)
Childhood	180 (62.9)	Kazakhs	2 (0.7)
Adolescent	64 (22.4)	Moldovans	2 (0.7)
Adult	37 (12.9)	Bashkirs	1 (0.4)
Median age at sample accession [Q1; Q3]	8.4 [3.8; 14.9]	Lezgians	1 (0.4)

Table 2. Categories of clinical symptoms (*N* = 286)

Symptom category	Patients, no. (%)	Symptom category	Patients, no. (%)
Neurological disorders	118 (4.3)	Autoimmune or rheumatological disorders	8 (2.8)
Dysmorphic and congenital abnormality syndromes	35 (12.2)	Cardiovascular disorders	7 (2.4)
Metabolic disorders	32 (11.2)	Gastrointestinal and hepatic disorders	7 (2.4)
Hearing and ear disorders	16 (5.6)	Endocrine disorders	4 (1.4)
Skeletal disorders	15 (5.2)	Ophthalmological disorders	3 (1.1)
Renal and urinary tract disorders	12 (4.2)	Respiratory disorders	2 (0.7)
Neurodevelopmental disorders	9 (3.1)	Growth disorders	1 (0.4)
Hematological and immunological disorders	8 (2.8)	Tumor syndromes	1 (0.4)
Dermatological disorders	8 (2.8)		

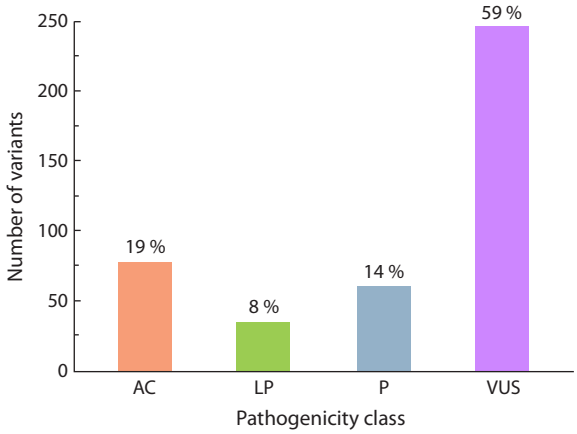


Fig. 1. Pathogenicity classes of the identified variants.

AC – asymptomatic carrier; LP – likely pathogenic; P – pathogenic; VUS – variant of unknown significance.

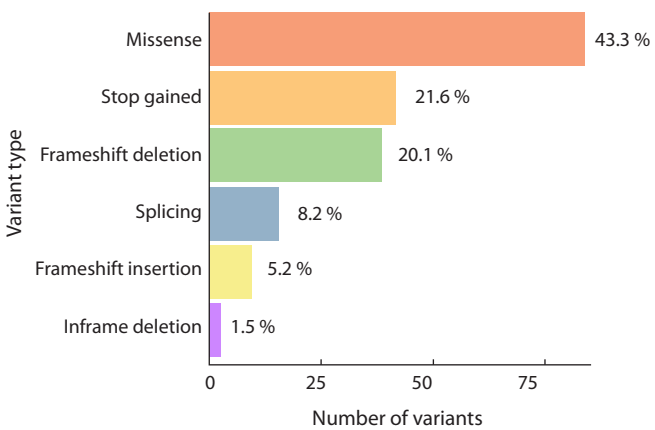


Fig. 2. Molecular classes of clinically significant variants.

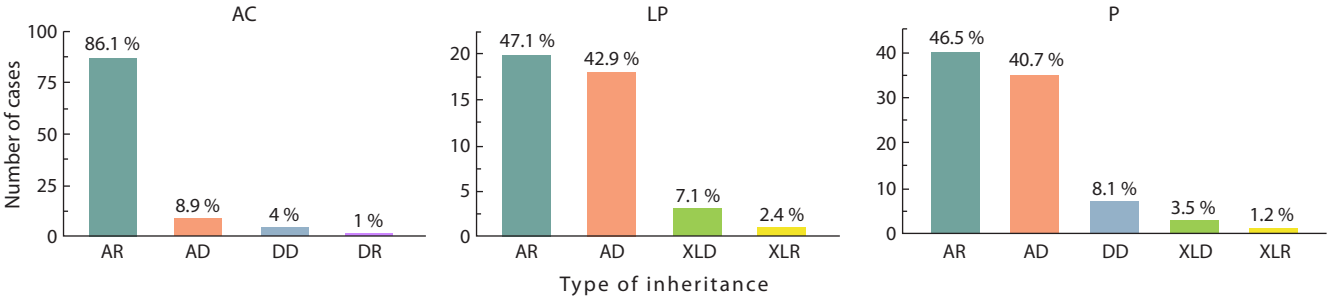


Fig. 3. Distribution of pathogenicity classes by inheritance types.

AR – autosomal recessive; AD – autosomal dominant; DD – digenic dominant; DR – digenic recessive; XLD – X-linked dominant; XLR – X-linked recessive.

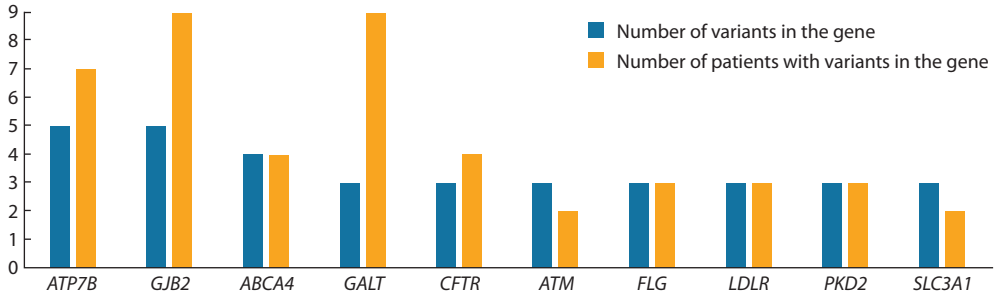


Fig. 4. Human genes with the largest numbers of variants in the studied cohort.

Table 3. The most common clinically significant variants in the studied cohort

cDNA variant	Patho-genicity	Inheritance	Position (hg19)	gnomAD AF, %	Mutation type	No. of patients	Relationship with diseases
ATP7B c.3207C>A	P, AC	AR	chr13:52518281G>T	0.1524	Missense	6	Wilson disease
GJB2 c.109G>A	P, AC	AR, DD	chr13:20763612C>T	0.1392	Missense	4	GJB2-related deafness
GJB2 c.35delG	P, AC	AR, DD	chr13:20763685AC>A	0.9578	Frameshift deletion	4	GJB2-related deafness
PAH c.1222C>T	P, AC	AR	chr12:103234271G>A	0.1719	Missense	4	Phenylketonuria
NEB c.23989C>T	AC	AR	chr2:152357937G>A	0.05838	Stop gained	3	Nemaline myopathy 2
C2 c.841_ 868del	AC	AR	chr6:31902065ATGGTGGACAG GGTCAGGAATCAGGAGTC>A	0.7159	Frameshift deletion	2	C2-complement deficiency
CFTR c.274G>A	AC	AR	chr7:117170953G>A	NA	Missense	2	Cystic fibrosis
KIAA0586 c.392del	AC	AR	chr14:58899156AG>A	0.4773	Frameshift deletion	2	Joubert syndrome
OTOA c.2359G>T	LP	AR	chr16:21747639G>T	0.03104	Stop gained	2	Autosomal recessive deafness 22

weight in the cohort studied (47.6 % for LP and 46.5 % for P). Autosomal dominant inheritance is observed for a substantial number of P/LP variants (40.7 and 42.9 %, respectively). X-linked inheritance has a relatively low prevalence in the cohort, accounting for no more than 10 % of all likely pathogenic variants and no more than 5 % of all pathogenic variants.

Among the genes with the greatest number of detected variants, as shown in Figure 4, the most prominent are *ATP7B*, *GJB2*, *ABCA4*, and *GALT*, which are associated with Wilson's disease, GJB2-related deafness, Stargardt disease, and galactosemia, respectively.

The most frequent clinically significant variant, as presented in Table 3, is c.3207C>A in the *ATP7B* gene, associated with Wilson's disease. However, the majority of the most frequent variants were detected in the AC (asymptomatic carrier) state. All most frequent variants detected are described in the literature as pathogenic, including the *CFTR* c.274G>A (p.Glu92Lys) variant (Chuvash mutation), which is absent from the gnomAD database.

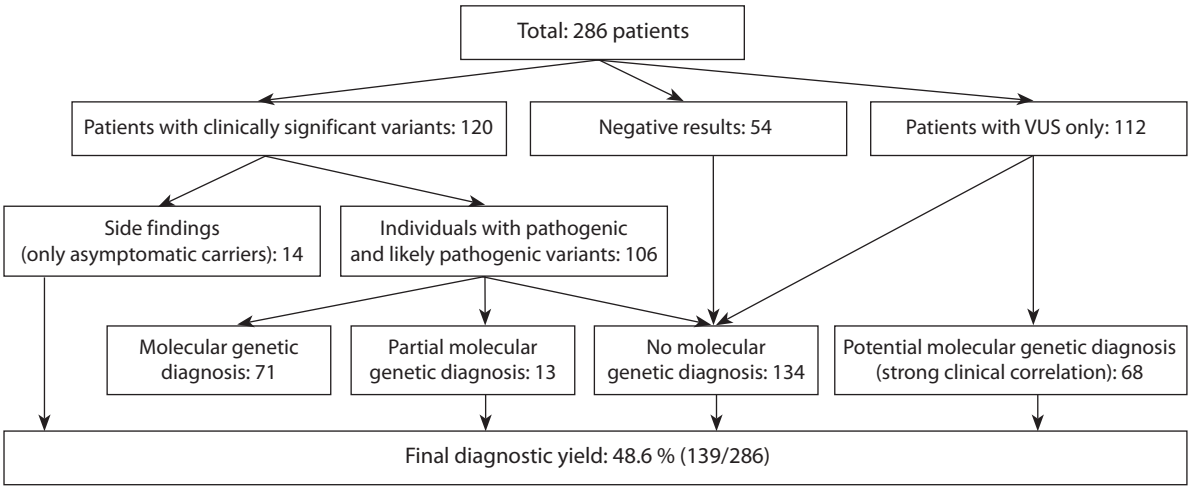


Fig. 5. WES testing summary.

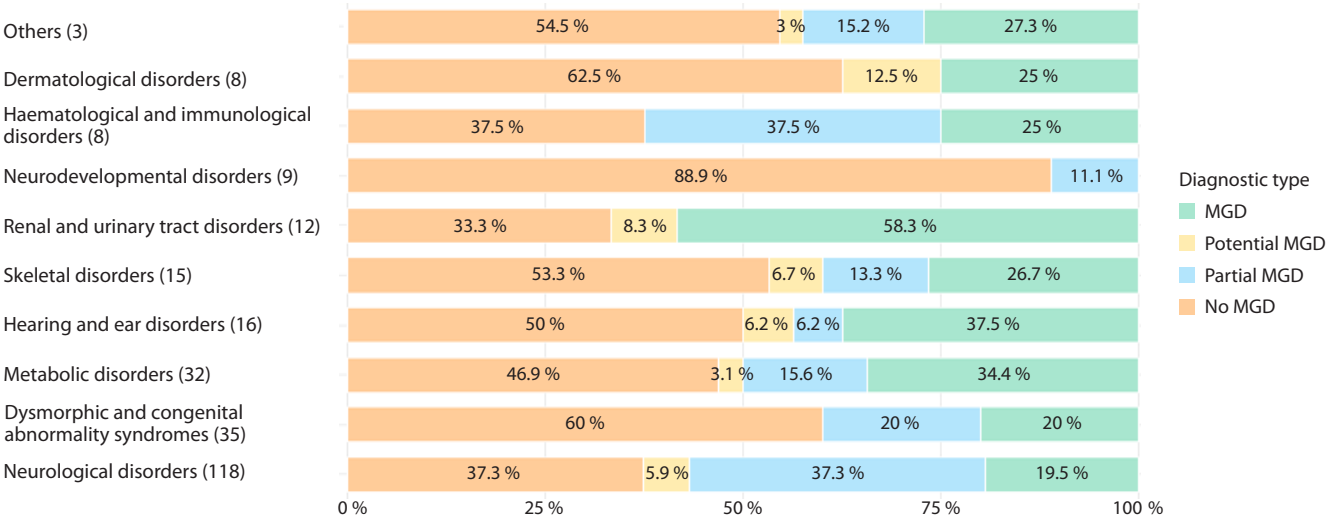


Fig. 6. Types of molecular genetic diagnosis (MGD) in the most common disorder groups.

Clinical effectiveness of WES in the examined cohort

As shown in Figure 5, WES results produced no findings in 54 out of 286 examined patients. Variants of P/LP and AC classes were found in 120 patients, with 14 of them having only AC as a secondary finding. Among 106 patients with phenotype-relevant variants, 71 received a molecular genetic diagnosis, and 13 had only a partial diagnosis. Among the 112 studied samples with only VUS, 68 patients had variants that allowed for a potential molecular genetic diagnosis. Thus, the diagnostic effectiveness of WES in the examined cohort was 48.6 % (139/286).

The distribution of molecular genetic diagnosis types varies significantly across different symptom categories (Fig. 6). The highest percentages of patients with confirmed molecular genetic diagnosis belong to the following categories: *Renal and urinary tract disorders*, 58.3 % (7/12 patients); *Hearing and ear disorders*, 37.5 % (6/16 patients), and *Metabolic disorders*, 34.4 % (11/32 patients). The lowest diagnostic effectiveness among the most common categories is observed in the *Neurodevelopmental disorders* group, where 88.9 % of patients (8/9) received no molecular genetic diagnosis.

The most ambiguous molecular diagnoses are observed in *Haematological and immunological disorders* – 37.5 % with potential MD (3/8 patients) and *Neurological disorders* – 37.3 % with potential MD (44/118 patients). The presence of a large number of variants of uncertain clinical significance (VUS) explains the ambiguity in these groups.

Discussion

We performed whole-exome sequencing for 286 probands with suspected hereditary diseases to clarify the molecular genetic diagnosis. The majority of patients (86.7 %) belonged to the child-adolescent age group, which is associated with the early manifestation of hereditary diseases. Male subjects were more numerous in the group studied (55.6 %), owing to the presence of X-linked recessive disorders.

The diagnostic effectiveness of the method was 24.8 % for cases with complete molecular genetic diagnosis and increased up to 48.6 % when taking into account variants of uncertain significance that have a close relationship with the observed phenotype. These variants may be reclassified using updated information in the foreseeable future, as with similar findings from studies in other countries (Han et al., 2025).

As long as WES testing was conducted only for probands, the experience of other researchers brought us to the assumption that diagnostic effectiveness could be increased significantly through the implementation of duo- and trio-WES (Lai et al., 2024), mainly due to more reliable identification of compound heterozygous variants and *de novo* variants (Tan et al., 2019). This is particularly relevant for the studied population, where the proportion of autosomal recessive pathogenic or likely pathogenic variants is close to 50 %

Additionally, diagnostic effectiveness could be improved through re-analysis of exome data due to the constant emergence of new data about the relationships between variants and phenotypes (Arteche-López et al., 2022). Finally, since approximately 13 % of genetic variability is associated with copy number variations (CNVs) (Stankiewicz, Lupski, 2010), we conceive that it would be desirable for patients with negative WES results to undergo whole-genome sequencing (WGS) or low-coverage genome sequencing in combination with re-analysis of WES (Moey et al., 2025).

We found that the most common symptom categories of hereditary diseases in Yugra were *Neurological disorders* (118 out of 286 cases), *Dysmorphic and congenital abnormality syndromes* (35/286); and *Metabolic disorders* (32/286).

Our study shows significant differences in WES diagnostic effectiveness among different symptom categories: from zero yield for percentage of confirmed molecular genetic diagnosis cases in the *Neurodevelopmental disorders* group to 58.3 % in the *Renal and urinary tract disorders* group. Thus, the high occurrence of neurological disorders in our cohort presumes that just nervous system disorders pose the greatest challenge to productive WES analysis. Such differences may be related to yet uncharacterized molecular causes of these diseases (Adams, Eng, 2018), and generally they are consistent with results obtained from the analysis of 3,040 WES samples, where a wide range from 4 to 55 % of positive reports (Retterer et al., 2016) was also observed in different disease groups.

It is also important to consider the relevance of applying special molecular genetic methods to specific inherited diseases, such as search for duplications on chromosome 17 in cases of hereditary motor-sensory neuropathies (Shchagina et al., 2020), which was not conducted within the scope of our study. Other studies note the effectiveness of duo- and trio-WES for patients with intellectual disability, developmental delays, and epilepsy due to the high number (up to 80 %) of *de novo* variants in these cases (Lai et al., 2024).

It is particularly important to emphasize the significance of providing sufficient data on clinical picture by referring physicians (mostly by geneticists), as incomplete understanding of the symptoms by the interpreter may lead to false elimination of many genes from the search criteria and produce false-negative results. Additionally, for multiethnic populations determining the patient's ethnicity (at least based on a questionnaire) is an important aspect, which was considered in our study, involving 16 ethnic groups.

Furthermore, the importance of secondary findings classified in our study as the asymptomatic carrier state (about 20 % of all identified variants) should not be underestimated, as they provide insight into the genetic burden of the population and should be appropriately considered during genetic counseling of families, particularly for planning future pregnancies.

It can be noted that 8.9 % of asymptomatic carrier cases among the examined patients are referred to genes with an autosomal dominant inheritance pattern. However, for these cases no significant effect on the phenotype was observed, which may be due to the different inheritance patterns of the variant depending on its belonging to a particular protein domain, as in the case of *GJB2* (Xiang et al., 2023). Another consideration is the risk of mutation manifestation at an older age, as in the case of mutations in the *BRCA1* and *BRCA2* genes (Azzollini et al., 2016).

Conclusions

In our study of 286 residents of the KHMAO-Yugra region, we accurately established molecular genetic diagnoses in 24.8 % of patients by using the whole-exome sequencing approach and identified the possible genetic nature of inherited diseases in 48.6 % of cases within the ethnically heterogeneous group. We characterized 420 gene variants, excluding benign and likely benign ones. Thus, whole exome sequencing should be considered a sufficiently effective first-tier diagnostic method for unveiling the genetic causes of a broad range of presumably hereditary diseases. However, some specific cases may demand duo- or trio-WES, whole-genome sequencing, or additional specific methods well proven in revealing the molecular background of certain classes of molecular disorders.

References

Adams D.R., Eng C.M. Next-generation sequencing to diagnose suspected genetic disorders. *N Engl J Med*. 2018;379(14):1353-1362. doi 10.1056/NEJMra1711801

Adzhubei I.A., Schmidt S., Peshkin L., Ramensky V.E., Gerasimova A., Bork P., Kondrashov A.S., Sunyaev S.R. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4): 248-249. doi 10.1038/nmeth0410-248

Arteche-López A., Ávila-Fernández A., Riveiro Álvarez R., Almoguera B., Bustamante Aragonés A., Martin-Merida I., López Martínez M.A., … Blanco-Kelly F., Tahsin Swafiri S., Lorda Sánchez I., Trujillo Tiebas M.J., Ayuso C. Five years' experience of the clinical exome sequencing in a Spanish single center. *Sci Rep*. 2022;12(1): 19209. doi 10.1038/s41598-022-23786-6

Azzollini J., Scuvera G., Bruno E., Pasanisi P., Zaffaroni D., Calvello M., Pasini B., Ripamonti C.B., Colombo M., Pensotti V., Radice P., Peissel B., Manoukian S. Mutation detection rates associated with specific selection criteria for BRCA1/2 testing in 1854 high-risk families: a monocentric Italian study. *Eur J Intern Med*. 2016;32: 65-71. doi 10.1016/j.ejim.2016.03.010

Barbitoff Y.A., Khmelkova D.N., Pomerantseva E.A., Slepchenkov A.V., Zubashenko N.A., Mironova I.V., Kaimonov V.S., … Aseev M.V., Shcherbak S.G., Glotov O.S., Isaev A.A., Predeus A.V. Expanding the Russian allele frequency reference via cross-laboratory data integration: insights from 7452 exome samples. *Natl Sci Rev*. 2024; 11(10):nwae326. doi 10.1093/nsr/nwae326

Cingolani P., Platts A., Wang le L., Coon M., Nguyen T., Wang L., Land S.J., Lu X., Ruden D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; *iso*-2; *iso*-3. *Fly* (*Austin*). 2012;6(2):80-92. doi 10.4161/fly.19695

Davydov E.V., Goode D.L., Sirota M., Cooper G.M., Sidow A., Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol.* 2010; 6(12):e1001025. doi 10.1371/journal.pcbi.1001025

Glotov O.S., Donnikov M.Yu., Glotov A.S., Popova E.A., Suchko P.A., Kolbasin L.N., Belotserkovtseva L.D., Kovalenko L.V. Genetic variants identified in patients with rare and multifactorial diseases in the Khanty-Mansi Autonomous Okrug – Yugra based on exome sequencing: registered database No. 2025622641. 2025. Available: https://new.fips.ru/publication-web/publications/document?type=doc&tab=PrEVM&id=60EDD49E-F82F-481E-BFB7-474B01087E34, accessed on 28.09.2025 (in Russian)

Han H., Seo G.H., Hyun S.I., Kwon K., Ryu S.W., Khang R., Lee E., … Yang S., Lee S., Jang S., Lee J., Lee H. Exome sequencing of 18,994 ethnically diverse patients with suspected rare Mendelian disorders. *NPJ Genom Med*. 2025;10(1):6. doi 10.1038/s41525-024-00455-3

Ioannidis N.M., Rothstein J.H., Pejaver V., Middha S., McDonnell S.K., Baheti S., Musolf A., … Bailey-Wilson J.E., Radivojac P., Thibodeau S.N., Whittemore A.S., Sieh W. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877-885. doi 10.1016/j.ajhg.2016.08.016

Kumar P., Henikoff S., Ng P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073-1081. doi 10.1038/nprot.2009.86

Lai G., Gu Q., Lai Z., Chen H., Chen J., Huang J. The application of whole-exome sequencing in the early diagnosis of rare genetic diseases in children: a study from Southeastern China. *Front Pediatr*. 2024;12:1448895. doi 10.3389/fped.2024.1448895

Li H. Exploring single-sample SNP and INDEL calling with whole-genome *de novo* assembly. *Bioinformatics*. 2011;28(14):1838-1844. doi 10.1093/bioinformatics/bts280

Moey L.H., Seo G.H., Cheah B.E., Keng W.T., Lee H., Ch'ng G.S. A first large study of whole-exome sequencing (WES) in 489 patients with suspected rare genetic disorders at a tertiary centre in Malaysia. *Rare*. 2025;3:100102. doi 10.1016/j.rare.2025.100102

Okuneva E.G., Kozina A.A., Baryshnikova N.V., Krasnenko A.Yu., Klimchuk O.I., Stetsenko I.F., Plotnikov N.A., Surkova E.I., Ilinsky V.V. The utility of exome sequencing in diagnosis of hereditary diseases. *Med Genet*. 2020;19(12):18-24. doi 10.25557/2073-7998.2020.12.18-24 (in Russian)

Pashaei E., Yilmaz A., Ozen M., Aydin N. Prediction of splice site using AdaBoost with a new sequence encoding approach. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary. IEEE, 2016;003853-003858. doi 10.1109/SMC.2016.7844835

Petersen B.S., Fredrich B., Hoeppner M.P., Ellinghaus D., Franke A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet*. 2017;18(1):14. doi 10.1186/s12863-017-0479-5

Quang D., Chen Y., Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*. 2015; 31(5):761-763. doi 10.1093/bioinformatics/btu703

Retterer K., Juusola J., Cho M.T., Vitazka P., Millan F., Gibellini F., Vertino-Bell A., … Richard G., Brandt T., Haverfield E., Chung W.K., Bale S. Clinical application of whole-exome sequencing across clinical indications. *Genet Med*. 2016;18(7):696-704. doi 10.1038/gim.2015.148

Richards S., Aziz N., Bale S., Bick D., Das S., Gastier-Foster J., Grody W.W., Hegde M., Lyon E., Spector E., Voelkerding K., Rehm H.L.; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi 10.1038/gim.2015.30

Ross J.P., Dion P.A., Rouleau G.A. Exome sequencing in genetic disease: recent advances and considerations. *F1000Res*. 2020;9(F1000 Faculty Rev):336. doi 10.12688/f1000research.19444.1

Ryzhkova O.P., Kardymon O.L., Prohorchuk E.B., Konovalov F.A., Maslennikov A.B., Stepanov V.A., Afanas'ev A.A., … Kostareva A.A., Pavlov A.E., Golubenko M.V., Polyakov A.V., Kucev S.I. Guidelines for the interpretation of massive parallel sequencing variants (update 2018, v2). *Med Genet.* 2019;18(2):3-24. doi 10.25557/2073-7998.2019.02.3-23 (in Russian)

Shchagina O.A., Ryzhkova O.P., Chuhrova A.L., Milovidova T.B., Gundorova P., Mironovich O.L., Orlova A.A., Orlova M.D., Polyakov A.V. Diagnostic utility of exome sequencing for inherited peripheral neuropathies. *Neuromuscular Disord*. 2020;10(4):12-26. doi 10.17650/2222-8721-2020-10-4-12-26 (in Russian)

Stankiewicz P., Lupski J.R. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010;61:437-455. doi 10.1146/annurev-med-100708-204735

Stranneheim H., Lagerstedt-Robinson K., Magnusson M., Kvarnung M., Nilsson D., Lesko N., Engvall M., … Soller M.J., Nordgren A., Wirta V., Lindstrand A., Wedell A. Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. *Genome Med*. 2021;13(1):40. doi 10.1186/s13073-021-00855-5

Strauch Y., Lord J., Niranjan M., Baralle D. CI-SpliceAI-improving machine learning predictions of disease causing splicing variants using curated alternative splice sites. *PLoS One.* 2022;17(6):e0269159. doi 10.1371/journal.pone.0269159

Sundaram L., Gao H., Padigepati S.R., McRae J.F., Li Y., Kosmicki J.A., Fritzilas N., Hakenberg J., Dutta A., Shon J., Xu J., Batzoglou S., Li X., Farh K.K. Predicting the clinical impact of human mutation with deep neural networks. *Nat Genet*. 2018;50(8):1161-1170. doi 10.1038/s41588-018-0167-z

Tan T.Y., Lunke S., Chong B., Phelan D., Fanjul-Fernandez M., Marum J.E., Kumar V.S., … Martyn M., Goranitis I., Thorne N., Gaff C.L., White S.M. A head-to-head evaluation of the diagnostic efficacy and costs of trio versus singleton exome sequencing analysis. *Eur J Hum Genet.* 2019;27(12):1791-1799. doi 10.1038/s41431-019-0471-9

Van der Auwera G.A., O'Connor B.D. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media, 2020

Wang Q., Shashikant C.S., Jensen M., Altman N.S., Girirajan S. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Sci Rep*. 2017;7(1):885. doi 10.1038/s41598-017-01005-x

Xiang J., Sun X., Song N., Ramaswamy S., Abou Tayoun A.N., Peng Z. Comprehensive interpretation of single-nucleotide substitutions in *GJB2* reveals the genetic and phenotypic landscape of *GJB2*-related hearing loss. *Hum Genet*. 2023;142(1):33-43. doi 10.1007/s00439-022-02479-0

Zinchenko R.A., Kutsev S.I., Aleksandrova O.Yu., Ginter E.K. Main methodological approaches to the identification and diagnosis of monogenic hereditary diseases and problems in the organization of medical care and unified preventive programs. *Probl Social Hyg Public Health Hist Med.* 2019;27(5):865-877. doi 10.32687/0869-866X-2019-27-5-865-877 (in Russian)

Conflict of interest. The authors declare no conflict of interest.
Received September 29, 2025. Revised December 24, 2025. Accepted February 9, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):490-501

doi 10.18699/vjgb-26-53

Rare variants in cholesterol transporter genes in patients with lipid metabolism disorders

A.D. Izyumchenko [1, 2] ✉, M.N. Grunina [1, 2], K.V. Dracheva [1], O.A. Chumakova[1], K.O. Tanayants [2], K.V. Legostaeva [1], A.N. Kulikov [1], O.A. Berkovich [1], E.I. Baranova [1], S.N. Pchelina [1, 2], V.V. Miroshnikova [1, 2]

[1] Pavlov First St. Petersburg State Medical University, St. Petersburg, Russia
[2] Petersburg Nuclear Physics Institute named by B.P. Konstantinov of National Research Centre "Kurchatov Institute", Gatchina, Leningrad region, Russia
✉ artemiz98@yandex.ru

Abstract. Cardiovascular diseases are the leading cause of death both in Russia and in the world. One of the factors predisposing to the development of cardiovascular diseases is lipid metabolism disorders (dyslipidemias), which contribute to the progression of atherosclerosis. Currently, there are known genes associated with the development of monogenic forms of lipid metabolism disorders characterized by marked changes in lipid levels. However, identifying individuals with an increased genetic risk of dyslipidemia remains an unsolved problem, due to the polygenic nature of most cases. The aim of this work was to study the spectrum of rare variants in the cholesterol transporter genes *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8* and *NPC1L1* that occur in patients with lipid metabolism disorders in the population of the Northwestern region of Russia. The search for rare variants (gnomAD frequency less than 1 %) in the *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8* and *NPC1L1* genes was performed using targeted sequencing data for 169 patients with lipid metabolism disorders. 14 variants were identified in the *ABCA1* gene (17 patients); 4 variants, in the *ABCG1* gene (5 patients); 11 variants, in the *ABCG5* gene (18 patients); and 7 variants, in the *ABCG8* gene (11 patients). The frequency of some of them, according to the RUSeq database, is higher than in the global population. 19 patients (11 %) were carriers of the p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) haplotype in the *NPC1L1* gene, which may be specific to the Russian population, meaning that these variants are not rare, but polymorphic, and occur more frequently in patients with impaired lipid metabolism. Influence of the p.(Val177Ile) variant of the *NPC1L1* gene on the development of atherosclerosis was assessed using additional sample sets (a group of patients with atherosclerosis, a control group), but no significant differences in genotype frequencies were revealed. Thus, at present, there are insufficient data to support the role of the p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) haplotype of the *NPC1L1* gene in the development of dyslipidemia and atherosclerosis. The study draws attention to the population specificity of a number of variants in cholesterol transporter genes, in particular in the *NPC1L1* gene, for the Northwestern region of Russia. The data can be further used for design and calculation of genetic risk scores for dyslipidemia.

Key words: familial hypercholesterolemia; dyslipidemia; targeted sequencing; genetic risk scale; reverse cholesterol transport

For citation: Izyumchenko A.D., Grunina M.N., Dracheva K.V., Chumakova O.A., Tanayants K.O., Legostaeva K.V., Kulikov A.N., Berkovich O.A., Baranova E.I., Pchelina S.N., Miroshnikova V.V. Rare variants in cholesterol transporter genes in patients with lipid metabolism disorders. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):490-501. doi 10.18699/vjgb-26-53

Funding. This work was supported by the Russian Science Foundation (RSF) grant No. 25-25-00351: "The impact of pathogenic variants in the LDLR and APOB genes on the transcriptomic profile of macrophages in familial hypercholesterolemia".

Acknowledgements. The authors would like to express their gratitude to all doctors (cardiologists and pediatric specialists) who referred patient samples for genetic diagnosis of lipid metabolism disorders to the Department of Molecular Genetic and Nanobiological Technologies of Pavlov First Saint Petersburg State Medical University; this contribution is important for accumulation of the data.

Редкие варианты в генах транспортеров холестерина у пациентов с нарушениями липидного обмена

А.Д. Изюмченко [1, 2] ✉, М.Н. Грунина [1, 2], К.В. Драчева [1], О.А. Чумакова[1], К.О. Танаянц [2], К.В. Легостаева [1], А.Н. Куликов [1], О.А. Беркович [1], Е.И. Баранова [1], С.Н. Пчелина [1, 2], В.В. Мирошникова [1, 2]

[1] Первый Санкт-Петербургский государственный медицинский университет им. академика И.П. Павлова, Санкт-Петербург, Россия
[2] Петербургский институт ядерной физики им. Б.П. Константинова Национального исследовательского центра «Курчатовский институт», Гатчина, Ленинградская область, Россия
✉ artemiz98@yandex.ru

Аннотация. Сердечно-сосудистые заболевания являются основной причиной смерти в России и мире. Один из предрасполагающих к развитию сердечно-сосудистых заболеваний факторов – нарушения липидного обмена (дислипидемии), способствующие прогрессированию атеросклероза. На настоящий момент известны гены, ассоциированные с развитием моногенных форм нарушений липидного обмена, характеризующихся выраженным изменением уровня липидов. Однако выявление лиц с повышенным генетическим риском развития дислипидемии остается нерешенной задачей, что связано с полигенной природой большинства случаев. Целью настоящей работы было изучить спектр редких вариантов в генах транспортеров холестерина *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8* и *NPC1L1*, которые встречаются у пациентов с нарушениями липидного обмена в популяции Северо-Западного региона России. Проведен поиск редких вариантов (частота gnomAD менее 1 %) в генах *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8* и *NPC1L1* с использованием данных таргетного секвенирования для 169 пациентов с нарушениями липидного обмена. Выявлено 14 вариантов в гене *ABCA1* (17 пациентов), 4 варианта в гене *ABCG1* (5 пациентов), 11 вариантов в гене *ABCG5* (18 пациентов) и 7 вариантов в гене *ABCG8* (11 пациентов). Частота некоторых из них, согласно базе данных RUSeq, была выше, чем в общемировой популяции. 19 пациентов (11 %) были носителями гаплотипа p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) гена *NPC1L1*, который предположительно может быть специфичен для российской популяции, т.е. эти варианты являются не редкими, а полиморфными и встречаются чаще у пациентов с нарушениями липидного обмена. Для варианта p.(Val177Ile) гена *NPC1L1* был проведен анализ его вклада в развитие атеросклероза с использованием дополнительных выборок (группа пациентов с атеросклерозом, контрольная группа), который не выявил достоверных различий в частотах генотипов. Таким образом, в настоящее время данных в пользу влияния гаплотипа p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) гена *NPC1L1* на развитие дислипидемии и атеросклероза недостаточно. Проведенное исследование заставляет обратить внимание на популяционную специфичность ряда вариантов в генах транспортеров холестерина, в частности в гене *NPC1L1*, для Северо-Западного региона России. Полученные данные в дальнейшем могут быть учтены в разработке шкал генетического риска развития дислипидемий.

Ключевые слова: семейная гиперхолестеринемия; дислипидемия; таргетное секвенирование; шкала генетического риска; обратный транспорт холестерина

Introduction

Cardiovascular diseases (CVDs) are the leading cause of death in Russia and worldwide (Danilova et al., 2021; Heron, 2021). The multifactorial nature of CVDs underscores the importance of studying new markers of the pathological process initiation, including genetic ones. The contribution of the genetic component to the development of heart and vascular diseases is estimated at 40–50 % (McPherson, Tybjaerg-Hansen, 2016; Roberts et al., 2021). Hypercholesterolemia, which often has a hereditary nature, promotes the formation of atherosclerotic plaques in the coronary vessels, which in turn leads to the developvent of coronary artery disease (CAD) and acute myocardial infarction (AMI) (Prasad, Mishra, 2022). Familial hypercholesterolemia (FH) occupies a special place among hereditary lipid metabolism disorders, as it significantly increases the risk of developing CVDs (Tokgozoglu, Kayikcioglu, 2021). We and other authors have shown that in most cases, FH is caused by pathogenic variants in the *LDLR* and *APOB* genes. However, depending on the severity of hypercholesterolemia, the genetic cause of the disease cannot be identified in 40–60 % of cases (Shakhtshneider et al., 2021; Miroshnikova et al., 2023a, 2025). Thus, discovering rare genetic variants, which, in combination, can have a cumulative effect on the development of polygenic cases of hypercholesterolemia and associated CVDs, remains relevant.

In a study by A.N. Meshkov and co-authors, an increased risk of CAD in carriers of rare and low-frequency variants in genes associated with lipid metabolism disorders was demonstrated (Meshkov et al., 2022). Rare functional genetic variants associated with cholesterol metabolism were found in 60 % of patients with AMI (Pan-Lizcano et al., 2022). It has been shown that 25 % of loci associated with AMI belong to lipid metabolism genes (Musunuru, Kathiresan, 2016). Rare genetic variants have a population prevalence of <1 % and may not be statistically associated with diseases of interest in large sample sets. However, even a small increase in allele frequency (1–5 %) in patients may indicate its influence on complex diseases and traits (Cross et al., 2022; Li et al., 2024). Moreover, such genetic variants may be population-specific and should be taken into account during development of genetic risk scores for a specific population.

Atherogenic imbalance of blood plasma lipid fractions manifests as an increase in the concentration of total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), and a decrease in high-density lipoprotein cholesterol (HDL-C). Transmembrane cholesterol transporters play an important role in regulation of cellular cholesterol levels, in the formation of lipoprotein particles, intestinal cholesterol absorption, and cholesterol excretion from the body (Yu, Tang, 2022). ATP-binding cassette (ABC) transporters – ABCA1 and ABCG1 – transport cholesterol to anti-atherogenic high-density lipoproteins (HDLs) (Yu, Tang, 2022). Mutations in the *ABCA1* gene lead to the development of an autosomal recessive disorder – Tangier disease – characterized by an almost complete absence of HDLs in blood plasma and premature development of atherosclerosis (Koseki et al., 2021). Transporters ABCG5 and ABCG8 regulate the absorption of cholesterol and plant sterols in the intestine. Mutations in these genes lead to the development of sitosterolemia, which has clinical manifestations similar to FH (Tada et al., 2022). The NPC1L1 transporter, known as Niemann-Pick C1-like protein 1, is also involved in intestinal cholesterol absorption and is a target for the lipid-lowering drug ezetimibe (Zhang et al., 2022). It can be hypothesized that combinations of rare variants in cholesterol transporter genes may contribute to the development of dyslipidemias and CVDs (Ghaleb et al., 2022; Meshkov et al., 2022).

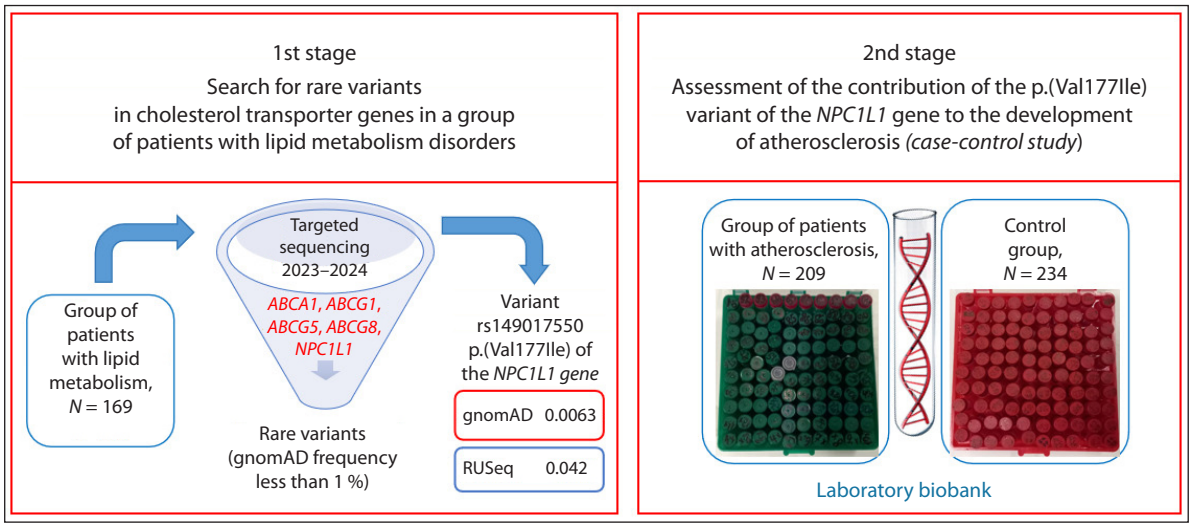


Fig. 1. Research design.

The aim of this study was to investigate the spectrum of rare variants in the cholesterol transporter genes *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8*, and *NPC1L1* that occur in patients with lipid metabolism disorders in the population of the Northwestern region of Russia.

Materials and methods

The study was approved by the Local Ethics Committee of the Pavlov First Saint Petersburg State Medical University of the Ministry of Health of the Russian Federation (Protocol No. 275 dated September 4, 2023). All participants provided written informed consent.

The overall study design is presented in Figure 1.

Sequencing data analysis

The search for rare variants in the cholesterol transporter genes *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8* and *NPC1L1* was carried out using NGS sequencing data obtained for 169 patients with lipid metabolism disorders: 13 adults (aged 24 to 85 years, mean age 52.3 ± 14.6; 65 men and 69 women); 35 children (aged 3 to 17 years, mean age 9.7 ± 3.8; 18 boys and 17 girls). These patients were referred for genetic diagnosis of hereditary dyslipidemias to the Department of Molecular Genetic and Nanobiological Technologies from various clinical diagnostic centers in Saint Petersburg (including Pavlov First Saint Petersburg State Medical University) during 2023–2024.

Criteria for genetic testing prescription were as follows: 1) possible/probable/definite FH in patients over 18 years of age in accordance with Dutch diagnostic criteria (Kukharchuk et al., 2020); 2) probable/definite FH in patients under 18 years of age in accordance with Simon Broome criteria (Yezhov et al., 2019); 3) pronounced hypertriglyceridemia (individuals with triglyceride concentrations from 3.3 to 10.5 mmol/L were included).

Libraries were prepared using a set of Prep&Seq reagents (Parseq Lab Co, Russia) and a custom panel "Dyslipidemia and CVD risk", including coding regions of the following genes: *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8*, *ANGPTL3*, *APOA1*, *APOA4*, *APOA5*, *APOB*, *APOC2*, *APOC3*, *APOE*, *CETP*, *CREB3L3*, *GCK*, *CYP27A1*, *CYP7A1*, *GPD1*, *GPIHBP1*, *HNF1A*, *LCAT*, *LDLR*, *LDLRAP1*, *LIPA*, *LIPC*, *LIPG*, *LMF1*, *LPL*, *LRP6*, *MTTP*, *MYLIP*, *NPC1L1*, *PCSK9*, *PNPLA5*, *SAR1B*, *SCARB1*, *SORT1*, *STAP1* and *TTR* (VariFind LM assay IL-v1.1.1, Parseq Lab Co, Russia). Sequencing was performed on a MiSeq instrument (Illumina, Inc., USA). The sequencing data were processed using the automated Seq&Go Software (Parseq Lab Co, Russia). Identified genetic variants were annotated and described according to the guidelines of the Human Genome Variation Society (HGVS) (www.hgvs.org) and presented in tabular format. Next, we selected rare variants (gnomAD frequency less than 1 %) in the cholesterol transporter genes *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8* and *NPC1L1*. The frequencies of the selected variants were compared with the Russian database of genetic information RUSeq (http://ruseq.ru/#/) (Barbitoff et al., 2024). The OMIM (https://omim.org/), gnomAD (https://gnomad.broadinstitute.org/), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), HGMD (https://www.hgmd.cf.ac.uk/ac/index.php), LOVD (https://www.lovd.nl/) databases and literature data were used to assess the clinical relevance of the identified nucleotide sequence variants. The assessment of the clinical significance (pathogenicity) of the identified variants was carried out on the basis of Russian recommendations for the interpretation of data obtained by mass parallel sequencing methods, as well as the recommendations of ACMG2015 (Ryzhkova et al., 2019).

The verification of the p.(Val177Ile), p.(His221Tyr) and p.(Ala271Phe) variants of the *NPC1L1* gene, which were considered to compose a haplotype, was carried out by Sanger sequencing on a Nanophor-05 sequencer (Syntol, Russia) using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). The results were analyzed using the Mutation Surveyor software (Soft Genetics, USA). The primers for sequencing were selected using the online program Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Table 1).

Table 1. The nucleotide sequences of the primers used for Sanger sequencing

Genetic variant	Sequence 5′–3′
2 exon, p.(Val177Ile)	F-TTGGGACTCATTGCAACGTG R-CCCCAATCAGAGCCTCTTCA
2 exon, p.(His221Tyr)	F-CCTTCTTGGGGTCCACCATC R-ATGTGTGGCGTGTATGGCTCT
2 exon, p.(Ala271Phe)	F-CCTTCTTGGGGTCCACCATC R-CTTCCACCTCTTGGAGCCTG

Evaluating the contribution of the p.(Val177Ile) variant of the *NPC1L1* gene to the development of atherosclerosis

Characteristics of the groups. Genotyping of the p.(Val177Ile) variant of the *NPC1L1* gene was performed for a total of 443 patients who were examined or treated at different times at Pavlov First Saint Petersburg State Medical University and whose DNA was available from the biobank of the Department of Molecular Genetic and Nanobiological Technologies: 1) patients with atherosclerosis of various localization (N = 209), 2) control group (N = 234).

The group of patients with early atherosclerosis included 209 patients (146 (70 %) men and 63 (30 %) women; average age 54.6 ± 8.5 years; the average age of the first clinical manifestations was 48.2 ± 6.8 years) with atherosclerosis of the arteries of various localization (coronary, cerebral, lower limb arteries), confirmed by angiographic examination (Table 2). The selection criterion was the presence of hemodynamically significant stenoses in at least one artery of one of the three main arterial basins – the cerebral, coronary, and lower extremities. It should be noted as a limitation that the sample set was formed based on the principle of confirmed atherosclerosis, regardless of whether patients had dyslipidemia, since the duration and effectiveness of statin treatment in most cases was difficult to evaluate.

The control group included 234 individuals (154 (66 %) men and 80 (34 %) women; average age 48.3 ± 5.8 years – the age corresponded to the average age of onset of the disease in patients with atherosclerosis) without lipid metabolism disorders, obesity and cardiovascular diseases. Control group individuals were examined by a cardiologist, and the examinations performed included electrocardiography, bicycle ergometry, and echocardiography.

Identification of the p.(Val177Ile) variant of the *NPC1L1* gene using polymerase chain reaction and restriction analysis. Genotyping of the rs149017550 chr7:44539868C>T c.529G>A p.(Val177Ile) variant of the *NPC1L1* gene was performed by polymerase chain reaction (PCR) and subsequent restriction analysis. The primers for PCR were selected using the online program Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The nucleotide sequence of the primers was (5′–3′): forward – TTGGGACTCATTGCAAC GTG, reverse – CCCCAATCAGAGCCTCTTCA. As a result of amplification with these primers, a PCR product with a size of 352 bp was obtained.

For restriction analysis, the PCR product was incubated with 1 unit of Fok I endonuclease in buffer Y containing: 33 mmol Tris acetate (pH 7.9 at 25 °C), 10 mmol magnesium acetate, 66 mmol potassium acetate, 1 mmol DTT, at +37 °C overnight. The results were visualized using polyacrylamide gel electrophoresis: depending on the genotype, DNA fragments of 352, 224, and 128 bp in length were obtained (for GG, only a fragment of 352 bp; for GA, fragments of 352, 224, and 128 bp; for AA, fragments of 224 and 128 bp).

Table 2. Clinical and biochemical characteristics of the studied groups

Parameters	A group of patients with atherosclerosis (N = 209)	Control group (N = 234)	p
Lipid profile, mmol/L*			
TC	5.20 (4.46 ÷ 6.15)	4.56 (4.02 ÷ 5.14)	0.000
LDL-C	2.76 (2.45 ÷ 3.39)	2.66 (2.19 ÷ 3.23)	0.129
HDL-C	1.06 (0.80 ÷ 1.10)	1.45 (1.26 ÷ 1.66)	0.000
TG	2.00 (1.18 ÷ 2.28)	0.81 (0.59 ÷ 1.02)	0.000
Parameters of atherosclerotic lesions, N (%)			
Coronary atherosclerosis	112 (53.6)	–	–
Cerebral atherosclerosis	43 (20.6)	–	–
Atherosclerosis of the lower extremities arteries	76 (36.4)	–	–
Lesions in 2–3 arterial basins	21 (10)	–	–
Multivessel atherosclerosis (3 or more arteries)	164 (78.5)	–	–

Note. TC – total cholesterol, LDL-C – cholesterol of low-density lipoproteins, HDL-C – cholesterol of high-density lipoproteins, TG – triglycerides.
* The median and interquartile range (IQR) are indicated.

Statistical analysis

The statistical analysis was performed using the SPSS version 17.0 software. The chi-square test was used to compare categorical variables. The correspondence of the data to the normal distribution was checked using the Kolmogorov–Smirnov test. The nonparametric Mann–Whitney test was used to compare quantitative indicators between two independent groups (patient-control). The assessment of the possible influence of the *NPC1L1* genotype on the development of atherosclerosis, adjusted for gender and age, was performed using multifactorial logistic regression analysis.

Results

Analysis of the prevalence of rare variants in the *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8* and *NPC1L1* genes in patients with lipid metabolism disorders

We analyzed targeted sequencing data for 169 patients with hyperlipidemia, specifically searching for rare variants (gnomAD frequency <1 %) in the cholesterol transporter genes – *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8* and *NPC1L1*. The results are presented in Tables 3–7.

14 variants in the *ABCA1* gene were identified in 17 patients, and four variants in the *ABCG1* gene were found in five patients (Tables 3 and 4). All identified variants are classified as benign in public databases. The ABCA1 transporter is an important regulator of HDL biogenesis, mediating the transfer of cholesterol from cells to nascent pre-beta-HDL particles. Therefore, *ABCA1* gene variants affecting protein function may be associated with reduced plasma HDL-C levels, contributing to the development of atherogenic dyslipidemia (Shim et al., 2021). A female patient with a clinical presentation of FH and low HDL-C levels (1.2 mmol/L, below the lower normal limit for women) was a carrier of two rare *ABCA1* variants: rs187652566 p.(Cys887Phe) and rs138422574 p.(Val1674Ile). It is noteworthy that the frequencies of two *ABCA1* variants in the Russian population are higher than the gnomAD frequencies: rs764824326 p.(Glu450Lys) and rs138422574 p.(Val1674Ile) (Table 3).

The ABCG1 transporter provides saturation of mature HDL cholesterol for subsequent transport to the liver. According to the RUSeq database, two variants of the *ABCG1* gene are more common in the Russian population: rs4148108 (c.292+4G>A) and rs143199611 p.(Ala265Thr) (Table 4).

Table 3. *ABCA1* rare variants identified in patients with lipid metabolism disorders

Number of patients	rsID	The genomic coordinate (GRCh38)	Substitution NM_013389.3/ protein	Frequency gnomAD	Frequency RUSeq
1	rs199976989	chr9:104903705T>G	c.–26A>C	0.0001511	–
1	rs1408196930	chr9:104928007A>T	c.–165T>A	0.000006569	–
1	rs1015644827	chr9:104840350T>C	c.983A>G p.(Asn328Ser)	0.000001368	0.0002962
1	rs764824326	chr9:104832735C>T	c.1348G>A p.(Glu450Lys)	0.00001232	0.0002998
1	rs186911476	chr9:104832706T>G	c.1377A>C p.(Thr459=)	0.0002719	0.001574
1	rs142342160	chr9:104832616G>A	c.1467C>T p.(Asn489=)	0.0000314	0.0005928
1	rs145582736	chr9:104825781T>C	c.2444A>G p.(Glu815Gly)	0.0001826	0.0008339
2	rs370414528	chr9:104824558G>A	c.2563C>T p.(Pro855Ser)	0.00001232	0.0001044
1	rs187652566	chr9:104822664C>A	c.2660G>T p.(Cys887Phe)	0.0003932	0.001469
1	rs41277767	chr9:104820060G>A	c.2970C>T p.(Val990=)	0.00003765	0.0002089
1	rs138422574	chr9:104798522C>T	c.5020G>A p.(Val1674Ile)	0.0004845	0.001780
1	rs13306077	chr9:104796059G>A	c.5376C>T (Thr1792=)	0.0024642	0.001195
3	rs142688906	chr9:104791982C>T	c.5774G>A p.(Arg1925Gln)	0.002204	0.009880
1	rs144588452	chr9:104784371C>T	c.6730G>A p.(Val2244Ile)	0.0002198	0.0004175

Note. All variants are likely benign.

Table 4. *ABCG1* rare variants identified in patients with lipid metabolism disorders

Number of patients	rsID	The genomic coordinate (GRCh38)	Substitution NM_013389.3/ protein	Frequency gnomAD	Frequency RUSeq
1	rs141619254	chr21:42225766G>A	c.144G>A p.(Thr46=)	0.001088	0.0002413
2	rs4148108	chr21:42225918G>A	c.292+4G>A	0.001545	0.005481
1	rs143199611	chr21:g.42284612G>A	c.793G>A p.(Ala265Thr)	0.0001834	0.0009681
1	rs138421137	chr21:42290169C>T	c.1350C>T p.(Phe450=)	0.001220	0.002301

Note. All variants are likely benign.

Table 5. *ABCG5* rare variants identified in patients with lipid metabolism disorders

Number of patients	rsID	The genomic coordinate (GRCh38)	Substitution NM_013389.3/ protein	Frequency gnomAD	Frequency RUSeq	Clinical significance according to the ClinVar database
3	rs56204478	chr2:43838600C>G	c.80G>C p.(Gly27Ala)	0.00308	0.004534	Likely benign
1	rs560839317	chr2:43838755G>A	c.–76C>T	0.0006295	–	Uncertain significance
1	rs373819340	chr2:43837861G>T	c.238C>A p.(Gln80Lys)	0.00007	0.0001042	
2	rs145164937	chr2:43832056G>C	c.293C>G p.(Ala98Gly)	0.00237	0.002706	
1	rs1250295912	chr2:43831947C>T	c.402G>A p.(Gln134=)	0.000003578	0.0001044	Likely benign
1	rs1044946422	chr2:43828090C>T	c.527G>A p.(Ser176Asn)	0.000007	–	Uncertain significance
1	rs141828689	chr2:43828024C>T	c.593G>A p.(Arg198Gln)	0.00140	0.004739	
1	rs72796720	chr2:43826460G>A	c.696C>T p.(Val232=)	0.00199	0.002380	Likely benign
1	rs552803459	chr2:43824988C>T	c.805G>A p.(Gly269Arg)	0.000007	–	Uncertain significance
1	rs150716811	chr2:43813266G>A	c.1806C>T p.(Phe602=)	0.0003297	0.0004180	Likely benign
5	rs140374206	chr2:43813208T>C	c.1864A>G p.(Met622Val)	0.00602	0.007414	

11 variants were identified in the *ABCG5* gene in 18 patients, and seven variants in the *ABCG8* gene were found in 11 patients, including two pathogenic variants (*ABCG8* rs137852987 p.(Trp361Ter) in three and *ABCG8* rs769576789 p.(Leu572Pro) in two patients) and nine variants of uncertain significance (VUS) (Tables 5–6). Thus, pathogenic variants in the *ABCG8* gene were identified in five patients (3 %). Variants of uncertain clinical significance, rs141828689 p.(Arg198Gln) and rs145164937 p.(Ala98Gly) in the *ABCG5* gene, previously described in patients with FH (Totoń-Żuranska et al., 2023), were identified in this study in two patients with hypercholesterolemia and early-onset CVD; the frequency of these genetic variants is higher in the Russian population according to the RUSeq database (Table 5). However, there is currently insufficient data to classify these variants as pathogenic. In addition, our study identified variants of uncertain clinical significance: rs1167870780 p.(Leu195Gln), rs776335488 p.(Ser569Pro) and rs113005049 p.(Ala642Thr). Variants rs1167870780 p.(Leu195Gln) and rs776335488 p.(Ser569Pro) were previously described in patients with sitosterolemia (Meašić et al., 2021; Chubykina et al., 2025), variant rs113005049 p.(Ala642Thr) – in a patient with FH (Averina et al., 2018). Variants in the *ABCG8* gene, rs776335488 p.(Ser569Pro), rs189249032 p.(Tyr613His) and rs113005049 p.(Ala642Thr),

Table 6. *ABCG8* rare variants identified in patients with lipid metabolism disorders

Number of patients	rsID	The genomic coordinate (GRCh38)	Substitution NM_013389.3/ protein	Frequency gnomAD	Frequency RUSeq	Clinical significance according to the ClinVar database
1	rs1167870780	chr2:43852376T>A	c.584T>A p.(Leu195Gln)	0.000003810	–	Uncertain significance
3	rs137852987	chr2:43872094G>A	c.1083G>A p.(Trp361Ter)	0.00102	0.0005935	Pathogenic
1	rs115227860	chr2:43873940C>T	c.1365C>T p.(Ile455=)	0.00169	0.0003131	Likely benign
1	rs776335488	chr2:43875362T>C	c.1705T>C p.(Ser569Pro)	0.00001	0.0003132	Uncertain significance
2	rs769576789	chr2:43875372T>C	c.1715T>C p.(Leu572Pro)	0.00007	0.0005924	Pathogenic
1	rs189249032	chr2:43877641T>C	c.1837T>C p.(Tyr613His)	0.00006	0.0002962	Uncertain significance
2	rs113005049	chr2:43877815G>A	c.1924G>A p.(Ala642Thr)	0.0008841	0.008137	

Table 7. *NPC1L1* rare variants identified in patients with lipid metabolism disorders

Genetic variant	Number of patients	The level of TC in blood plasma, mmol/L	The level of LDL-C in blood plasma, mmol/L	Manifestation of atherosclerosis	Frequency gnomAD	Frequency RUSeq
Haplotype p.(Val177Ile) /p.(His221Tyr) /p.(Ala271Phe)§	13 adults 6 children	8.5 ± 0.9 7.3 ± 0.9	6.2 ± 1.3 5.8 ± 0.8	7 of 13 (54 %) adults	0.0063*	0.042*
Haplotype p.(His221Tyr) /p.(Ala271Phe)	2 adults 1 child	8.7 8.0 6.9	6.2 4.9 4.3	No	–	–
rs375614485 chr7:44541254C>T c.6G>A p.(Ala2=)	1	5.1	3.4	Yes	0.0018	0.0015
rs116204045 chr7:44539876C>T c.521G>A p.(Arg174His)	2	9.1 9.1	6.8 7.4	No	0.0009	0.0011
rs757263723 chr7:g.44538977G>A c.1420C>T p.(Pro474Ser)	1	–	–	–	0.00001	–
rs758137107 chr7:44534597C>T c.2016G>A p.(Gly672=)	1	9.3	5.5	Yes	0.0003	0.0005
rs137889714 chr7:44533515G>A c.2325C>T p.(Thr775=)	2	10.0 9.3	7.0 6.6	Yes	0.0005	0.0022
chr7:44533487G>A c.2353C>T p.(Leu785Phe)	1 child	6.7	–	–	–	–

Note. § Combination of genetic variants: chr7:44539868C>T c.529G>A p.(Val177Ile) (rs149017550), chr7:44539736G>A c.661C>T p.(His221Tyr) (rs114376659) and chr7:44539585_44539586GC>AA c.811_812delinsTT p.(Ala271Phe) (rs117724326/rs139533378).
* TC – total cholesterol; LDL-C – low-density lipoprotein cholesterol. The frequency is stated for rs149017550 chr7: 44539868C>T c.529G>A p.(Val177Ile).

as well as the pathogenic variant rs1167870780 p.(Leu572Pro), are found in the Russian population more often than it is indicated in the gnomAD database (Table 6).

Nine *NPC1L1* rare variants were identified in 29 patients (Table 7). 19 patients were carriers of three variants – p.(Val177Ile), p.(His221Tyr) and p.(Ala271Phe) – which, as we assume, are in linkage disequilibrium and in combination compose a haplotype (Fath et al., 2020) with an increased frequency in the Russian population (Table 7). This assumption is supported by the results of Sanger sequencing performed in two families, where relatives of the patients were available for analysis (Fig. 2). We also identified a homozygous carrier of

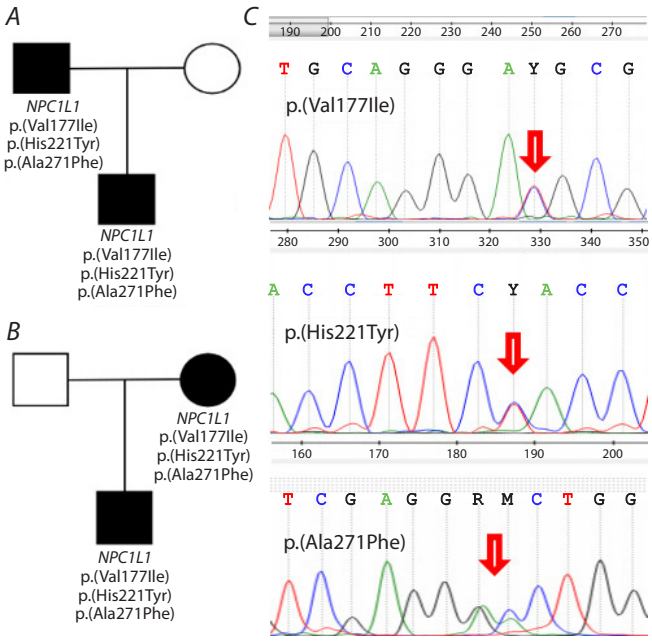


Fig. 2. Sanger sequencing for two familial cases of haplotype p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) in the *NPC1L1* gene.

A, *B* – pedigrees of patients carrying the p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) haplotype of the *NPC1L1* gene; *C* – example of Sanger verification of the p.(Val177Ile), p.(His221Tyr) and p.(Ala271Phe) variants of the *NPC1L1* gene.

the *NPC1L1* p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) haplotype. Note that the single-nucleotide substitutions described in the database – rs117724326 p.(Ala271Val) (frequency according to RUSeq: 0.04729) and rs139533378 p.(Ala271Ser) (frequency according to RUSeq: 0.04729) – in adjacent nucleotides are in fact a single variant, c.811_812delinsTT p.(Ala271Phe) (Fig. 3). It should be noted that the *NPC1L1* p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) haplotype is typically found in patients in whom monogenic dyslipidemia cannot be confirmed. In our study, only 1 carrier of this haplotype also had a mutation in the *LDLR* gene.

Evaluation of the contribution of the p.(Val177Ile) variant of *NPC1L1* gene to the development of atherosclerosis

The *NPC1L1* gene variants p.(Val177Ile), p.(His221Tyr) and p.(Ala271Phe) are located in the N-terminal domain of the NPC1L1 protein, which is responsible for cholesterol binding (Hu et al., 2021). Carriage of this haplotype may lead to enhanced intestinal cholesterol absorption and may be associated with the development of dyslipidemia and arterial atherosclerosis (Fath et al., 2020). A method based on PCR and restriction analysis was developed for the identification of the p.(Val177Ile) variant of the *NPC1L1* gene. Genotyping was performed in a group of patients with atherosclerosis of various localizations, as well as in a control group (Table 8).

The frequency of the p.(Val177Ile) variant of the *NPC1L1* gene in the control group was 0.043, what is comparable with the frequency reported in the RUSeq database (0.042). In the group of patients with atherosclerosis, the frequency was 0.067, which is higher than in the control group; however, the differences did not reach statistical significance. When performing regression analysis using the entire dataset – and additionally accounting for factors such as sex, age, and the presence of atherosclerosis – we were unable to demonstrate an association between the p.(Val177Ile) variant of the *NPC1L1* gene and the development of atherosclerosis on the base on dyslipidemia. This result may be attributed to the insufficient size of the study sample sets; further research is required.

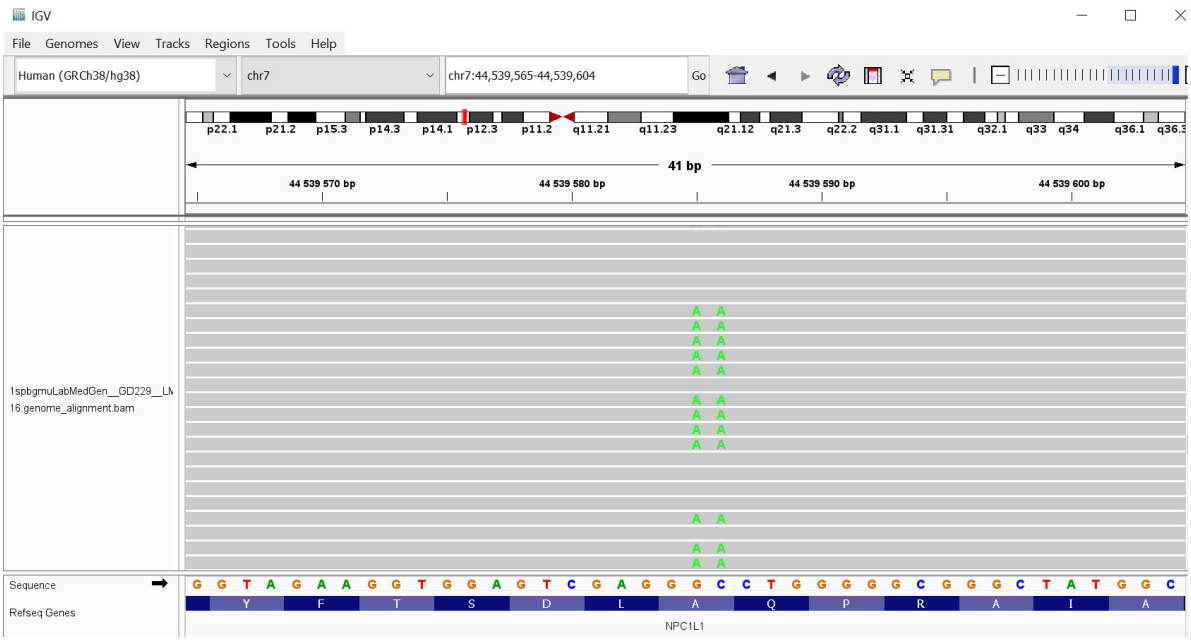


Fig. 3. Results of NGS-sequencing for the p.(Ala271Phe) variant of the *NPC1L1* gene (bam-file).

Table 8. Comparison of the frequency of the rs149017550 p.(Val177Ile) variant of the *NPC1L1* gene across all studied groups

Genotype	Homozygote GG (Val/Val)	Carriers of the p.(Val177Ile) variant (including homozygous)	Frequency of the p.(Val177Ile) variant
Control group, *N* = 234	214	20 (0)	0.043
Group of patients with atherosclerosis, *N* = 209	182	27 (1)*	0.067
Group of patients with lipid metabolism disorders (sequencing data), *N* = 169	150	19 (1)**	0.059

* $p = 0.136$ when compared with the control group; the groups are comparable in age and gender.
** $p = 0.366$ when compared with the control group; the groups are not comparable in age and gender.

Discussion

Due to the high prevalence of CVDs, the impact of various factors on the development of this pathological process remains a relevant issue. A significant genetic contribution to the progression of CVDs highlights the prospects of clarifying the status of rare variants associated with lipid accumulation and subsequent vascular damage (Meshkov et al., 2022). The complexity of the genetic architecture of dyslipidemias underscores the importance of implementing modern diagnostic approaches for screening and early identification of carriers of rare genetic variants (Kalwick, Roth, 2025).

The most common hereditary dyslipidemia is FH (Miroshnikova et al., 2023a). FH is an autosomal dominant genetic disorder characterized by high levels of cholesterol and LDL-C in blood plasma. Typical phenotypic manifestations include xanthomas, xanthelasmas and corneal arcus. Genetic diagnosis in patients with FH can be discovered in only 40–60 % of cases with a monogenic nature of the disease (Medeiros et al., 2024; Miroshnikova et al., 2025). The prevalence of FH varies from 1:250 to 1:173 depending on the population (Meshkov et al., 2021; Toft-Nielsen et al., 2022). Additionally, there are more rare monogenic dyslipidemias, such as lipoprotein lipase deficiency (Hegele et al., 2015), dysbetalipoproteinemia (Heidemann et al., 2022), sitosterolemia (Miroshnikova et al., 2023b), and others (Ivanova et al., 2020). The remaining cases of lipid metabolism disorders may be of polygenic nature (Futema et al., 2015).

As previously shown in GWAS studies, common genetic variants play an important role in predisposing to lipid metabolism disorders (Ripatti et al., 2016), but rare and low-frequency variants may also significantly contribute to the development of dyslipidemia and subsequent progression of atherosclerosis (Hindy et al., 2022). The population-specific distribution of genetic variants highlights the need to assess allele frequencies separately for each population (Read et al., 2021; Senftleber et al., 2024; Fan et al., 2025). In this study, we examined the spectrum of rare variants (gnomAD frequency <1 %) in cholesterol transporter genes – *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8* and *NPC1L1* – in patients with lipid metabolism disorders in the population of the Northwestern region of Russia.

We identified several variants in the *ABCG5* and *ABCG8* genes previously reported in patients with FH or sitosterolemia (*ABCG5* rs141828689 (p.Arg198Gln); *ABCG8* rs776335488 p.(Ser569Pro), rs769576789 p.(Leu572Pro), rs189249032 p.(Tyr613His), rs113005049 p.(Ala642Thr)), several among them were more prevalent in the Russian population (Tables 5 and 6). Homozygous and compound heterozygous pathogenic variants in these genes cause a rare autosomal recessive genetic disorder, sitosterolemia, with an estimated prevalence of 1:200,000 (Pshenichnikova et al., 2024). This condition is characterized by elevated blood levels of plant sterols, as well as total cholesterol and LDL-C. Patients with sitosterolemia often present with xanthomas and early-onset cardiovascular diseases, making the clinic signs similar to FH. Heterozygous carriage of pathogenic *ABCG5* or *ABCG8* variants does not cause sitosterolemia but may increase the risk of hypercholesterolemia (Reeskamp et al., 2020). In our study, heterozygous carriers of pathogenic and likely pathogenic variants in *ABCG5* and *ABCG8* accounted for 3 %, which is consistent with previously published data (Reeskamp et al., 2020; Medeiros et al., 2024).

Our study allowed us to pay attention to the increased frequency of the p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) haplotype of the *NPC1L1* gene both in patients with lipid metabolism disorders in the Northwestern region of Russia and in the Russian population overall. Active investigation of the NPC1L1 transporter is driven by the potential for dyslipidemia therapy. Currently, ezetimibe – an NPC1L1 inhibitor that blocks intestinal cholesterol absorption – is widely used. Ezetimibe acts as an allosteric inhibitor of NPC1L1, inducing a "closed" conformation of the transporter, which disrupts cholesterol binding (Valdivia et al., 2023). Polymorphic variants in the *NPC1L1* gene have been shown to account for inter-individual differences in sensitivity to ezetimibe (Liao et al., 2022) and may lead to a complete lack of response (Mauriello et al., 2023). Therefore, studying *NPC1L1* genetic variants as risk factors for dyslipidemia and as pharmacogenetic markers for predicting the efficacy of lipid-lowering therapy is an important goal (Liao et al., 2022). Typically, variants in this gene are associated with reduced protein activity and are therefore protective, but variants that increase cholesterol absorption are also known – for example, p.(Arg174His), which we identified in two patients (Mokhtar et al., 2022).

Substitutions p.(Arg174His), p.(Val177Ile), p.(His221Tyr), and p.(Ala271Phe) are located in the N-terminal domain of the NPC1L1 protein, which plays a key role in cholesterol uptake (Valdivia et al., 2023; Yoon et al., 2023). Variability in cholesterol absorption and plasma LDL-C levels has been pre-

viously demonstrated to depend on a combination of rare variants in the *NPC1L1* gene (Simonen et al., 2023). In the study by F. Fath et al., the *NPC1L1* p.(Val177Ile), p.(His221Tyr) and p.(Ala271Phe) variants were described in a patient with hypercholesterolemia, and their combination was classified by the authors as a possible cause of the disease (Fath et al., 2020), which is consistent with our findings. For the Russian population, the p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) haplotype has been shown to be population-specific: it is not rare but rather polymorphic. We also assessed for the first time the contribution of one of the haplotype variants, p.(Val177Ile), to atherosclerosis development. However, differences in genotype frequencies between patients with atherosclerosis and the control group did not reach statistical significance, which does not allow us to draw a conclusion about the impact of the p.(Val177Ile) variant in the *NPC1L1* gene on atherosclerosis. Our results do not indicate a definitive absence of contribution from either the individual variant or the p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) haplotype to the polygenic background; thus, expanding the sample size in future studies may allow a more precise verification of the significance of these variants. Further research is needed to investigate the contribution of both individual variants and the haplotype to the polygenic risk of hypercholesterolemia and the efficacy of ezetimibe treatment.

Conclusion

Currently, in the absence of monogenic variants, polygenic hypercholesterolemia is assumed: when no significant pathogenic genetic variants are present, the accumulation of common and rare genetic variants with small effects may lead to dyslipidemia. This study examined the spectrum of rare variants in cholesterol transporter genes – *ABCA1*, *ABCG1*, *ABCG5*, *ABCG8* and *NPC1L1* – in patients with lipid metabolism disorders who underwent genetic testing. Our work indicates a higher frequency of some rare variants in these genes – particularly *NPC1L1* – in the Russian population. However, to determine their contribution to CVD risk, their prevalence must be assessed in larger patient groups. Additional studies are needed to develop population-specific genetic risk scores that account for the cumulative contribution of risk and protective alleles in lipid metabolism-related genes and can be used to predict individual risk of dyslipidemia and cardiovascular diseases.

Study limitations. The limitations of the study include: relatively small sample sizes for investigating population-specific patterns of genetic variant distribution; specific features of sample formation. In particular, the first sample set included individuals with severe dyslipidemia who were prescribed genetic testing. Consequently, the proportion of individuals with monogenic forms of dyslipidemia – primarily familial hypercholesterolemia (FH) – in this sample set was relatively high, accounting for 30 %. On the other hand, it should be noted that the haplotype p.(Val177Ile)/p.(His221Tyr)/p.(Ala271Phe) of the *NPC1L1* gene was detected predominantly in individuals without an identified monogenic cause of dyslipidemia, with the exception of one patient. A limitation of the sample set used in the second stage of the study is that the group of patients with atherosclerosis was formed based on the presence of the disease, regardless of whether the patients initially had dyslipidemia. This is because the duration and effectiveness of statin treatment were difficult to track in most cases.

References

Averkova A.O., Brazhnik V.A., Speshilov G.I., Rogozhina A.A., Koroleva O.S., Zubova E.A., Galyavich A.S., Tereshenko S.N., Boyeva O.I., Zateyshchikov D.A. Targeted sequencing in patients with clinically diagnosed hereditary lipid metabolism disorder and acute coronary syndrome. *Bull Rus State Med Univ*. 2018;5:80-86. doi 10.24075/brsmu.2018.061

Barbitoff Y.A., Khmelkova D.N., Pomerantseva E.A., Slepchenkov A.V., Zubashenko N.A., Mironova I.V., Kaimonov V.S., … Aseev M.V., Shcherbak S.G., Glotov O.S., Isaev A.A., Predeus A.V. Expanding the Russian allele frequency reference via cross-laboratory data integration: insights from 7452 exome samples. *Natl Sci Rev*. 2024;11(10):nwae326. doi 10.1093/nsr/nwae326

Chubykina U., Vasiluev P., Ivanova O., Ezhov M. The mysterious masks of hypercholesterolemia: a unique clinical case. *Circulation*. 2025; 151(11):799-803. doi 10.1161/CIRCULATIONAHA.124.071638

Cross B., Turner R., Pirmohamed M. Polygenic risk scores: an overview from bench to bedside for personalised medicine. *Front Genet*. 2022;13:1000667. doi 10.3389/fgene.2022.1000667

Danilova I., Rau R., Barbieri M., Grigoriev P., Jdanov D.A., Meslé F., Vallin J., Shkolnikov V.M. Subnational consistency in cause-of-death data: the cases of Russia, Germany, the United States, and France. *Population*. 2021;76(4):645-674. doi 10.3917/popu.2104.0693

Ezhov M.V., Bazhan S.S., Ershova A.I., Meshkov A.N., Sokolov A.A., Kukharchuk V.V., Gurevich V.S., … Balakhonova T.V., Filippov A.E., Akhmedzhanov N.M., Aleksandrova O.Yu., Lipovetsky B.M. Clinical guidelines for familial hypercholesterolemia. *Ateroscleroz*. 2019;15(1):58-98 (in Russian)

Fan H.Y., Tsai M.C., Lai C.J., Yeh C.L., Hsu H.Y., Lai P.J., Hsu H.C., Su T.C., Lin H.J., Lin Y.F., Lu T.P., Chien K.L. Genetic variants in severe hypertriglyceridemia among Taiwanese participants – insights from genome-wide association and whole-exome sequencing analyse. *Circ J*. 2025;89(3):331-339. doi 10.1253/circj.CJ-24-0491

Fath F., Bengeser A., Barresi M., Binner P., Schwab S., Ray K.K., Krämer B.K., Fraass U., März W. FH ALERT: a feasibility study of a novel approach to identify patients with familial hypercholesterolemia. *Res Square*. 2020. doi 10.21203/rs.3.rs-108034/v1

Futema M., Shah S., Cooper J.A., Li K.W., Whittall R.A., Sharifi M., Goldberg O., … Roeters van Lennep J.E., Sijbrands E.J.G., Whittaker J.C., Talmud P.J., Humphries S.E. Refinement of variant selection for the LDL cholesterol genetic risk score in the diagnosis of the polygenic form of clinical familial hypercholesterolemia and replication in samples from 6 countries. *Clin Chem*. 2015;61(1):231-238. doi 10.1373/clinchem.2014.231365

Ghaleb Y., Elbitar S., Philippi A., El Khoury P., Azar Y., Andrianirina M., Loste A., … Deleuze J.F., Rabès J.P., Boileau C., Abifadel M., Varret M. Whole exome/genome sequencing joint analysis of a family with oligogenic familial hypercholesterolemia. *Metabolites*. 2022; 12(3):262. doi 10.3390/metabo12030262

Hegele R.A., Ban M.R., Cao H., McIntyre A.D., Robinson J.F., Wang J. Targeted next-generation sequencing in monogenic dyslipidemias. *Curr Opin Lipidol*. 2015;26(2):103-113. doi 10.1097/MOL.0000000000000163

Heidemann B.E., Koopal C., Baass A., Defesche J.C., Zuurbier L., Mulder M.T., Roeters van Lennep J.E., Riksen N.P., Boot C., Marais A.D., Visseren F.L.J. Establishing the relationship between familial dysbetalipoproteinemia and genetic variants in the *APOE* gene. *Clin Genet*. 2022;102(4):253-261. doi 10.1111/cge.14185

Heron M. Deaths: leading causes for 2018. *National Vital Stat Rep*. 2021;70(4):1-115. doi 10.15620/cdc:104186

Hindy G., Dornbos P., Chaffin M.D., Liu D.J., Wang M., Selvaraj M.S., Zhang D., … Willer C.J., Natarajan P., Flannick J.A., Khera A.V., Peloso G.M. Rare coding variants in 35 genes associate with circulating lipid levels – a multi-ancestry analysis of 170,000 exomes. *Am J Hum Genet*. 2022;109(1):81-96. doi 10.1016/j.ajhg.2021.11.021

Hu M., Yang F., Huang Y., You X., Liu D., Sun S., Sui S.F. Structural insights into the mechanism of human NPC1L1-mediated cholesterol uptake. *Sci Adv*. 2021;7(29):eabg3188. doi 10.1126/sciadv.abg3188

Ivanova O.N., Vasiliev P.A., Zakharova E.Y. Molecular basis of primary monogenic dyslipidemia. *Medical Genetics*. 2020;19(12):4-17. doi 10.25557/2073-7998.2020.12.4-17 (in Russian)

Kalwick M., Roth M. A comprehensive review of the genetics of dyslipidemias and risk of atherosclerotic cardiovascular disease. *Nutrients*. 2025;17(4):659. doi 10.3390/nu17040659

Koseki M., Yamashita S., Ogura M., Ishigaki Y., Ono K., Tsukamoto K., Hori M., Matsuki K., Yokoyama S., Harada-Shiba M. Current diagnosis and management of tangier disease. *J Atheroscler Thromb*. 2021;28(8):802-810. doi 10.5551/jat.RV17053

Kukharchuk V.V., Ezhov M.V., Sergienko I.V., Arabidze G.G., Bubnova M.G., Balakhonova T.V., Gurevich V.S., … Sokolov A.A., Sumarokov A.B., Gornyakova N.B., Obrezan A.G., Shaposhnik I.I. Diagnosis and correction of lipid metabolism disorders in order to prevent and treat atherosclerosis. Russian recommendations VII revision. *J Atherosclerosis Dyslipidemias*. 2020;1(38):7-40. doi 10.34687/2219-8202.JAD.2020.01.0002 (in Russian)

Li W., Wang Y., Huang R., Lian F., Xu G., Wang W., Xue S. Rare and common coding variants in lipid metabolism-related genes and their association with coronary artery disease. *BMC Cardiovasc Disord*. 2024;24(1):97. doi 10.1186/s12872-024-03759-5

Liao J., Yang L., Zhou L., Zhao H., Qi X., Cui Y., Ouyang D. The *NPC1L1* gene exerts a notable impact on the reduction of low-density lipoprotein cholesterol in response to hyzetimibe: a factorial-designed clinical trial. *Front Pharmacol*. 2022;13:755469. doi 10.3389/fphar.2022.755469

Mauriello A., Ascrizzi A., Molinari R., Falco L., Caturano A., D'Andrea A., Russo V. Pharmacogenomics of cardiovascular drugs for atherothrombotic, thromboembolic and atherosclerotic risk. *Genes*. 2023;14(11):2057. doi 10.3390/genes14112057

McPherson R., Tybjaerg-Hansen A. Genetics of coronary artery disease. *Circ Res*. 2016;118(4):564-578. doi 10.1161/CIRCRESAHA.115.306566

Meašić A.M., Bobinec A., Sansović I., Boban L., Pavić A.M., Pušeljić S., Barišić I. 96 homozygous ABCG8 mutation in a 14-year-old boy with sitosterolemia. *Arch Dis Child*. 2021;106(Suppl.2):A1-A216

Medeiros A.M., Alves A.C., Miranda B., Chora J.R., Bourbon M., Rato Q., Gaspar A., … Correia S., Vassalo T., Pack T., Martins V., Vieira V.F. Unraveling the genetic background of individuals with a clinical familial hypercholesterolemia phenotype. *J Lipid Res*. 2024;65(2):100490. doi 10.1016/j.jlr.2023.100490

Meshkov A.N., Ershova A.I., Kiseleva A.V., Shalnova S.A., Drapkina O.M., Boytsov S.A.; on behalf of the FH-ESSE-RF investigators. The prevalence of heterozygous familial hypercholesterolemia in selected regions of the Russian Federation: the FH-ESSE-RF study. *J Pers Med*. 2021;11(6):464. doi 10.3390/jpm11060464

Meshkov A.N., Kiseleva A.V., Yershova A.I., Sotnikova E.A., Smetnev S.A., Limonova A.S., Zharikova A.A., Zaichenoka M., Ramensky V.E., Drapkina O.M. *ANGPTL3*, *ANGPTL4*, *APOA5*, *APOB*, *APOC2*, *APOC3*, *LDLR*, *PCSK9*, *LPL* gene variants and coronary artery disease risk. *Russian Journal of Cardiology*. 2022;27(10):22-26. doi 10.15829/1560-4071-2022-5232 (in Russian)

Miroshnikova V.V., Pchelina S.N., Donnikov M.Yu., Vorobyev A.S., Tsai V.V., Kovalenko L.V., Glotov O.S. The NGS panel for genetic testing in cardiology: from the evaluation of disease risk to pharmacogenetics. *Farmakogenetika i Farmakogenomika = Pharmacogenetics and Pharmacogenomics*. 2023a;1:7-19. doi 10.37489/2588-0527-2023-1-7-19 (in Russian)

Miroshnikova V.V., Vasiluev P.A., Linkova S.V., Soloviov V.M., Ivanova O.N., Tolmacheva E.R., Udalova V.Y., … Grunina M.N., Smirnova N.N., Kuchina A.S., Zakharova E.Y., Pchelina S.N. Pediatric patients with sitosterolemia: next-generation sequencing and biochemical examination in clinical practice. *J Pers Med*. 2023b;13(10):1492. doi 10.3390/jpm13101492

Miroshnikova V.V., Izyumchenko A.D., Muzalevskaya M.V., Legostaeva K.V., Grunina M.N., Dracheva K.V., Urazgildeeva S.A., Berkovich O.A., Baranova E.I., Glotov O.S., Kulikov A.N., Gurevich V.S., Pchelina S.N. Genetic architecture of familial hypercholesterolemia: a cohort of St. Petersburg residents. *Russian Journal of Cardiology*. 2025;30(10):87-94. doi 10.15829/1560-4071-2025-6432 (in Russian)

Mokhtar F.B.A., Plat J., Mensink R.P. Genetic variation and intestinal cholesterol absorption in humans: a systematic review and a gene network analysis. *Prog Lipid Res*. 2022;86:101164. doi 10.1016/j.plipres.2022.101164

Musunuru K., Kathiresan S. Surprises from genetic analyses of lipid risk factors for atherosclerosis. *Circ Res*. 2016;118(4):579-585. doi 10.1161/CIRCRESAHA.115.306398

Pan-Lizcano R., Mariñas-Pardo L., Núñez L., Rebollal-Leal F., López-Vázquez D., Pereira A., Molina-Nieto A., Calviño R., Vázquez-Rodríguez J.M., Hermida-Prieto M. Rare variants in genes of the cholesterol pathway are present in 60 % of patients with acute myocardial infarction. *Int J Mol Sci*. 2022;23(24):16127. doi 10.3390/ijms232416127

Prasad K., Mishra M. Mechanism of hypercholesterolemia-induced atherosclerosis. *Rev Cardiovasc Med*. 2022;23(6):212. doi 10.31083/j.rcm2306212

Pshenichnikova I.I., Zakharova I.N., Korchagina Yu.V., Pupykina V.V., Okulova O.A., Teleznikova N.D., Vasiliev P.A., Ivanova O.N., Baranova P.V., Zakharova E.Yu., Ezhov M.V. Sitosterolemia (phytosterolemia): diagnosis, treatment, and prognosis. *Meditsinskiy Sovet = Medical Council*. 2024;18(19):198-205. doi 10.21518/ms2024-432 (in Russian)

Read R.W., Schlauch K.A., Lombardi V.C., Cirulli E.T., Washington N.L., Lu J.T., Grzymski J.J. Genome-wide identification of rare and common variants driving triglyceride levels in a Nevada population. *Front Genet*. 2021;12:639418. doi 10.3389/fgene.2021.639418

Reeskamp L.F., Volta A., Zuurbier L., Defesche J.C., Hovingh G.K., Grefhorst A. *ABCG5* and *ABCG8* genetic variants in familial hypercholesterolemia. *J Clinical Lipidology*. 2020;14(2):207-217.e7. doi 10.1016/j.jacl.2020.01.007

Ripatti P., Rämö J.T., Söderlund S., Surakka I., Matikainen N., Pirinen M., Pajukanta P., … Wilson R.K., Palotie A., Freimer N.B., Taskinen M.R., Ripatti S. The contribution of GWAS loci in familial dyslipidemias. *PLoS Genet*. 2016;12(5):e1006078. doi 10.1371/journal.pgen.1006078

Roberts R., Chang C.C., Hadley T. Genetic risk stratification: a paradigm shift in prevention of coronary artery disease. *JACC Basic Transl Sci*. 2021;6(3):287-304. doi 10.1016/j.jacbts.2020.09.004

Ryzhkova O.P., Kardymon O.L., Prokhorchuk E.B., Konovalov F.A., Maslennikov A.B., Stepanov V.A., Afanasyev A.A., … Kostareva A.A., Pavlov A.E., Golubenko M.V., Polyakov A.V., Kutsev S.I. Guidelines for the interpretation of massive parallel sequencing variants (update 2018, v2). *Med Genet*. 2019;18(2):3-23. doi 10.25557/2073-7998.2019.02.3-23 (in Russian)

Senftleber N.K., Andersen M.K., Jørsboe E., Stæger F.F., Nøhr A.K., Garcia-Erill G., Meisner J., … Jørgensen M.E., Zeggini E., Moltke I., Hansen T., Albrechtsen A. GWAS of lipids in Greenlanders finds association signals shared with Europeans and reveals an independent PCSK9 association signal. *Eur J Hum Genet*. 2024;32(2):215-223. doi 10.1038/s41431-023-01485-8

Shakhtshneider E., Ivanoshchuk D., Timoshchenko O., Orlov P., Semaev S., Valeev E., Goonko A., Ladygina N., Voevoda M. Analysis of rare variants in genes related to lipid metabolism in patients with familial hypercholesterolemia in Western Siberia (Russia). *J Pers Med*. 2021;11(11):1232. doi 10.3390/jpm11111232

Shim S.Y., Yoon H.Y., Yee J., Han J.M., Gwak H.S. Association between *ABCA1* gene polymorphisms and plasma lipid concentration: a systematic review and meta-analysis. *J Pers Med*. 2021;11(9):883. doi 10.3390/jpm11090883

Simonen P., Öörni K., Sinisalo J., Strandberg T.E., Wester I., Gylling H. High cholesterol absorption: a risk factor of atherosclerotic cardiovascular diseases? *Atherosclerosis*. 2023;376:53-62. doi 10.1016/j.atherosclerosis.2023.06.003

Tada H., Kojima N., Takamura M., Kawashiri M.A. Sitosterolemia. *Adv Clin Chem*. 2022;110:145-169. doi 10.1016/bs.acc.2022.06.006

Toft-Nielsen F., Emanuelsson F., Benn M. Familial hypercholesterolemia prevalence among ethnicities – systematic review and meta-analysis. *Front Genet*. 2022;13:840797. doi 10.3389/fgene.2022.840797

Tokgozoglu L., Kayikcioglu M. Familial hypercholesterolemia: global burden and approaches. *Curr Cardiol Rep*. 2021;23(10):151. doi 10.1007/s11886-021-01565-5

Totoń-Żurańska J., Wołkow P., Kapusta M., Wójcik M., Starzyk J., Kawalec E., Idzior-Waluś B., Waluś-Miarka M. Targeted sequencing of a gene panel in patients with familial hypercholesterolemia from Southern Poland. *Pol Arch Intern Med*. 2023;133(6):16417. doi 10.20452/pamw.16417

Valdivia A., Luque F.J., Llabrés S. Binding of cholesterol to the N-terminal domain of the *NPC1L1* transporter: analysis of the epimerization-related binding selectivity and loop mutations. *J Chem Inf Model*. 2023;64(1):189-204. doi 10.1021/acs.jcim.3c01319

Yoon H.J., Lee Y., Jeong J., Jang S., Lee H.H., Kim G.S. Binding free energy of several sterols to the N-terminal domain of Niemann-Pick C1-like 1 protein due to mutation: molecular dynamics study. *J Chin Chem Soc*. 2023;70(3):539-546. doi 10.1002/jccs.202200315

Yu X.H., Tang C.K. ABCA1, ABCG1, and cholesterol homeostasis. In: Zheng L. (Ed.) HDL Metabolism and Diseases. Advances in Experimental Medicine and Biology. Vol. 1377. Springer, Singapore, 2022;95-107. doi 10.1007/978-981-19-1592-5_7

Zhang R., Liu W., Zeng J., Meng J., Shi L., Yang S., Chang J., Wang C., Xing K., Wen J., Liu N., Liang B., Xing D. Recent advances in the screening methods of NPC1L1 inhibitors. *Biomed Pharmacother*. 2022;155:113732. doi 10.1016/j.biopha.2022.113732

Conflict of interest. The authors declare no conflict of interest.
Received October 16, 2025. Revised January 15, 2026. Accepted February 17, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):502-514

doi 10.18699/vjgb-26-21

Mathematical and computational modeling of biosystems at different levels of organization

S.A. Lashin, R.A. Ivanov, Y.G. Matushkin

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

Abstract. Modern biology increasingly relies on mathematical and computational modeling to describe complex hierarchically organized biological systems. This review considers models that cover the main levels of biological organization, from the molecular-genetic and cellular levels to tissue/organ, organismal, population and ecological ones. The aim of the work is to systematize the key modeling approaches at each of these levels, to analyze their capabilities and limitations, and to discuss strategies for constructing multiscale and hybrid models that consistently link processes operating at different spatial and temporal scales. We survey classical deterministic and stochastic models based on ordinary and partial differential equations, logical and graph-based models of regulatory networks, cellular automata, agent-based models, as well as flux-balance approaches. Typical examples are given for the modeling of gene regulatory and metabolic networks, chemotaxis, tissue and organ growth, population dynamics and genetic structure, and ecosystem functioning. Special attention is paid to comparing approaches with respect to the scale of description, complexity of modeled processes, data requirements, computational cost and interpretability of results. The analysis shows that hybrid and multiscale models provide an adequate framework to account for nonlinearity, stochasticity and structural heterogeneity of biosystems, but require substantial computational resources and careful data-driven calibration. Methodological and technological trends are outlined, including the development of specialized platforms and model repositories, standards for model representation and tools for reuse of model components.

Key words: mathematical modeling; biological systems; multiscale models; computational biology

For citation: Lashin S.A., Ivanov R.A., Matushkin Y.G. Mathematical and computational modeling of biosystems at different levels of organization. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):502-514. doi 10.18699/vjgb-26-21

Funding. The study was supported by the budget project No. FWNR-2025-0032.

Математическое и компьютерное моделирование биологических систем на разных иерархических уровнях организации

С.А. Лашин, Р.А. Иванов, Ю.Г. Матушкин

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

Аннотация. Современная биология все в большей степени опирается на математическое и компьютерное моделирование для описания сложных иерархически организованных биосистем. В данном обзоре рассматриваются математические модели, охватывающие основные уровни биологической организации – от молекулярно-генетического и клеточного до тканевого/органного, организменного, популяционного и экологического. Цель работы состоит в систематизации ключевых подходов к моделированию на каждом из этих уровней, анализе их возможностей и ограничений, а также в обсуждении стратегий построения многомасштабных и гибридных моделей, связывающих воедино процессы разных пространственно-временных масштабов. Рассматриваются классические детерминированные и стохастические модели на основе дифференциальных уравнений в частных производных, логические и графовые модели регуляторных сетей, клеточные автоматы, агентно-ориентированные и индивидуально-ориентированные модели и подходы, базирующиеся на балансе потоков. Приводятся типичные примеры моделирования молекулярно-генетических сетей, метаболизма и хемотаксиса, роста тканей и органов, динамики популяций и генетической структуры, а также функционирования экосистем. Особое внимание уделяется сопоставлению подходов по критериям масштабов описания, сложности моделируемых процессов, доступности

исходных данных, вычислительной трудоемкости и интерпретируемости результатов. Обзор обобщает отечественный и зарубежный опыт, подчеркивая вклад российских и, в частности, новосибирских коллективов в развитие гибридных методов моделирования, построения многомасштабных моделей и реализации программных платформ для системной биологии. В результате проведенного анализа показано, что гибридные и многомасштабные модели позволяют наиболее полно учесть нелинейность, стохастичность и структурную неоднородность биологических систем, но требуют значительных вычислительных ресурсов и тщательной калибровки по данным. Отмечаются методические и программно-технологические тенденции, включая развитие специализированных платформ и репозиториев моделей, средств стандартизации описания и повторного использования модельных компонентов.
Ключевые слова: математическое моделирование; биологические системы; многомасштабные модели; компьютерная биология

Introduction

Biological systems are complex, hierarchically organized systems where differences in the characteristic sizes of objects span up to ten orders of magnitude, and differences in the characteristic timescales of processes span up to 18 orders of magnitude (Riznichenko, 2003; Shumnyi et al., 2006).

Highlighting the levels of biological organization that are most developed in terms of mathematical description, we note the following:

- Molecular-genetic;
- Cellular;
- Tissue/organ;
- Organismal;
- Population;
- Ecological/biocenotic.

Each of these levels is characterized by specific objects and processes and, accordingly, requires its own mathematical and computer models for description. It should also be noted that concepts such as *genotype* and *phenotype*, being among the most frequently used terms in biology, can be described from the perspective of mathematical modeling in completely different ways. Depending on this description, they may formally belong to different levels of biological organization. Depending on the context of the situation under consideration, these allelic variants can be described as symbolic sequences, numbers, etc. The diversity of ways to describe phenotypic traits of an organism is even broader – these can be discrete traits (e. g., eye or hair color, flower color) or continuous traits (e. g., height or weight). Traits can be the result of measurements conducted using both simple instruments like rulers or scales, and complex physical and/or biochemical instruments (e. g., morphophysiological traits).

The aim of this review is to systematize the main mathematical and computer methods for modeling biological systems at various levels of organization, analyze their capabilities and limitations, and consider approaches to constructing multiscale and hybrid models that combine several levels within a unified conceptual and computational framework.

Molecular-genetic level of organization

The Molecular-Genetic Control System (MGCS) of a cell is the set of its irregular polymers (DNA, RNA, and proteins), as well as molecular subsystems performing various biochemical processes on these irregular polymers (Ratner et al., 1985):

- Synthesis;
- Transformation;
- Decay;
- Transport, etc.

Among the methodological approaches for studying the MGCS, three main ones can be noted:

1. Structural-functional approach – focuses on the material properties of macromolecules and the MGCS in relation to their function. It studies structural, physicochemical, and energy patterns of macromolecule structure, thermodynamics and kinetics of processes, etc.
2. Information-cybernetic approach – focuses on identifying the principles of organization and control of the MGCS, abstracting from their structural features. It studies self-reproduction, information processes, coding, memory, reliability, regulation systems, etc.
3. Evolutionary approach – identifies the paths of origin and evolution of the MGCS as a whole, various subsystems and fractions of macromolecules, as well as evolution factors, types of evolutionary dynamics, etc. (Ratner et al., 1985).

A particular case of the MGCS is the so-called *gene networks* – groups of coordinately functioning genes interacting with each other both through their primary products (RNA and proteins) and through various metabolites and other secondary products of gene network functioning (Kolchanov et al., 2013).

Models of biological systems at the molecular-genetic level include simple ODEs (Ordinary Differential Equations) and systems of ODEs consisting of several equations, discrete models built using various formalisms (Boolean networks, Petri nets, cellular automata, etc.), discrete-continuous models, as well as various computer simulation and agent-based models.

The simplest MGCS models were built based on chemical kinetics equations of the form (Eq. (1)), representing various ways of presenting the *kinetic law of mass action* and the Generalized Chemical Kinetic Modeling Method (GCKMM) proposed by Vitaly Likhoshvai (Likhoshvai et al., 2000):

$$\frac{dX}{dt} = V(Y, K), \qquad (1)$$

where X – vector (list) of controlled variables, Y – vector (list) of controlling variables, K – list of parameters.

The inclusion of the same variables in both lists is allowed, but, in general, the lists X and Y do not coincide and may not intersect at all. Variables usually represent concentrations of substances or probabilities of realizing selected states of substances. Variables from the list Y not included in the list X are parameters for the current elementary model. The functional V describes the law of rates of change of concentrations of substances from the list X.

The kinetic law of mass action is derived based on collision theory. Let there be a biochemical reaction:

$$A + B \leftrightarrow C. \quad (2)$$

Then the rate $V_{A+B\to C}$ of formation of complex C at current time t is equal to:

$$V_{A+B\to C}(t) = k_1[A][B], \quad (3)$$

and the rate $V_{C\to A+B}$ of decay of complex C into components A and B at current time t is equal to:

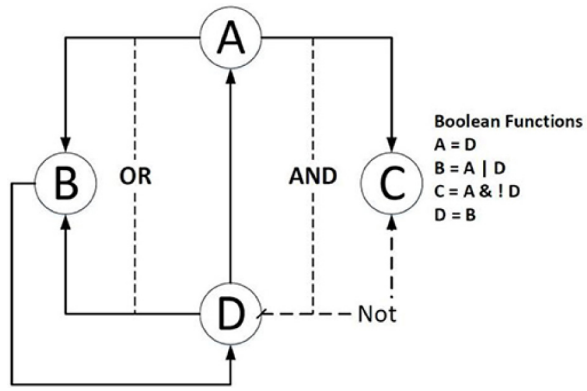
$$V_{C\to A+B}(t) = k_2[C]. \quad (4)$$

If there is a specific biochemical reaction scheme of the form (Eq. (2)), then the instantaneous rate of change of concentration of any substance equals to the sum of local rates of change of concentration of this substance in each reaction in which this substance participates (Kazantsev et al., 2009; Akberdin et al., 2013). This simple rule allows easily writing down the final system of differential equations describing the target biochemical scheme, using only first and second-order polynomials as the right-hand sides of the system equations (Eq. (1)). The theoretical basis for it is the Korzukhin theorem, which is crucial for modeling chemical kinetics: "For any set of non-negative curves defined on a finite time interval, and any given accuracy, there exists such (perhaps not the one) biochemical scheme, composed only of bimolecular and monomolecular reactions, that the mathematical model built according to this biochemical scheme approximates the given set of curves with the given accuracy" (Zhabotinskii, 1974).

The most important element in constructing more complex MGCS models using GCKMM (Likhoshvai et al., 2000) is the rule of summation of local rates of biochemical reactions: the total rate of change of component concentrations in the system is the sum of the rates of change of concentration of this component in all elementary processes (see Supplementary Materials, Table S1 for details)[1].

Logical approaches for modeling MGCS, introduced back in the 60s of the 20th century, are based on applying logic and discrete mathematics terms to describe molecular-genetic mechanisms (Kauffman, 1969). Let us give an example of describing gene network operation in terms of Boolean logic. A gene network is represented as a Boolean network – a directed graph where vertices represent gene states, and edges represent regulatory events, i. e., the action of genes on other genes (Ratushnyi et al., 2005; Tran, 2016; Barbuti et al., 2020).

[1] Supplementary Table S1 and Figures S1–S7 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Lashin_Engl_30_3.pdf



Example of a simple Boolean gene network.

The right side lists Boolean functions of state changes; the dashed line means the source gene are inhibited; the solid line means the source gene are activated (according to: (Chen et al., 2018)).

The current state of the gene network is described by a list of expressed (working) and non-expressed (non-working) genes at discrete moments in time. It is assumed that genes can be in only two states: "expressed" (true,1), or "not expressed" (false,0). The state of the gene network, thus, represents a Boolean vector:

$$X = (A_i), \quad i \in \overline{1, N}, \quad (5)$$

where A_i is the state of the i-th gene. The change of network states is described using a Boolean vector-function f:

$$X(t+1) = f(X(t)) = \\ = (f_{A_1}(X(t)), f_{A_2}(X(t)), \ldots, f_{A_N}(X(t))), \quad (6)$$

where $f_{A_i}: \{0,1\}^N \to \{0,1\}$ – describes the impact of expression of all genes on the expression of the i-th gene. An example of a simple Boolean gene network is shown in the Figure.

It should be noted that the approaches mentioned above have limited applicability in quantitative forecasting due to their simplification. Therefore, at present, the most frequently used approaches for describing MGCS are various *hybrid methods* combining continuous, discrete, and stochastic modeling methods. One of the first such methods is the generalized threshold modeling method developed in Novosibirsk by Rustem Churaev back in the 70s of the 20th century (Churaev, Ratner, 1972; Churaev, 2005). This method combines approaches of automata theory and linear ODEs. The core idea lies in dividing the system's phase space into regions, within which a system behavior is described by linear ODEs, while transition conditions between regions are described by Boolean functions. The solutions of linear systems are "stitched" at zone boundaries. Another example of hybrid approaches is a *rule-based modeling*, where the model is defined indirectly via a specific set of rules (Blinov et al., 2004; Harris et al., 2016). Further development of GCKMM and numerical modeling methods of MGCS based on it at the Institute of Cytology and Genetics was also carried out within the paradigm of hybrid approaches (Kazantsev et al., 2009, 2018).

Cellular level of organization

Mathematical modeling at the cellular level of organization considers biological processes such as transport of substances and energy from the environment into the cell and back, their metabolism, cell movement in space, and their division. Regarding modeling metabolic processes, the cellular level of organization is directly linked to the molecular-genetic level discussed in the previous section. The term "Electronic Cell", which appeared in the late 20th–early 21st century (Tomita et al., 1999; Tomita, 2001; Ishii et al., 2004; Price et al., 2004; Karr et al., 2012; Akberdin et al., 2013), implies modeling the cell primarily at the molecular-genetic level of biological organization.

Nevertheless, unlike the models discussed in the previous section, where the MGCS themselves, their states, functioning modes, etc., are of interest, cell models can focus on other events and processes, considering the MGCS models included in them as "handy tools". For example, Japanese systems biologist and bioinformatician Masaru Tomita, the leader of the E-CELL project – one of the first successful projects modeling an electronic bacterial cell – highlights among cellular level organization processes that can be modeled using MGCS processes: substance transport across the membrane, cell cycle and cell division, as well as the development of pathological states (e. g., for human erythrocyte cells) (Tomita et al., 1999; Tomita, 2001; Hucka et al., 2003). Programs developed within the E-CELL project use a hybrid approach to modeling cell viability; in particular, version E-CELL 3 uses ODE systems, stochastic modeling using the Gillespie algorithm, as well as special algorithms for "stitching" solutions.

Another important cellular process for modeling which the MGCS models can also be used is *chemotaxis* – cell movement. Bacterial chemotaxis represents one of the simplest and well-studied examples of microbial behavior. It allows swimming bacterial cells to follow chemicals in the environment along a concentration gradient. Molecular mechanisms of chemotaxis in the model bacterium *E. coli* have been studied in great detail over the last 50 years, using a wide spectrum of experimental, mainly biophysical, methods (Berg, Purcell, 1977; Vladimirov, Sourjik, 2009; Kaizu et al., 2014). The accumulation of experimental data in this area led to the creation of several chemotaxis models. For example, the model published by the group (Bray et al., 1993) is a hybrid block-modular model where the molecular-genetic component is described using chemical kinetics equations, and the physical component – rotation of special motor proteins – using a finite automaton. A schematic diagram of this model is provided in Figure S1. The model reproduces the behavior of over 30 mutants, indicating its high quality. The model was developed and implemented as the BCT (BacterialChemoTaxis) software package (Bray, Bourret, 1995), which subsequently allowed describing over 60 mutants (Bray et al., 2007). Additionally, agent-based models combining this approach with stochastic modeling of MGCS were used to study chemotaxis (AgentCell (Emonet et al., 2005)); hybrid modeling combining ODE blocks and stochastic modeling, as well as mean-field approximation for the Monod–Wyman–Changeux model[2] (RapidCell (Vladimirov et al., 2008), see also Fig. S2).

Molecular mechanisms of movement in more complex eukaryotic cells differ fundamentally from those in bacteria. Nevertheless, the methodological repertoire for mathematical modeling of these processes is generally the same. For example, a mathematical model of hair cell regulation – receptors of the auditory system and vestibular apparatus of animals and humans, describing ion transport, membrane potential, and cell movement (O'Beirne, Patuzzi, 2007), was implemented based on the Boltzmann equation.

Despite the significant background in modeling biological systems at the cellular level of organization, in most studies currently being conducted, single-cell models are used as auxiliary tools. This applies to levels of biological organization considering models of cell ensembles – tissue, organismal, and population levels.

Tissue/organ level of organization

Biological *tissue* is a system of cells similar in origin, structure, and functions performed in the organism, as well as intercellular substances and structures – products of their vitality (Gilyarov et al., 1986). *Animal tissues* are divided into four main types: connective, muscle, nervous, and epithelial. *Plant tissues* are divided into *simple*, consisting of cells of one type (e. g., collenchyma), and *complex*, consisting of different types (e. g., epidermis). Depending on the classification, two or three types of plant tissues are distinguished. In the first case, tissues are subdivided into *meristematic* (actively dividing cells, e. g., in roots or stem tips) and *permanent* (having lost the ability to divide). In the second case, tissues are divided into *tissue systems*: *epidermis*, *mechanical tissue*, and *conducting tissue* (Gilyarov et al., 1986). A combination of various interacting tissues forms organs.

Given such diversity of properties and functions of biological tissues and organs, the repertoire of mathematical and computer models representing this level of biological organization is extremely large. The methodological arsenal for constructing such models is broad and includes both classical modeling methods using systems of ODEs and partial differential equations (more often used for modeling biological tissues), and modern hybrid modeling methods including agent-based approaches, etc.

Mathematical and computer modeling of biological tissues and organs arouses the greatest interest among medics and biologists in the following contexts: (1) modeling development, i. e. *morphogenesis*, *embryogenesis*, etc.; (2) modeling

[2] The Monod–Wyman–Changeux model describes the regulation of enzymatic activity in a protein composed of identical subunits through allosteric structural changes (Monod et al., 1965).

pathologies (diseases) and strategies for their correction. In some cases, the same model can be considered in both contexts simultaneously, for example, a cancer tumor development model allowing numerical investigation of various treatment strategies.

The classical model of morphogenesis theory is the "reaction-diffusion" model proposed in the mid-20th century by Alan Turing (Turing, 1952):

$$\begin{cases} \frac{\partial x}{\partial t} = P(x,y) + D_x \frac{\partial^2 x}{\partial r^2}, \\ \frac{\partial y}{\partial t} = Q(x,y) + D_y \frac{\partial^2 y}{\partial r^2}, \end{cases} \quad (7)$$

where r – spatial coordinate, $D_x\partial^2 x/\partial r^2$ and $D_y\partial^2 y/\partial r^2$ describe diffusion of substances x and y along this coordinate. Diffusion of non-linearly linked components x and y in this system leads not to averaging, but to a distribution periodic in time and non-uniform in space (Riznichenko, 2003).

Modifications of the "reaction-diffusion" model have been actively studied and applied to a wide range of biological tasks, including modeling tissues and organs. In particular, a mathematical model implemented as system (8) was used to study animal skin coloration patterns. Depending on model parameters, which correspond to different combinations of morphogens (substances affecting individual organism development), coloration patterns vary within a wide range (Fig. S3).

$$\frac{\partial u}{\partial t} = \gamma f(u,v) + \nabla^2 u, \frac{\partial v}{\partial t} = \gamma g(u,v) + d\nabla^2 v,$$
$$f(u,v) = a - u - h(u,v), g(u,v) = \alpha(b-v) - h(u,v), \quad (8)$$
$$h(u,v) = \frac{\rho uv}{1 + u + Ku^2},$$

where f, g, h – functions describing reaction kinetics; a, b, α, ρ and K – positive parameters describing reaction kinetics, representing ratios of kinetic constants (detailed derivation of these parameters is provided in (Murray, 2003)); d – ratio of diffusion coefficients; γ determines the region size.

Another approach to modeling biological systems at the tissue and organ level of organization was proposed by Vitaly Likhoshvai and later developed in works under the supervision of Victoria Mironova on modeling plant tissues (Likhoshvai et al., 2007; Fadeev et al., 2008; Mironova et al., 2010, 2012; Novoselova et al., 2013; Pasternak et al., 2019). Its essence lies in describing plant tissue as a system of ODEs of sufficiently large dimension – from hundreds to several thousand equations. Here, a separate tissue cell corresponds to several equations; in most works – just four equations describing the dynamics of substances of interest to the researcher: proteins and hormones. Cells are considered as systems with complete mixing. Redistribution of substances between cells via diffusion and active transport processes is described by substance transport equations. An example of such equations for the hormone auxin is given in the following fragment of system (9) from the article (Likhoshvai et al., 2007) (model scheme is shown below in Fig. S4).

$$\begin{aligned} \frac{da_n}{dt} &= \alpha + P_t a_{n-1} - P_t a_n - K_d a_n + K_0 a_n f(a_n), \\ \frac{da_i}{dt} &= P_t(a_{i+1} + a_{i-1}) + K_0 a_{i+1} f(a_{i+1}) - \\ &\quad - 2P_t a_i - K_d a_i - K_0 a_i f(a_i), \quad i = \overline{n-1,2}, \\ \frac{da_1}{dt} &= -P_t a_1 - K_d a_1 + P_t a_2 + K_0 a_2 f(a_2), \end{aligned} \quad (9)$$

where a_i – concentration of hormone auxin in the i-th cell of the plant root (the root itself is considered in this model as a one-dimensional array of cells, see also Fig. S4*a*); parameter α describes constant auxin inflow into the system; P_t – auxin diffusion between cells; K_d – auxin degradation parameter in the cell; function $f(a_i)$ describes active directed transport of auxin across the membrane via special transporter proteins (Eq. (10)):

$$f(a_i) = \left(\frac{\left(\frac{a_i}{q_{11}}\right)^{p_1}}{1+\left(\frac{a_i}{q_{12}}\right)^{p_1}}\right) \times \left(\frac{1}{1+\left(\frac{a_i}{q_2}\right)^{p_2}}\right). \quad (10)$$

In this function, in turn, q_{11} – activation threshold constant for auxin-dependent transport; q_{12} – saturation threshold constant for auxin-dependent transport; q_2 – inhibition threshold constant for auxin-dependent transport; p_1 and p_2 – nonlinearity coefficients of activation and inhibition mechanisms, respectively. In the aforementioned series of works on modeling plant organs and tissues, both one-dimensional (1D) and two-dimensional (2D) models were considered (Fig. S4). Although these models represent systems of ODEs (often of large dimension), ideologically they are close to cellular automaton models, which are also frequently used for modeling biological systems at tissue and organ levels of organization.

Thus, modeling using cellular automata was used to study ontogenesis processes (Markus et al., 1999; Akberdin et al., 2007; Vitvitsky, 2014; Pałubicki et al., 2019) and vegetation (Komarov et al., 2003; Colasanti et al., 2007) of plants. This approach is also actively used for modeling tissues and organs of animals and humans, in particular, pathological states such as oncological diseases (Gevertz, Torquato, 2006; Szabó, Merks, 2013; Brüningk et al., 2019; Salguero et al., 2019), immune (Bezzi et al., 1997), infectious (Slimi et al., 2009) and others (Talaminos-Barroso et al., 2020).

In most of the works listed above, as well as in many others not included in this review, the cellular automaton modeling methodology is used in combination with other approaches, in particular, agent-based modeling and rule-based modeling (Fig. S5), mentioned earlier in section "Molecular-genetic level of organization". Such hybrid models, used primarily for modeling tissues and organs, received the name *Cellular Potts Models* (Glazier, Graner, 1993; Marée et al., 2007; Voss-Böhme, 2012).

Further development of methods for modeling biological tissues and organs led to the emergence of software packages, libraries, and platforms adapting Potts models and agent-based modeling for solving specific content-related tasks in biology.

We note such software packages as CellSys (Hoehme, Drasdo, 2010), EPISIM (Sütterlin et al., 2013, 2017), CompuCell3D (Swat et al., 2012), and others. We separately note projects on modeling whole organs, for example, the platform for modeling the human liver VirtualLiver (Holzhütter et al., 2012; Drasdo et al., 2014). The development of methods for modeling biological tissues and organs prepared the background for the emergence of computer models of whole multicellular organisms, which will be discussed in the next section.

Organismal level of organization

The history of mathematical and computer modeling of individual organism functioning dates back to the 60s of the 20th century. The first and simplest tree model was presented by Igor Poletaev and his students in 1965 (Poletaev, 1965, 1966). Despite its simplicity, the model, based on physical principles, answered the question "why does a tree not grow infinitely in height?" Subsequently, more complex model variants were built and investigated based on the simplest model (Karev, Skomorovsky, 1999; Kolobov, Frisman, 2008). The real flourishing of detailed models of individual multicellular organisms occurred in the last 10–15 years. This was largely due to the colossal progress achieved in computer performance, namely – processing speed, RAM volume, and storage device capacities. It became possible to create realistic models of living organisms. The term "Digital Twins of Biological Organisms" appeared (Barnabas, Raj, 2020; Tellechea-Luzardo et al., 2020; Miehe et al., 2021), etc.

The methodological arsenal used for modeling at the organismal level of biological organization naturally includes all approaches discussed in previous sections. Depending on the goals and tasks that authors of "digital organisms" set for themselves, some levels of biological organization, as well as corresponding biological processes, may be described in the model in more detail, and others – less. For example, in the well-known OpenWorm project, dedicated to creating a computer model of the worm *Caenorhabditis elegans* (Szigeti et al., 2014; Sarma et al., 2018; Palyanov, 2019), main attention is paid to modeling movement biomechanics and biophysics of neuroimpulse transmission. Whereas, for example, the Digital Salmon project (Omholt et al., 2013) is more focused on salmon metabolism and ontogenesis processes.

It should be noted that the work on creating computer models of whole organisms – "digital twins" – is currently being carried out by large scientific teams, and often by consortia consisting of many teams. As a rule, within these works, entire software packages and even software platforms are developed, which contribute to the development of the methodological base of mathematical and systems biology.

Population level of organization

Biological *population* – a collection of individuals of one species possessing a common gene pool and occupying a certain territory (Gilyarov et al., 1986). The history of mathematical biology is primarily linked to modeling biological populations. Starting with Leonardo Fibonacci, who in his arithmetic book "Liber Abaci" proposed a model of rabbit population size change over time (the model solution – the famous Fibonacci numbers), continuing with Thomas Robert Malthus's work "An Essay on the Principle of Population" (Malthus, 1978) and Pierre François Verhulst's "Notice on the Law that Population Follows in its Growth" (Verhulst, 1838), it is precisely population dynamics modeling that becomes the driving force of mathematical biology development. The models described in the mentioned works are, of course, very simple (see Eq. (11) and (12)), but they have served as a foundation for more complex models not only in biology but also in other fields of science – physics, chemistry, etc. Below is the Malthus equation:

$$\frac{dN}{dt} = aN, \quad a>0, \tag{11}$$

where N – population size, a – population growth rate coefficient.

The Verhulst equation looks as follows:

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right), \quad r>0, K>0, \tag{12}$$

where r – population growth rate coefficient, K – maximum population size.

Models of interacting populations were first proposed by Alfred J. Lotka (Lotka, 1909, 1920, 1925) and Vito Volterra (Volterra, 1928, 1976) and subsequently received the names "Lotka–Volterra models" or "predator–prey models":

$$\begin{aligned} \frac{dN}{dt} &= x(\varepsilon_x - \gamma_{xy} y), \\ \frac{dy}{dt} &= y(\gamma_{yx} x - \varepsilon_y), \end{aligned} \tag{13}$$

where x – number of prey, y – number of predators; coefficients: ε_x – natural prey growth, ε_y – natural predator decline (in absence of prey), γ_{xy} – impact of predators on prey numbers, γ_{yx} – impact of prey on predator numbers.

The generalized Volterra model (Eq. (14)) allows considering other types of interactions between two populations besides "predator–prey" relationships.

$$\begin{aligned} \frac{dN_1}{dt} &= a_1 N_1 + b_{12} N_1 N_2 - c_1 N_1^2, \\ \frac{dN_2}{dt} &= a_2 N_2 + b_{21} N_1 N_2 - c_2 N_2^2, \end{aligned} \tag{14}$$

where N_1, N_2 – size (density) of corresponding population, a_1, a_2 – growth rate of corresponding population, c_1, c_2 – mortality coefficient of corresponding population, b_{12}, b_{21} – coefficients of population influence on each other. Depending on the values of parameters b_{12} and b_{21}, the type of interaction between populations is determined (see the Table).

Functioning modes realized in models (13, 14) boil down to two types – stationary states of the system and oscillating modes (Riznichenko, 2002, 2017). When several (more than

Types of interactions between populations considered in the generalized Volterra model (Eq. (14)), according to: (Odum, 1975)

Type of relationship	Influence of species 1	Influence of species 2	Parameters
Symbiosis	+	+	$b_{12}, b_{21} > 0$
Commensalism	+	0	$b_{12} > 0, b_{21} = 0$
Predator–prey	+	–	$b_{12} > 0, b_{21} < 0$
Amensalism	0	–	$b_{12} = 0, b_{21} < 0$
Competition	–	–	$b_{12}, b_{21} < 0$
Neutralism	0	0	$b_{12}, b_{21} = 0$

two) populations interact with each other, the richness of dynamic modes of models increases sharply. For example, in the "predator–two prey" system, the possibility of existence of chaotic modes was shown (Aponina et al., 1982).

In the book by Alexander Bazykin "Nonlinear Dynamics of Interacting Populations", an exhaustive analysis of models of three interacting populations of the generalized Volterra model type is conducted (Bazykin, 2003). Discrete analogs of the continuous population dynamics models discussed above – Moran and Ricker models – consider population size as a discrete quantity changing at certain discrete moments in time, which corresponds to experimental data on census of real populations. If we assume that population size at time t (N_t, $t = 0, 1, 2, \ldots$) depends on sizes at some preceding moments in time, then to describe population dynamics one can apply the apparatus of recurrent or difference equations (mappings):

$$N_t = F(N_{t-1}, N_{t-2}, \ldots, N_{t-k}). \quad (15)$$

The solution of this equation is a sequence of values N_t satisfying equation (15) at each t. The Moran and Ricker model (Eq. (16)) was proposed to describe population dynamics of insects (Moran, 1950) and fish (Ricker, 1958):

$$N_t = N_{t-1} \exp\left\{r\left(1 - \frac{N_{t-1}}{K}\right)\right\}. \quad (16)$$

Interestingly, even in such a simple model, very different functioning modes are found (Fig. S6).

Another large direction in population modeling – population genetics modeling – was laid by classics of mathematical biology: Ronald Fisher, John Haldane, Sewall Wright, and others (Haldane, 1924, 1926, 1990; Fisher, 1930; Wright, 1931, 1949). Unlike population dynamics models describing changes in population sizes, population genetics models focus on describing changes in allele frequencies (gene variants) in populations. The mathematical apparatus used in classical population genetics models largely resembles that used in population dynamics models – these are either recurrent equations or ODEs. As an example, below is a model of a Mendelian asexual diploid panmictic population with one diallelic locus (a gene having only two states – A_1 or A_2):

$$\frac{dp}{dt} = \frac{sp(1-p)}{1-s(1-p)} \approx sp(1-p), \quad (17)$$

where p – frequency of allele A_1 in the population (then frequency of allele A_2, $q = 1 - p$), gamete fitness is defined as w_1, w_2, and their difference: $s = w_1 - w_2$. Equation (17) describes the change in allele frequencies over time.

Another class of models – matrix models of population structure dynamics with complex age-sex structure or populations including individuals of one species with differing physiological or biological characteristics, first proposed by Patrick H. Leslie (Leslie, 1945, 1948), were thoroughly investigated in studies (Gimelfarb et al., 1974; Logofet, Belova, 2008). The model is represented by an equation of the form:

$$x(t+1) = Lx(t), \quad t = 0, 1, 2, \ldots, \quad (18)$$

where column-vector $x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$ describes population structure (numbers of separate groups of individuals), and matrix L, also called the Leslie matrix, has the form:

$$L = \begin{pmatrix} b_1 & b_2 & \cdots & b_{n-1} & b_n \\ s_1 & 0 & & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \cdots & s_{n-1} & 0 \end{pmatrix},$$

where b_i – birth rate coefficients, s_i – survival coefficients.

Until recently, due to the lack of large-scale genomic data, such models were built primarily based on certain biologically meaningful assumptions, which allowed conducting theoretical research in this area only at a qualitative level. The development of sequencing methods and the subsequent emergence of large volumes of experimental data led to the appearance of computer models of population-genetic processes taking these data into account at a quantitative level. Simulation modeling of genetic sequence evolution received the name "coalescence simulation".

Methodologically, coalescence modeling represents a variety of stochastic modeling using various approaches (Monte Carlo methods, Markov chains, etc.) (Salem et al., 2005). Most works in this area are based on the development and modification of classical population genetics models, such as the

Wright–Fisher model (Hudson, 2002), island model (Wakeley, 2001), and others. A scheme of sequential complication of the population-genetic model by adding additional biological parameters to it was proposed (Schaffner et al., 2005) together with an algorithm for verifying values of these parameters. Currently, a number of software packages have been created for such modeling: SIMCOAL 2.0 (Laval, Excoffier, 2004), GENOME (Liang et al., 2007), Migrate-n (Beerli, Palczewski, 2010), CoaSim (Mailund et al., 2005), and others. Biologically significant results were obtained using these packages. In work (Bataillon et al., 2006), an assessment of the effective population size of Iceland, recombination rate, and a number of other population parameters was conducted. Testing new methods of analyzing genetic associations with human diseases using computer modeling is provided in (Guan et al., 2009). Statistical assessment of alternative human evolution scenarios using modeling was conducted in study (Fagundes et al., 2007).

Ecological/biocenotic level of organization

Ecology (from Ancient Greek *oikos* – "house" and *logos* – "study"), according to Ernst Haeckel – is the science of relationships of organisms and their populations with each other and with the habitat (Haeckel, 1866). Ecology studies *biocenoses* and *ecosystems* as a result of interdependent evolution of organisms (biota) and biocenological environment, taking into account activities of populations carried out at different trophic levels, determining the power of energy and substance flows in ecosystems and the general circulation of substances, as well as autoregulation of ecosystems and their role in the planet's biosphere (Bykov, 1983). Currently, the functioning of the planet's ecosystems depends on social factors and anthropogenic influences. In any limited space, usually many species inhabit, between which constant and complex relationships have been established. In other words, various types of organisms existing in a certain space with a complex of physicochemical conditions form a complex system, more or less persistently preserved in nature. In ecology, they are called *ecosystems* (Tansley, 1935) or *biogeocenoses* (Sukachev, 1972).

The term "*ecological modeling*" includes consideration of both models of interaction of individual organisms with the environment (*autecology*), interaction of population with the environment (*demecology*), and whole communities or biocenoses (*synecology*). Ecological models are based, first of all, on describing the transfer of substance, energy, and information between different parts of the ecosystem. Main attention is paid to how these parts interact, how they are connected to each other and influence each other, including the physical environment.

Dimensional units used in ecological modeling are usually the amount of energy or matter moving through the system. This is one of the main differences of ecological models from population ones, where measurement units are usually population size (Jørgensen, 2009).

The methodological arsenal of ecological modeling largely repeats methods used for describing molecular-genetic, cellular, organismal, and population levels discussed in previous sections of this work. In particular, ecological modeling uses the *flux balance analysis method* (FBA) (Allen, Gillooly, 2009; Orth et al., 2010). The main idea of the method lies in describing substance conversion flows as a linear programming task with constraints, which gives opportunities to estimate synthesis and degradation rates of these substances. An example of a schematic image of a model based on flux balance principles is provided in Figure S7.

ODE methods (Tskhai et al., 2001; Owolabi, Patidar, 2016; Lavaud et al., 2020) and systems of PDEs (Holmes et al., 1994; Tskhai et al., 2001), cellular automata (Gómez Esteban, Rodríguez-Patón, 2011), stochastic modeling (Kutalik et al., 2005; Phillips et al., 2006; Khatri et al., 2012), graph analysis (Fath et al., 2007), and other methods are also actively used in ecological modeling.

Ecological modeling is also one of the directions in which multiscale and multilevel/multilayer modeling received the strongest impulse for development (Grimm et al., 2005, 2010; Grimm, Berger, 2016).

A particular case of multiscale models are agent-based models, which in ecological modeling are traditionally called individual-based models. Being essentially simulation models, they cover a significant spectrum of ecosystem functioning questions both at the level of individual organisms and their populations and communities (Kreft et al., 1998; Doebeli, Dieckmann, 2003, 2004; Hellweger et al., 2016; Widder et al., 2016). In the individual-based paradigm, modern software packages have been developed, in particular, for modeling bacterial communities. For example, the simulator program BacSim (Kreft et al., 1998, 2001), as well as the software package developed in work (Xavier et al., 2005), describe such bacterial life processes as substrate uptake (transport), metabolism, cell division, and cell death, highlighting a separate cell as an object, considering communities as ensembles of such objects. They are oriented, first of all, towards studying bacterial communities in the form of biofilms. The program Micro-Gen Bacteria Simulator models the life cycle of a growing bacterial culture and its interaction with various molecules, for example, antibiotics (Murphy, Walshe, 2011). In these programs, ecological, metabolic, and population components are described in detail, but description of genetic processes and inheritance is absent.

Conclusion

Modern methods of modeling biological systems at different hierarchical levels of organization are based on both traditional approaches (differential, algebraic, and stochastic equations, graph theory, cellular automata, etc.) and hybrid techniques

combining object-oriented and agent-based (individual-based) approaches with the traditional ones mentioned above. Although collectively they cover practically all aspects of ecological (Jørgensen et al., 2009) and evolutionary (de Jong, 2002; Ferrer et al., 2008) processes, the combination of ecological and evolutionary components within one model still remains quite rare.

Comparing traditional approaches to building mathematical models with simulation modeling, we can see that in both cases certain limitations exist, significant for such a modeling object as a complexly organized biological system, such as, for example, a microbial community. In the first case, the static structure of the model acts as a limitation: the number of equations, variables, and model parameters does not change during the calculation. In the case of simulation modeling, the problem of model structure staticity is solved, as simulation models can contain a variable number of objects (e. g., individuals). However, simulation models of evolution and population dynamics are very demanding on RAM size and also have high computational complexity.

Based on the above, development of modeling methods for complex hierarchically organized biological systems, taking into account multiscale processes occurring in these systems, as well as taking into account their evolution – is an important task of modern mathematical biology. No less important is the task of developing software packages allowing effective solving of content-related biology tasks using mathematical and computer modeling.

It should be separately emphasized that the role of mathematical and computer models in biotechnology will only grow. Modeling biotechnological processes – from kinetics of enzymatic reactions and bioreactor operation to rational design of metabolic pathways and optimization of producer strains – allows not only reducing the cost and duration of experimental search but also purposefully forming the solution space. Such models serve as a basis for *in silico* screening of cultivation conditions and genetic modification constructs, supporting technological decision-making and scaling processes from laboratory to industrial levels. Inclusion of biotechnological applications into the contour of systemic modeling of complex biological systems appears to be an important direction for further development of interdisciplinary research at the intersection of mathematics, biology, and engineering sciences.

References

Akberdin I.R., Ozonov E.A., Mironova V.V., Omelyanchuk N.A., Likhoshvai V.A., Gorpinchenko D.N., Kolchanov N.A. A cellular automaton to model the development of primary shoot meristems of *Arabidopsis thaliana*. *J Bioinf Comput Biol.* 2007;05(02b):641-650. doi 10.1142/S0219720007002862

Akberdin I.R., Kazantsev F.V., Ermak T.V., Timonov V.S., Khlebodarova T.M., Likhoshvai V.A. *In silico* cell: challenges and perspectives. *Mathematical Biology and Bioinformatics*. 2013;8(1):295-315. doi 10.17537/2013.8.295 (in Russian)

Allen A.P., Gillooly J.F. Towards an integration of ecological stoichiometry and the metabolic theory of ecology to better understand nutrient cycling. *Ecol Lett*. 2009;12(5):369-384. doi 10.1111/j.1461-0248.2009.01302.x

Aponina E.A., Aponin Yu.M., Bazykin A.D. Analysis of complex dynamic behavior in the "predator – two prey" model. *Problems of Ecological Monitoring and Ecosystem Modeling*. 1982;5:163-180 (in Russian)

Barbuti R., Gori R., Milazzo P., Nasti L. A survey of gene regulatory networks modelling methods: from differential equations, to Boolean and qualitative bioinspired models. *J Membr Comput*. 2020;2(3): 207-226. doi 10.1007/s41965-020-00046-y

Barnabas J., Raj P. The human body: a digital twin of the cyber physical systems. *Adv Comput*. 2020;117(1):219-246. doi 10.1016/bs.adcom.2019.09.004

Bataillon T., Mailund T., Thorlacius S., Steingrimsson E., Rafnar T., Halldorsson M.M., Calian V., Schierup M.H. The effective size of the Icelandic population and the prospects for LD mapping: inference from unphased microsatellite markers. *Eur J Hum Genet.* 2006; 14(9):1044-1053. doi 10.1038/sj.ejhg.5201669

Bazykin A.D. Nonlinear Dynamics of Interacting Populations. Izhevsk, 2003 (in Russian)

Beerli P., Palczewski M. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*. 2010;185(1):313-326. doi 10.1534/genetics.109.112532

Berg H.C., Purcell E.M. Physics of chemoreception. *Biophys J.* 1977; 20(2):193-219. doi 10.1016/S0006-3495(77)85544-6

Bezzi M., Celada F., Ruffo S., Seiden P.E. The transition between immune and disease states in a cellular automaton model of clonal immune response. *Physica A Statist Mech Appl*. 1997;245(1-2):145-163. doi 10.1016/S0378-4371(97)00290-2

Blinov M.L., Faeder J.R., Goldstein B., Hlavacek W.S. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*. 2004;20(17): 3289-3291. doi 10.1093/bioinformatics/bth378

Bray D., Bourret R.B. Computer analysis of the binding reactions leading to a transmembrane receptor-linked multiprotein complex involved in bacterial chemotaxis. *Mol Biol Cell*. 1995;6(10):1367-1380. doi 10.1091/mbc.6.10.1367

Bray D., Bourret R.B., Simon M.I. Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Mol Biol Cell*. 1993;4(5):469-482. doi 10.1091/mbc.4.5.469

Bray D., Levin M.D., Lipkow K. The chemotactic behavior of computer-based surrogate bacteria. *Curr Biol*. 2007;17(1):12-19. doi 10.1016/j.cub.2006.11.027

Brüningk S.C., Ziegenhein P., Rivens I., Oelfke U., Haar G.T. A cellular automaton model for spheroid response to radiation and hyperthermia treatments. *Sci Rep*. 2019;9(1):17674. doi 10.1038/s41598-019-54117-x

Bykov B.A. Ecological Dictionary. Alma-Ata: Nauka Publ., 1983 (in Russian)

Chen L., Kulasiri D., Samarasinghe S. A novel data-driven Boolean model for genetic regulatory networks. *Front Physiol*. 2018;9:1328. doi 10.3389/fphys.2018.01328

Churaev R.N. Gene and epigene networks: two levels of organization of the hereditary system. *Informatsionnyy Vestnik VOGiS = The Herald of Vavilov Society for Geneticists and Breeders*. 2005;9(2):199-208 (in Russian)

Churaev R.N., Ratner V.A. Modeling of molecular-genetic control systems in the language of automata theory. Report 1. Operons and operon systems. In: Studies in Theoretical Genetics. Novosibirsk, 1972;210-228 (in Russian)

Colasanti R.L., Hunt R., Watrud L. A simple cellular automaton model for high-level vegetation dynamics. *Ecol Modell*. 2007;203(3-4): 363-374. doi 10.1016/j.ecolmodel.2006.12.039

de Jong H. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*. 2002;9(1):67-103. doi 10.1089/10665270252833208

Doebeli M., Dieckmann U. Speciation along environmental gradients. *Nature*. 2003;421(6920):259-264. doi 10.1038/nature01274

Doebeli M., Dieckmann U. Adaptive dynamics of speciation: spatial structure. In: Adaptive Speciation. Cambridge Studies in Adaptive Dynamics. Cambridge University Press, 2004;140-168

Drasdo D., Bode J., Dahmen U., Dirsch O., Dooley S., Gebhardt R., Ghallab A., … Klingmüller U., Kuepfer L., Timmer J., Zerial M., Hengstler J.G. The virtual liver: state of the art and future perspectives. *Arch Toxicol*. 2014;88(12):2071-2075. doi 10.1007/s00204-014-1384-6

Emonet T., Macal C.M., North M.J., Wickersham C.E., Cluzel P. AgentCell: a digital single-cell assay for bacterial chemotaxis. *Bioinformatics*. 2005;21(11):2714-2721. doi 10.1093/bioinformatics/bti391

Fadeev S.I., Likhoshvai V.A., Kogai V.V., Omel'yanchuk N.A. On mathematical modeling of auxin distribution patterns in plant root. *Siberian Electronic Mathematical Reports.* 2008;5:25-41 (in Russian)

Fagundes N.J.R., Ray N., Beaumont M., Neuenschwander S., Salzano F.M., Bonatto S.L., Excoffier L. Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci USA*. 2007;104(45):17614-17619. doi 10.1073/pnas.0708280104

Fath B.D., Scharler U.M., Ulanowicz R.E., Hannon B. Ecological network analysis: network construction. *Ecol Modell*. 2007;208(1):49-55. doi 10.1016/j.ecolmodel.2007.04.029

Ferrer J., Prats C., López D. Individual-based modelling: an essential tool for microbiology. *J Biol Phys*. 2008;34(1-2):19-37. doi 10.1007/s10867-008-9082-3

Fisher R.A. The Genetical Theory of Natural Selection. Clarendon Press, Oxford, 1930. Available: https://archive.org/details/geneticaltheoryo031631mbp/page/n25/mode/2up

Gevertz J.L., Torquato S. Modeling the effects of vasculature evolution on early brain tumor growth. *J Theor Biol*. 2006;243(4):517-531. doi 10.1016/j.jtbi.2006.07.002

Gilyarov M.S., Baev A.A., Vinberg G.G., Zavarzin G.A., Ivanov A.V., Severin S.E., Simolin A.V., Sokolov V.E., Tatarinov L.P., Takhtadzhian A.L., Yablokov A.V. (Eds) Biological Encyclopedic Dictionary. Moscow: Sovetskaia Entsiklopediia Publ., 1986. (in Russian)

Gimelfarb A.A., Ginzburg L.R., Poluektov R.A., Pykh Yu.A., Ratner V.A. Dynamic Theory of Biological Populations. Moscow: Nauka Publ., 1974 (in Russian)

Glazier J.A., Graner F. Simulation of the differential adhesion driven rearrangement of biological cells. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*. 1993;47(3):2128-2154. doi 10.1103/physreve.47.2128

Gómez Esteban P., Rodríguez-Patón A. Simulating a Rock-Scissors-Paper bacterial game with a discrete cellular automaton. In: Ferrández J.M., Álvarez Sánchez J.R., de la Paz F., Toledo F.J. (Eds) New Challenges on Bioinspired Applications. IWINAC 2011. Lecture Notes in Computer Science. Vol. 6687. Berlin: Springer, 2011;363-370. doi 10.1007/978-3-642-21326-7_39

Grimm V., Berger U. Structural realism, emergence, and predictions in next-generation ecological modelling: synthesis from a special issue. *Ecol Modell*. 2016;326:177-187. doi 10.1016/j.ecolmodel.2016.01. 001

Grimm V., Revilla E., Berger U., Jeltsch F., Mooij W.M., Railsback S.F., Thulke H.-H., Weiner J., Wiegand T., DeAngelis D.L. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science*. 2005;310(5750):987-991. doi 10.1126/science.1116681

Grimm V., Berger U., DeAngelis D.L., Polhill J.G., Giske J., Railsback S.F. The ODD protocol: a review and first update. *Ecol Modell*. 2010;221(23):2760-2768. doi 10.1016/j.ecolmodel.2010.08.019

Guan W., Liang L., Boehnke M., Abecasis G.R. Genotype-based matching to correct for population stratification in large-scale case-control genetic association studies. *Genetic Epidemiol*. 2009;33(6):598-517. doi 10.1002/gepi.20403

Haeckel E. Generelle Morphologie der Organismen. Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von C. Darwin reformirte Descendenz-Theorie. Berlin: Verlag Von Georg Reimer, 1866. Available: https://archive.org/details/generellemorphol01haec/page/n9/mode/2up

Haldane J.B.S. A mathematical theory of natural and artificial selection. Part II. The influence of partial self-fertilisation, inbreeding, assortative mating, and selective fertilisation on the composition of Mendelian populations, and on natural selection. *Biol Rev*. 1924;1(3):158-163. doi 10.1111/j.1469-185X.1924.tb00546.x

Haldane J.B.S. A mathematical theory of natural and artificial selection. *Math Proc Cambridge Philos Soc*. 1926;23(4):363-372. doi 10.1017/S0305004100015176

Haldane J.B.S. A mathematical theory of natural and artificial selection – I. 1924. *Bull Math Biol*. 1990;52(1-2):209-240. doi 10.1007/BF02459574

Harris L.A., Hogg J.S., Tapia J.J., Sekar J.A., Gupta S., Korsunsky I., Arora A., Barua D., Sheehan R.P., Faeder J.R. BioNetGen 2.2: advances in rule-based modeling. *Bioinformatics*. 2016;32(21):3366-3368. doi 10.1093/bioinformatics/btw469

Hellweger F.L., Clegg R.J., Clark J.R., Plugge C.M., Kreft J.-U. Advancing microbial sciences by individual-based modelling. *Nat Rev Microbiol*. 2016;14(7):461-471. doi 10.1038/nrmicro.2016.62

Hoehme S., Drasdo D. A cell-based simulation software for multi-cellular systems. *Bioinformatics*. 2010;26(20):2641-2642. doi 10.1093/bioinformatics/btq437

Holmes E.E., Lewis M.A., Banks J.E., Veit R.R. Partial differential equations in ecology: spatial interactions and population dynamics. *Ecology*. 1994;75(1):17-29. doi 10.2307/1939378

Holzhütter H.-G., Drasdo D., Preusser T., Lippert J., Henney A.M. The virtual liver: a multidisciplinary, multilevel challenge for systems biology. *Wiley Interdiscip Rev Syst Biol Med.* 2012;4(3):221-235. doi 10.1002/wsbm.1158

Hucka M., Finney A., Sauro H.M., Bolouri H., Doyle J.C., Kitano H., Arkin A.P., … Stelling J., Takahashi K., Tomita M., Wagner J., Wang J. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*. 2003;19(4):524-531. doi 10.1093/bioinformatics/btg015

Hudson R.R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*. 2002;18(2):337-338. doi 10.1093/bioinformatics/18.2.337

Ishii N., Robert M., Nakayama Y., Kanai A., Tomita M. Toward large-scale modeling of the microbial cell for computer simulation. *J Biotechnol*. 2004;113(1-3):281-94. doi 10.1016/j.jbiotec.2004.04.038

Jørgensen S.E. (Ed.) Ecosystem Ecology. Elsevier, 2009

Jørgensen S.E., Chon T.-S., Recknagel F.A. (Eds) Handbook of Ecological Modelling and Informatics. WIT Press, 2009

Kaizu K., de Ronde W., Paijmans J., Takahashi K., Tostevin F., ten Wolde P.R. The Berg–Purcell limit revisited. *Biophys J*. 2014;106(4):976-985. doi 10.1016/j.bpj.2013.12.030

Karev G.P., Skomorovsky Yu.I. Simulation on the time course of the state of one-species tree stands. *Sibirskii Ekologicheskii Zhurnal = Siberian Journal of Ecology.* 1999;6(4):403-417 (in Russian)

Karr J.R., Sanghvi J.C., MacKlin D.N., Gutschow M.V., Jacobs J.M., Bolival B., Assad-Garcia N., Glass J.I., Covert M.W. A whole-cell computational model predicts phenotype from genotype. *Cell*. 2012;150(2):389-401. doi 10.1016/j.cell.2012.05.044

Kauffman S.A. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol*. 1969;22(3):437-467. doi 10.1016/0022-5193(69)90015-0

Kazantsev F.V., Akberdin I.R., Bezmaternykh K.D., Likhoshvai V.A. The tool for automatic generation of gene networks mathematical models. *Informatsionnyy Vestnik VOGiS = The Herald of Vavilov Society for Geneticists and Breeders*. 2009;13(1):163-169 (in Russian)

Kazantsev F., Akberdin I., Lashin S., Ree N., Timonov V., Ratushnyi A., Khlebodarova T., Likhoshvai V. MAMMOTh: a new database for curated mathematical models of biomolecular systems. *J Bioinf Comput Biol*. 2018;16(01):1740010. doi 10.1142/S0219720017400108

Khatri B.S., Free A., Allen R.J. Oscillating microbial dynamics driven by small populations, limited nutrient supply and high death rates. *J Theor Biol*. 2012;314:120-129. doi 10.1016/j.jtbi.2012.08.013

Kolchanov N.A., Ignat'eva E.V., Podkolodnaia O.A., Likhoshvai V.A., Matushkin Yu.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2013;17(4/2):833-850 (in Russian)

Kolobov A.N., Frisman E.Ya. Modelling of dynamics of the self-organizing processes in spatially distributed plant communities. *Mathematical Biology and Bioinformatics*. 2008;3(2):85-102 (in Russian)

Komarov A.S., Palenova M.M., Smirnova O.V. The concept of discrete description of plant ontogenesis and cellular automata models of plant populations. *Ecol Modell*. 2003;170(2-3):427-439. doi 10.1016/S0304-3800(03)00243-6

Kreft J., Booth G., Wimpenny J.W.T. BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology*. 1998;144(12):3275-3287. doi 10.1099/00221287-144-12-3275

Kreft J.U., Picioreanu C., Wimpenny J.W., Van Loosdrecht M.C. Individual-based modelling of biofilms. *Microbiology*. 2001;147(11):2897-2912. doi 10.1099/00221287-147-11-2897

Kutalik Z., Razaz M., Baranyi J. Connection between stochastic and deterministic modelling of microbial growth. *J Theor Biol*. 2005;232(2):285-299. doi 10.1016/j.jtbi.2004.08.013

Laval G., Excoffier L. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*. 2004;20(15):2485-2487. doi 10.1093/bioinformatics/bth264

Lavaud R., Filgueira R., Nadeau A., Steeves L., Guyondet T. A Dynamic Energy Budget model for the macroalga *Ulva lactuca*. *Ecol Modell*. 2020;418:108922. doi 10.1016/j.ecolmodel.2019.108922

Leslie P.H. On the use of matrices in certain population mathematics. *Biometrika*. 1945;33(3):183-212. doi 10.1093/biomet/33.3.183

Leslie P.H. Some further notes on the use of matrices in population mathematics. *Biometrika*. 1948;35(3/4):213. doi 10.2307/2332342

Liang L., Zöllner S., Abecasis G.R. GENOME: a rapid coalescent-based whole genome simulator. *Bioinformatics*. 2007;23(12):1565-1567. doi 10.1093/bioinformatics/btm138

Likhoshvai V.A. Mathematical Modeling and Computational Analysis of Gene Networks. Dr. Biol. Sci. Diss. Novosibirsk, 2008 (in Russian)

Likhoshvai V.A., Matushkin Yu.G., Vatolin Yu.N., Bazhan S.I. A generalized chemical-kinetic method for modeling complex biological systems. Computer model of bacteriophage Lambda ontogenesis. *Journal of Computational Technologies*. 2000;5(Special issue):87-99 (in Russian)

Likhoshvai V.A., Omel'yanchuk N.A., Mironova V.V., Fadeev S.I., Mjolsness E.D., Kolchanov N.A. Mathematical model of auxin distribution in the plant root. *Russ J Dev Biol*. 2007;38(6):374-382. doi 10.1134/S1062360407060057

Logofet D.O., Belova I.N. Nonnegative matrices as a tool to model population dynamics: classical models and contemporary expansions. *J Math Sci.* 2008;155(6):894-907. doi 10.1007/s10958-008-9249-2

Lotka A.J. Contribution to the theory of periodic reactions. *J Phys Chem*. 1909;14(3):271-274. doi 10.1021/j150111a004

Lotka A.J. Undamped oscillations derived from the law of mass action. *J Am Chem Soc*. 1920;42(8):1595-1599. doi 10.1021/ja01453a010

Lotka A.J. Elements of Physical Biology. Williams and Wilkins, Baltimore, 1925

Mailund T., Schierup M.H., Pedersen C.N.S., Mechlenborg P.J.M., Madsen J.N., Schauser L. CoaSim: a flexible environment for simulating genetic data under coalescent models. *BMC Bioinformatics*. 2005;6(1):252. doi 10.1186/1471-2105-6-252

Malthus T.R. An Essay on the Principle of Population. Cambridge University Press, 2008

Marée A.F.M., Grieneisen V.A., Hogeweg P. The Cellular Potts Model and biophysical properties of cells, tissues and morphogenesis. In: Anderson A.R.A., Chaplain M.A.J., Rejniak K.A. (Eds) Single-Cell-Based Models in Biology and Medicine. Mathematics and Biosciences in Interaction. Birkhäuser, Basel, 2007;107-136. doi 10.1007/978-3-7643-8123-3_5

Markus M., Böhm D., Schmick M. Simulation of vessel morphogenesis using cellular automata. *Math Biosci*. 1999;156(1-2):191-206. doi 10.1016/S0025-5564(98)10066-4

Miehe R., Horbelt J., Baumgarten Y., Bauernhansl T. Reprint of: Basic considerations for a digital twin of biointelligent systems: applying technical design patterns to biological systems. *CIRP J Manuf Sci Technol*. 2021;34:133-145. doi 10.1016/j.cirpj.2021.06.004

Mironova V.V., Omelyanchuk N.A., Yosiphon G., Fadeev S.I., Kolchanov N.A., Mjolsness E., Likhoshvai V.A. A plausible mechanism for auxin patterning along the developing root. *BMC Syst Biol*. 2010;4(3):98. doi 10.1186/1752-0509-4-98

Mironova V.V., Omelyanchuk N.A., Novoselova E.S., Doroshkov A.V., Kazantsev F.V., Kochetov A.V., Kolchanov N.A., Mjolsness E., Likhoshvai V.A. Combined *in silico*/*in vivo* analysis of mechanisms providing for root apical meristem self-organization and maintenance. *Ann Bot*. 2012;110(2):349-360. doi 10.1093/aob/mcs069

Monod J., Wyman J., Changeux J.-P. On the nature of allosteric transitions: a plausible model. *J Mol Biol*. 1965;12(1):88-118. doi 10.1016/S0022-2836(65)80285-6

Moran P.A.P. Some remarks on animal population dynamics. *Biometrics*. 1950;6(3):250-258. doi 10.2307/3001822

Murphy J.T., Walshe R. Modeling antibiotic resistance in bacterial colonies using agent-based approach. In: Dubitzky W., Southgate J., Fuß H. (Eds) Understanding the Dynamics of Biological Systems. Springer, New York, 2011;131-154. doi 10.1007/978-1-4419-7964-3_7

Murray J.D. (Ed.) Mathematical Biology. II. Spatial Models and Biomedical Applications. New York: Springer, 2003. doi 10.1007/b98869

Novoselova E.S., Mironova V.V., Omelyanchuk N.A., Kazantsev F.V., Likhoshvai V.A. Mathematical modeling of auxin transport in protoxylem and protophloem of *Arabidopsis thaliana* root tips. *J Bioinf Comput Biol*. 2013;11(1):1340010. doi 10.1142/S0219720013400106

O'Beirne G.A., Patuzzi R.B. Mathematical model of outer hair cell regulation including ion transport and cell motility. *Hear Res*. 2007;234(1-2):29-51. doi 10.1016/j.heares.2007.09.008

Odum E. Fundamentals of Ecology. Moscow: Mir Publ., 1975 (in Russian)

Omholt S.W., Gjuvsland A.B., Vik J.O. The Digital Salmon – key to both more science and more profit. In: Grue J., Almås K.A. (Eds) Food from the Ocean – Norway's Opportunities. Oslo, Norway, 2013;81-112

Orth J.D., Thiele I., Palsson B.Ø. What is flux balance analysis? *Nate Biotech*. 2010;28(3):245-248. doi 10.1038/nbt.1614

Owolabi K.M., Patidar K.C. Numerical simulations of multicomponent ecological models with adaptive methods. *Theor Biol Med Modell*. 2016;13(1):1. doi 10.1186/s12976-016-0027-4

Pałubicki W., Kokosza A., Burian A. Formal description of plant morphogenesis. *J Exp Bot*. 2019;70(14):3601-3613. doi 10.1093/jxb/erz210

Palyanov A.Yu. Methods and Algorithms for Solving a Number of Current Problems in Computational Neurobiology, Biomechanics and Molecular Biology. Novosibirsk, 2019 (in Russian)

Pasternak T., Groot E.P., Kazantsev F.V., Teale W., Omelyanchuk N., Kovrizhnykh V., Palme K., Mironova V.V. Salicylic acid affects root meristem patterning via auxin distribution in a concentration-dependent manner. *Plant Physiol*. 2019;180(3):1725-1739. doi 10.1104/pp.19.00130

Phillips S.J., Anderson R.P., Schapire R.E. Maximum entropy modeling of species geographic distributions. *Ecol Modell*. 2006;190(3-4):231-259. doi 10.1016/j.ecolmodel.2005.03.026

Poletaev I.A. On some models of biogeocenoses. In: Application of Mathematical Methods in Biology. Leningrad, 1965 (in Russian)

Poletaev I.A. On mathematical models of elementary processes in biocenoses. In: Problemy Kibernetiki. Iss. 16. Moscow: Nauka Publ., 1966;171-190 (in Russian)

Price N.D., Reed J.L., Palsson B.Ø. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol*. 2004;2(11):886-897. doi 10.1038/nrmicro1023

Ratner V.A., Zharkikh A.A., Kolchanov N.A., Rodin S.N., Solov'ev V.V., Shamin V.V. Problems of the Theory of Molecular Evolution. Novosibirsk: Nauka Publ., 1985 (in Russian)

Ratushnyi A.V., Likhoshvai V.A., Anan'ko E.A., Vladimirov N.V., Gunbin K.V., Lashin S.A., Nedosekina E.A., Nikolaev S.V., Omel'yanchuk L.V., Matushkin Yu.G., Kolchanov N.A. The Novosibirsk school of systems computational biology: historical review and future prospects. *Informatsionnyy Vestnik VOGiS = The Herald of Vavilov Society for Geneticists and Breeders*. 2005;9(2):232-261 (in Russian)

Ricker W.E. Handbook of Computations of Biological Statistics of Fish Populations. Ottawa: Queen's Printer and Controller of Stationery, 1958. Available: https://waves-vagues.dfo-mpo.gc.ca/library-bibliotheque/10161.pdf

Riznichenko G.Yu. Lectures on Mathematical Models in Biology. Moscow, Izhevsk, 2002 (in Russian)

Riznichenko G.Yu. Mathematical Models in Biophysics and Ecology. Moscow, Izhevsk, 2003 (in Russian)

Riznichenko G.Yu. Mathematical Modeling of Biological Processes: Models in Biophysics and Ecology. Moscow: Urait Publ., 2017 (in Russian)

Salem R.M., Wessel J., Schork N.J. A comprehensive literature review of haplotyping software and methods for use with unrelated individuals. *Hum Genom*. 2005;2(1):39. doi 10.1186/1479-7364-2-1-39

Salguero A.G., Capel M.I., Tomeu A.J. Parallel cellular automaton tumor growth model. In: Practical Applications of Computational Biology and Bioinformatics, 12th International Conference. PACBB-2018. Advances in Intelligent Systems and Computing. Vol. 803. Springer, 2019;175-182. doi 10.1007/978-3-319-98702-6_21

Sarma G.P., Lee C.W., Portegys T., Ghayoomie V., Jacobs T., Alicea B., Cantarelli M., … Khayrulin S., Lung D., Palyanov A., Watts M., Larson S.D. OpenWorm: overview and recent advances in integrative biological simulation of *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci*. 2018;373(1758):20170382. doi 10.1098/rstb.2017.0382

Schaffner S.F., Foo C., Gabriel S., Reich D., Daly M.J., Altshuler D. Calibrating a coalescent simulation of human genome sequence variation. *Genome Res*. 2005;15(11):1576-1583. doi 10.1101/gr.3709305

Shumnyi V.K., Shokin Yu.I., Kolchanov N.A., Fedotov A.M. (Eds) Biodiversity and Ecosystem Dynamics: Information Technologies and Modeling. Novosibirsk, 2006 (in Russian)

Slimi R., El Yacoubi S., Dumonteil E., Gourbière S. A cellular automata model for Chagas disease. *Appl Math Modell*. 2009;33(2):1072-1085. doi 10.1016/j.apm.2007.12.028

Sukachev V.N. Selected Works: in 3 vols. Leningrad: Nauka Publ., 1972 (in Russian)

Sütterlin T., Kolb C., Dickhaus H., Jäger D., Grabe N. Bridging the scales: semantic integration of quantitative SBML in graphical multi-cellular models and simulations with EPISIM and COPASI. *Bioinformatics*. 2013;29(2):223-229. doi 10.1093/bioinformatics/bts659

Sütterlin T., Tsingos E., Bensaci J., Stamatas G.N., Grabe N. A 3D self-organizing multicellular epidermis model of barrier formation and hydration with realistic cell morphology based on EPISIM. *Sci Rep*. 2017;7(1):43472. doi 10.1038/srep43472

Swat M.H., Thomas G.L., Belmonte J.M., Shirinifard A., Hmeljak D., Glazier J.A. Multi-scale modeling of tissues using CompuCell3D. *Methods Cell Biol.* 2012;110:325-366. doi 10.1016/B978-0-12-388403-9.00013-8

Szabó A., Merks R.M.H. Cellular potts modeling of tumor growth, tumor invasion, and tumor evolution. *Front Oncol*. 2013;3:87. doi 10.3389/fonc.2013.00087

Szigeti B., Gleeson P., Vella M., Khayrulin S., Palyanov A., Hokanson J., Currie M., Cantarelli M., Idili G., Larson S. OpenWorm: an open-science approach to modeling *Caenorhabditis elegans*. *Front Comput Neurosci*. 2014;8:137. doi 10.3389/fncom.2014.00137

Talaminos-Barroso A., Reina-Tosina J., Roa-Romero L.M. Models based on cellular automata for the analysis of biomedical systems. In: Control Applications for Biomedical Engineering Systems. Elsevier, 2020;405-445. doi 10.1016/B978-0-12-817461-6.00014-7

Tansley A.G. The use and abuse of vegetational concepts and terms. *Ecology*. 1935;16(3):284-307. doi 10.2307/1930070

Tellechea-Luzardo J., Winterhalter C., Widera P., Kozyra J., de Lorenzo V., Krasnogor N. Linking engineered cells to their digital twins: a version control system for strain engineering. *ACS Synth Biol*. 2020;9(3):536-545. doi 10.1021/acssynbio.9b00400

Tomita M. Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol*. 2001;19(6):205-210. doi 10.1016/S0167-7799(01)01636-5

Tomita M., Hashimoto K., Takahashi K., Shimizu T., Matsuzaki Y., Miyoshi F., Saito K., Tanida S., Yugi K., Venter J., Hutchison C. E-CELL: software environment for whole-cell simulation. *Bioinformatics*. 1999;15(1):72-84. doi 10.1093/bioinformatics/15.1.72

Tran Q.N. Finding long-term influence and sensitivity of genes using probabilistic genetic regulatory networks. In: Emerging Trends in Applications and Infrastructures for Computational Biology, Bioinformatics, and Systems Biology. Morgan Kaufmann, 2016;107-120. doi 10.1016/B978-0-12-804203-8.00008-0

Tskhai A.A., Pulian M., Beldeeva L.N., Ganulis J., Zherelina I.V., Kvon V.I., Kirillov V.V., Liakhova S.A., Nakhtnebel H.-P. Introduction to Ecological Modeling. Barnaul: Azbuka Publ., 2001 (in Russian)

Turing A.M. The chemical theory of morphogenesis. *Philos Trans R Soc Lond B Biol Sci.* 1952;237(641):37-72. doi 10.1098/rstb.1952.0012

Verhulst P.F. Notice sur la loi que population suit dans son accroissement. *Correspondance Mathématique et Physique*. 1838;10:113-121

Vitvitsky A.A. Cellular automata with dynamic structure to simulate the growth of biological tissues. *Numer Analys Appl*. 2014;7(4):263-273. doi 10.1134/S1995423914040016

Vladimirov N., Sourjik V. Chemotaxis: how bacteria use memory. *Biol Chem*. 2009;390(11):1097-1104. doi 10.1515/BC.2009.130

Vladimirov N., Løvdok L., Lebiedz D., Sourjik V. Dependence of bacterial chemotaxis on gradient shape and adaptation rate. *PLoS Comp Biol*. 2008;4(12):e1000242. doi 10.1371/journal.pcbi.1000242

Volterra V. Mathematical theory of the struggle for existence. *Uspekhi Fizicheskikh Nauk*. 1928;8(1):13-34. doi 10.3367/UFNr.0008.192801b.0013 (in Russian)

Volterra V. Mathematical Theory of the Struggle for Existence. Moscow: Nauka Publ., 1976 (in Russian)

Voss-Böhme A. Multi-scale modeling in morphogenesis: a critical analysis of the cellular Potts model. *PLoS One*. 2012;7(9):e42852. doi 10.1371/journal.pone.0042852

Wakeley J. The coalescent in an island model of population subdivision with variation among demes. *Theor Popul Biol*. 2001;59(2):133-144. doi 10.1006/tpbi.2000.1495

Widder S., Allen R.J., Pfeiffer T., Curtis T.P., Wiuf C., Sloan W.T., Cordero O.X., … Klapper I., Labarthe S., Smets B.F., Wang H., Soyer O.S. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J*. 2016;10(11):2557-2568. doi 10.1038/ismej.2016.45

Wright S. Evolution in Mendelian populations. *Genetics*. 1931;16(2):97-159. doi 10.1093/genetics/16.2.97

Wright S. The genetical structure of populations. *Ann Eugen*. 1951;15(4):323-354. doi 10.1111/j.1469-1809.1949.tb02451.x

Xavier J.B., Picioreanu C., van Loosdrecht M.C.M. A framework for multidimensional modelling of activity and structure of multispecies biofilms. *Environ Microbiol*. 2005;7(8):1085-1103. doi 10.1111/j.1462-2920.2005.00787.x

Zhabotinskii A.M. Concentration Self-oscillations. Moscow: Nauka Publ., 1974 (in Russian)

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

Received November 28, 2025. Revised February 4, 2026. Accepted February 4, 2026.

BIOINFORMATICS AND SYSTEMS BIOLOGY
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):515-526

doi 10.18699/vjgb-26-54

Evolutionary inferences from the analysis of mutation dynamics in the SARS-CoV-2 replication-transcription complex

A.Yu. Palyanov [1, 2] ✉, A.P. Devyaterikov[1], N.V. Palyanova [2], A.M. Shestopalov [2]

[1] A.P. Ershov Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Research Institute of Virology, Federal Research Center of Fundamental and Translational Medicine, Novosibirsk, Russia
✉ palyanov@iis.nsk.su

Abstract. The SARS-CoV-2 virus continues to evolve and remains a significant public health threat, while the worldwide monitoring and sequencing of its genomic variants provide a unique opportunity to study its evolution and better understand its molecular mechanisms. In our work, we analyze its replication-transcription complex (RTC) over a 5.5-year period (December 2019–July 2025). This complex is significantly more conserved (as any alteration impairing its function prevents viral replication) than the S-protein (directly impacting infectivity and immune evasion) but still dynamically evolving part of the genome. The study focuses on high-frequency substitutions, their temporal behavior, co-occurrence, and structural context. Using genomes from GISAID, we identified 22 amino acid point mutations present in at least 1 % of currently available sequences, analyzed their weekly dynamics, revealed three distinct temporal patterns, and enumerated frequent co-occurring groups (pairs, triplets, and larger sets) within the same genomes. We mapped the affected residues onto an RTC 3D structure and reviewed the literature to examine the reported functional consequences. Notably, all these substitutions were single-nucleotide. One of the mutations, nsp12:G671S, showed a unique dynamic feature: it emerged, dominated globally for months, disappeared twice, and in 2025 reappeared for the 3rd time, always accompanied with other mutations in the RTC. Thus, it was interesting to trace its dynamics as an indicator of probable changes. In addition, our analysis of mutation and variant timelines suggests that the Delta variant may have emerged 7–8 months earlier than commonly reported. Taken together, these results provide a consolidated view of recurrent RTC variation, its temporal classes, co-occurrence, and structural context, underscoring the value of systematic surveillance of nsp7–nsp14 alongside analyses focused on structural proteins.
Key words: SARS-CoV-2; RdRp; RTC; evolution; substitutions; mutations; dynamics; analysis; nsp12:G671S; Delta

For citation: Palyanov A.Yu., Devyaterikov A.P., Palyanova N.V., Shestopalov A.M. Evolutionary inferences from the analysis of mutation dynamics in the SARS-CoV-2 replication-transcription complex. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):515-526. doi 10.18699/vjgb-26-54

Funding. This research was funded by the Russian Science Foundation, project 23-64-00005.

Эволюционные аспекты динамики мутаций в репликационно-транскрипционном комплексе SARS-CoV-2

А.Ю. Пальянов [1, 2] ✉, А.П. Девятериков[1], Н.В. Пальянова [2], А.М. Шестопалов [2]

[1] Институт систем информатики им. А.П. Ершова Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Научно-исследовательский институт вирусологии, Федеральный исследовательский центр фундаментальной и трансляционной медицины, Новосибирск, Россия
✉ palyanov@iis.nsk.su

Аннотация. Вирус SARS-CoV-2 остается заметной угрозой для человечества, поскольку продолжает циркулировать и эволюционировать. При этом продолжающийся глобальный мониторинг и накопление геномных данных позволяют детально изучать эволюционные механизмы. В нашей работе мы анализируем репликационно-транскрипционный комплекс (RTC, nsp7–nsp14) – более консервативную, чем S-белок, напрямую связанный с инфекционностью и избеганием иммунитета, но все-таки динамично эволюционирующую часть генома – в течение 5.5-летнего периода (декабрь 2019 г. – июль 2025 г.). Исследование сосредоточено на аминокислотных заменах, присутствующих по крайней мере в 1 % доступных в базе GISAID геномов, их временно́й динамике, совместной встречаемости и структурном контексте. Мы идентифицировали 22 таких точечных аминокислотных замены, проанализировали их недельную динамику, выявили три различных временны́х паттерна и рассчитали частоты для групп мутаций (пары, тройки и т. д.), одновременно присутствующих в геномах. Примечательно, что все изученные замены были однонуклеотидными. Мы визуализировали аминокислотные остатки, соответствующие рассмотренным мутациям, на трехмерной структуре RTC, описали их особенности и обобщили данные литературы для изучения известных функциональных последствий. Одна из мутаций, nsp12:G671S, продемонстрировала уникальную динамику: она появлялась, доминировала в

глобальном масштабе в течение нескольких месяцев, затем исчезала дважды, а в 2025 г. появилась в третий раз. При этом она всегда сопровождалась набором сопутствующих мутаций в комплексе RTC, что делает ее потенциальным индикатором изменений в геноме, т.е. необходимо продолжить отслеживание ее динамики. Кроме того, наш анализ временны́х линий мутаций и вариантов позволяет предположить, что вариант Дельта мог появиться на 7–8 месяцев раньше, чем принято считать. В совокупности эти результаты дают целостное представление о повторяющихся вариациях в RTC, их временны́х классах и совместной встречаемости, а также о структурном контексте, подчеркивая ценность систематического мониторинга nsp7–nsp14 наряду с анализом, сфокусированным на структурных белках.

Ключевые слова: SARS-CoV-2; RdRp; RTC; эволюция; замены; мутации; динамика; анализ; nsp12:G671S; Дельта

Introduction

The COVID-19 pandemic catalyzed a sustained global effort in next-generation sequencing, virology, and bioinformatics, yielding an unprecedented corpus of SARS-CoV-2 genomes over the past five years. As the virus continues to evolve and remains a major public health concern, continuous genomic surveillance and further investigation of its molecular mechanisms are of considerable importance. The 29.9-kb positive-sense RNA genome encodes structural (S, E, M, N), nonstructural (nsp1–nsp16), and nine small accessory proteins (Bai et al., 2022; Yan W. et al., 2022), with approximately two-thirds of the coding capacity assigned to the nonstructural machinery on the 5′ end (Eriani, Martin, 2022). A similar genomic structure, in terms of both gene composition and gene order, is observed across multiple related coronaviruses, including those hosted by humans, bats, and pangolins (Brant et al., 2021, Fig. 1; Temmam et al., 2022, Fig. 1). Approximately two-thirds of the genome on the 5′ part codes for non-structural proteins and one-third of the genome on the 3′ part codes for structural and accessory proteins (Eriani, Martin, 2022).

Mutations in the S-protein gene (~13 % of the 29.9 kb genome) attract the greatest attention due to their direct impact on infectivity and immune evasion. However, mutations in other structural and nonstructural proteins, as well as recombination events and other genomic alterations, can also lead to diverse consequences that affect virus adaptability. SARS-CoV-2 exhibits a complex interplay between antigenicity, transmissibility, and virulence, which usually has unpredictable consequences for the future trajectory of evolution of the virus (Carabelli et al., 2023). In the present study, we continue our investigation of SARS-CoV-2 evolution, using genomic sequences that have been collected and deposited in public databases since the virus first emerged in late 2019. In our previous works, we analyzed SARS-CoV-2 evolution at several scales: globally (Palyanov, Palyanova, 2024), regionally across Siberia (Palyanova et al., 2023), and within the genomic variant landscape of the virus (Palyanov, Palyanova, 2023). Here, we focus on a mutational analysis of the replication-transcription complex, which carries out several essential viral functions, including genome replication and proofreading, and constitutes a substantial portion of the viral genome (Romano et al., 2020).

Among nonstructural proteins, the replication-transcription complex (RTC) is central to RNA synthesis and error control. The nsp12 RNA-dependent RNA polymerase (RdRp) operates with cofactors nsp7 and nsp8. Together they form a subcomplex (holoenzyme), which is the minimal core component for mediating coronavirus RNA synthesis (Peng et al., 2020; Singh et al., 2025). In this work we analyze an extended, structurally characterized RTC assembly spanning nsp7–nsp14 (PDB: 7EIZ) (Yan L. et al., 2021), which enables assessment of substitutions not only within RdRp but also across its binding interfaces and the exonuclease proofreading module. The analysis of this extended configuration may provide additional insights into the structural and functional consequences of mutations, particularly in cases where multiple substitutions occur in close spatial proximity. For example, point mutations at specific sites of the nsp12–nsp8 interface dramatically affect the RNA polymerization activity of SARS-CoV-2 (Ferrer-Orta et al., 2024). In the same year 2024, it was discovered that the combination of nsp8:A21V and nsp12:P323L mutations resulted in an approximately 50 % increase in polymerase activity (Danda et al., 2024). The authors described this as the first biochemical study demonstrating the functional impact of amino acid substitutions involving all components of the RdRp complex in emerging SARS-CoV-2 subvariants.

The evolution of SARS-CoV-2 proceeds through a mixture of point mutations, insertions, deletions, and recombinations, and its consequences are highly diverse and difficult to predict. For instance, mutations in nonstructural proteins such as nsp6:ΔSGF(3675–3677) (Feng et al., 2023) and nsp12:P323L/G671S (Kim et al., 2023) have been shown to increase viral replication efficiency. Next, nsp1:(Y154A/F157A) and nsp1:(R171E/R175E) mutations have been found to abolish protein translation inhibition in a cell free system (Schubert et al., 2020). The single mutation nsp14:F60S in the exoribonuclease (ExoN), responsible for replication proofreading, accelerates viral evolution by increasing the mutation rate (Mack et al., 2023). Moreover, the evolutionary rate of SARS-CoV-2 varies considerably among variants, and the emergence of new lineages often coincides with episodic accelerations of this rate. In certain periods, the evolutionary rate increased up to fourfold relative to the background phylogenetic rate, leading to the emergence of new variants within weeks rather than months, as would be expected from the baseline tempo of viral evolution (Tay et al., 2022).

While early global surveys mapped common substitutions across the genome (Abbasian et al., 2023), a focused and time-resolved analysis of RTC variation at population scale is still warranted given its functional importance. Genes encoding RTC are among the most conserved in viral genomes, as malfunction of the replication machinery prevents the production of viable virions. Nevertheless, SARS-CoV-2 (2019) exhibits four to eight point amino acid substitutions compared to its closest relatives (e. g., BANAL-20-52 (Temmam et al., 2022; Ou et al., 2023) and RaTG13 (Rahalkar, Bahulikar, 2020;

Zhou et al., 2020)), a range similar to the divergence observed between the RTC genes of the ancestral SARS-CoV-2 and of a typical 2023 Omicron descendant (Table S1)[1]. The RNA-dependent RNA polymerases (RdRps) of viruses and cellular organisms share a conserved structural core, most notably the canonical palm domain that forms the catalytic center of the polymerase, underscoring the universality of the RNA synthesis mechanism. However, they diverge significantly in their structural details and a variety of accessory proteins – so, there is scope for variation, and the virus exploits it during evolution. Therefore, it is crucial to monitor the rate and consequences of these changes to identify emerging, functionally significant RTC substitutions that rise above the background noise in global surveillance data.

In this study, we aimed to identify RTC substitutions that eventually attain population-level peaks and to characterize their post-peak behaviors, including distinct cycles of re-emergence and decline or extinction, while also assessing co-occurrence within the extended complex. Using all publicly available genomes from late 2019 to July 2025, we quantified amino acid substitutions in nsp7–nsp14 occurring in ≥1 % of sequences; profiled their weekly and monthly dynamics to classify persistent, transient, and recurrent patterns; and analyzed co-occurrence. We then interpreted these population-scale signals in structural and functional terms by centering the analysis on the structurally resolved RTC (nsp7–nsp14; PDB 7EIZ), with the broader genome-wide context provided by prior surveys (Abbasian et al., 2023).

Materials and methods

We performed a secondary analysis of SARS-CoV-2 genomic surveillance data in the public domain to quantify the amino acid (AA) substitution dynamics in the proteins of the replication-transcription complex (RTC: nsp7/8/9/10/12/13/14) from the start of the pandemic through July 15, 2025. The primary outcome was the weekly global fraction of genomes carrying a given AA substitution; secondary outcomes included (i) geographic stratification by continent, (ii) co-occurrence patterns among substitutions within RTC genes, and (iii) lineage/clade context for substitutions with notable temporal behavior. The pre-specified gene set and date horizon are described in the main text and the Results sections.

Data sources

1. GISAID (https://gisaid.org) (Khare et al., 2021): the source of genome sequences and metadata used for lineage context and targeted queries by mutation, date, and location. All GISAID queries used the web interface (EpiCov→Search) with the filters *complete* + *low coverage excluded* + *collection date complete* for reliability, as detailed in the Results and Materials sections of the manuscript. It retains 15.5 of 17.5 million of genome samples, whereas the stronger variant, *high coverage* instead of *low coverage excluded* (entries with <1 % of Ns vs entries with <5 % of Ns), keeps only 5.7 million genomes. Text input fields for *collection* (*from*) and *collection* (*to*) dates were used to specify the necessary time interval, *AA Substitutions*, to input single or multiple mutations in proteins (for example, nsp9_T35I, nsp12_G671S, nsp13_S36P), and *Nucl. Mutations*, to input single or multiple mutations in genome nucleotides.
2. CovidCG (https://covidcg.org) (Elbe, Buckland-Merrett, 2017): weekly counts of genomes carrying specified AA substitutions, and tools for comparing/combining substitutions within a gene (the *Compare AA mutations* module), used to obtain mutation frequency time-series and co-occurrence tallies. As of July 15, 2025, CovidCG reported 21,082,039 analyzable genomes (vs. 17,413,645 in the GISAID's own counter).
3. Nextclade (https://clades.nextstrain.org) (Aksamentov et al., 2021): used for clade assignment and recombinant flags on downloaded FASTA selections.
4. PDB structure 7EIZ from the Protein Data Bank, PDB (https://www.rcsb.org) and PyMOL v3.1 (https://www.pymol.org) were used solely for structure-mapping of residues corresponding to substitutions.
5. SARS-CoV-2 (COVID-19) stores the genome map of the Wuhan-Hu-1 isolate, with complete nucleotide and amino acid sequences (https://www.snapgene.com/plasmids/coronavirus_resources/SARS-CoV-2_(COVID-19)_Genome); together with GISAID's genome sequences lists of nucleotide and corresponding amino acid mutations available for all genome samples. We used it to create the table, showing which specific codon triplets in nucleotide sequences represent protein sequences in the reference genome.

Inclusion criteria and quality control

Sequences: human SARS-CoV-2 genomes flagged as complete, with low-coverage entries excluded, and complete collection dates (YYYY-MM-DD) required. These exact GISAID web interface options are enumerated in the Results/Methods narrative and were applied to every targeted query (e. g., *first occurrence* tables).

Time frame: December 2019 – July 15, 2025, matching the scope stated in the paper. Access dates for all web tools were within July 1–15, 2025 (final extraction and checks on July 15, 2025). All queries (filters, date ranges, and mutation lists) are specified in the main text (Tables/Figures).

Gene set: RTC proteins nsp7/8/9/10/12/13/14.

Data acquisition for single-point mutation frequencies

Data on amino acid substitution frequencies in SARS-CoV-2 RTC proteins (nsp7–nsp14) were obtained from the *Compare AA Mutations* section of the CovidCG platform (https://covidcg.org). This tool provides aggregated daily/weekly/monthly/annual counts of genomes carrying specific amino acid substitutions based on submissions to the GISAID database.

For each RTC protein (nsp7 through nsp14), amino acid–level data were selected. The grouping parameter was set to mutation, and the genomic coordinate system defined by protein. The analysis was performed using the full residue

[1] Supplementary Materials 1–4 are available at: https://vavilovj-icg.ru/download/pict-2026-30/appx28.pdf

range for each protein and included all geographic regions. The time range was set to cover the entire pandemic period (from December 2019 to July 2025) according to the *Since pandemic start* preset.

We extracted the information on all single amino acid substitutions with a total global frequency exceeding 1 % from the resulting datasets. This threshold is not a biological or statistical boundary. It was chosen as a pragmatic, operational cutoff to suppress locality-driven and unstable low-frequency noise and to keep the population-scale analysis tractable. Functionally important sub-1 % changes can exist, but these fall outside the scope of our population-scale focus. For each mutation, the CovidCG interface provides the number and proportion of genomes carrying the substitution relative to the total number of genomes available for the corresponding week, and the figures were used in further analysis of temporal dynamics.

Construction of weekly frequency-time series

Building on the dataset described above, we investigated the weekly dynamics of each identified mutation using the *AA Mutation Co-occurrence* module of the mentioned CovidCG platform. For each selected amino acid substitution within RTC proteins (nsp7–nsp14), we examined the New AA Mutation Percentages by Week chart, which displays the proportion of SARS-CoV-2 genomes carrying a given substitution over time. The data are shown as percentages and grouped by week to visualize the mutation prevalence dynamics throughout the entire observation period. The resulting time series for each substitution were exported via the CovidCG *Download* function and used to assess temporal trends in mutation frequencies.

Classification of dynamic types

Each mutation identified within the RTC (nsp7–nsp14) was represented as a weekly series reflecting the proportion of genomes carrying that substitution relative to the total number of genomes reported globally for the same week.

The observed temporal trajectories were categorized into three distinct dynamic types based on their characteristic shapes and relative frequency changes over time.

Persistent or fixation-like pattern. The frequency of the mutation exhibited a rapid increase from near 0 % to ≥90–100 %, followed by a consistently high level for the remainder of the observation period. These trajectories correspond to mutations that became fixed or near-fixed in the global SARS-CoV-2 population.

Transient pattern. The mutation frequency increased noticeably (typically reaching 10–60 %) but subsequently declined to near 0 %, indicating the rise and disappearance of a temporary lineage or variant in which the substitution was predominant.

Recurrent pattern. The mutation frequency showed multiple distinct peaks separated by intervals of decline, i. e., a pattern of emergence, near-complete disappearance, and subsequent re-establishment of high frequency (≥80–100 %). In our dataset, this behavior was observed only for the nsp12:G671S substitution. This mutation was accompanied by different sets of co-occurring substitutions during individual peaks, each exhibiting highly correlated frequency profiles throughout their respective intervals.

Lineage and clade context

Nextclade was run on representative FASTA samples exported from the GISAID for periods of rising and falling phases to obtain Nextstrain clades and recombinant flags. The Pango lineage labels shown in tables/figures were taken from the GISAID metadata of the same records (e. g., B.1.617.2/Delta, XBB.*, JN.1, etc.).

Structural mapping

Substitutions with a weekly peak frequency of over 25 % were mapped onto 7EIZ (extended RTC) with residue atoms rendered as spheres in PyMOL v3.1 (The PyMOL Molecular Graphics System, Version 3.1 Schrödinger, LLC, https://www.pymol.org). Figures were rendered directly from these models and used solely to visualize the spatial dispersion and proximity to RNA. No structural inference beyond visualization was performed.

Results and discussion

Mutations in SARS-CoV-2 RTC (nsp7–nsp14) proteins with overall average frequencies >1 % in the population

The list of mutations in nsp7–nsp14 proteins with the total average frequency exceeding 1 % in the global dataset during the period from January 1, 2020 to July 15, 2025 was obtained from CovidCG.org (Elbe, Buckland-Merrett, 2017), as described in "Data acquisition for single-point mutation frequencies" (section Materials and Methods). The results are presented in Table 1. The rightmost column contains values of peak weekly frequency of the same mutations, obtained as described in "Construction of weekly frequency-time series" (section Materials and Methods). In total, 22 mutations were identified.

Dynamics of SARS-CoV-2 mutation frequencies within RTC

The data on the weekly dynamics of mutation frequencies (from CovidCG.org) presented in Table 1 were obtained as described in "Construction of weekly frequency-time series" (section Materials and Methods) . The analyzed time interval is from January 1, 2020 to July 15, 2025.

We classified the curves presented in Figure 1 into three dynamic patterns: (1) persistent or fixation-like, (2) transient and (3) recurrent, as described in "Classification of dynamic types". We noticed only one mutation out of 22, nsp12:G671S, with two specific properties that distinguished it from all the others.

First, upon its emergence, it reached a 100 % frequency, remained dominant for months and then disappeared; however, the cycle repeated, and the mutation appears to be undergoing this process for the third time at present. Second, during each round of emergence, it was accompanied by other RTC mutations during either a single tide (nsp13:S36P, nsp14:A394V) or two subsequent tides (nsp13:S36P). Their mutation frequency curves during this period also reach high values (up to 90–100 %) and closely resemble that of nsp12:G671S.

Table 1. High-frequency substitutions in the RTC proteins

RTC protein	Protein length, aa	Mutation name	Number of genomes with this mutation	% of genomes with this mutation	Peak % in weekly dynamics
nsp7	83	–	–	–	–
nsp8	198	N118S	218,846	1.0	12.9
nsp9	113	T35I	1,181,871	6.8	100.0
nsp10	139	–	–	–	–
nsp11	14	–	–	–	–
nsp12 (RdRp)	932	F192V	484,275	2.3	9.6
		P227L	313,050	1.5	17.2
		Y273H	199,613	3.1	53.6
		P323L	20,916,671	99.3	100.0
		G671S	7,424,696	35.3	100.0
		F694Y	328,845	1.6	10.9
		L838I	325,059	1.5	8.1
nsp13 (helicase)	601	S36P	1,452,092	6.9	89.4
		P77L	5,810,196	27.6	100.0
		T127N	501,924	2.4	23.0
		H164Y	370,389	1.8	6.4
		M233I	792,615	3.8	53.9
		N268S	490,535	2.3	40.6
		I334V	600,127	2.8	13.1
		E341D	207,872	1.0	10.0
		R392C	8,622,554	40.9	100.0
		K460R	335,412	1.6	15.7
nsp14 (ExoN)	527	I42V	11,613,863	55.1	100.0
		N129D	274,952	1.3	24.0
		A394V	5,197,852	24.7	95.6

Note. Substitutions present in at least 1 % of genomes were collected globally for the period from December, 2019 to July 15, 2025. For each mutation, the number of genomes with this mutation, the percentage of genomes with this mutation in the global population, and the weekly peak percentage are presented.

The dynamics of nsp12:G671S frequency in the population – two and a half tides

The first tide of nsp12:G671S

The atypical dynamics of nsp12:G671S frequency in the population showed two complete "tides" (by analogy with the SARS-CoV-2 pandemic waves), each consisting of an emergence phase, growth, a plateau, and a decline to near-zero level. Early in 2025, a third tide appeared to have begun, and it is still ongoing. As we needed to better understand the process dynamics, we conducted an investigation to determine the connection between these events and the dominance periods of the Delta, Omicron, and other variants.

Since the frequency curves of genomes with nsp12:G671S, nsp13:P77L and nsp14:A394V almost completely overlapped during the first tide (May 2021 – May 2022), rising from near 0 % to nearly 100 % level and back, it is highly probable that these three mutations tended to occur together. Indeed, among 4.12 million genomes with nsp12:G671S, 4.19 million with nsp13:P77L and 3.81 million with nsp14:A394V (collection dates between May 1, 2021 and May 1, 2022 in GISAID), 3.75 million contained all the three amino acid substitutions simultaneously. Moreover, all samples collected in mid-May 2021 (during the rise of the tide) belonged to the Pango lineage B.1.617.2+AY.*, also known as the Delta variant. Our analysis aimed to determine whether the genomes carrying the nsp12:G671S + nsp13:P77L + nsp14:A394V combination or the Delta variant appeared first, or whether they appeared simultaneously. The results indicate that the latter is likely. Supporting details and evidence for this conclusion are shown in Figure 2.

By March 2020, the SARS-CoV-2 genome variants carrying nsp12:G671S, nsp13:P77L, and nsp14:A394V already existed. However, according to GISAID data, none of the samples contained them all simultaneously. When this combination first appeared in August 2020, the earliest genome identified as belonging to the Delta variant was detected, and it carried precisely these three mutations. Nearly all Delta genomes collected and sequenced thereafter also contained this mutation trio, suggesting that these mutations may have been required for the emergence and/or fitness advantage of Delta. Literature analysis revealed that this mutation trio had been mentioned in two articles describing variants detected in Iran (Ahmadi et al., 2023) and India (Bokolia, Gadepalli, 2022), but apparently received little attention at the time.

Timing the emergence of the Delta variant

The analysis of samples corresponding to the Delta variant (including early events classified as B.1.617.2 / Delta based on GISAID Pango metadata) and containing the trio of muta-

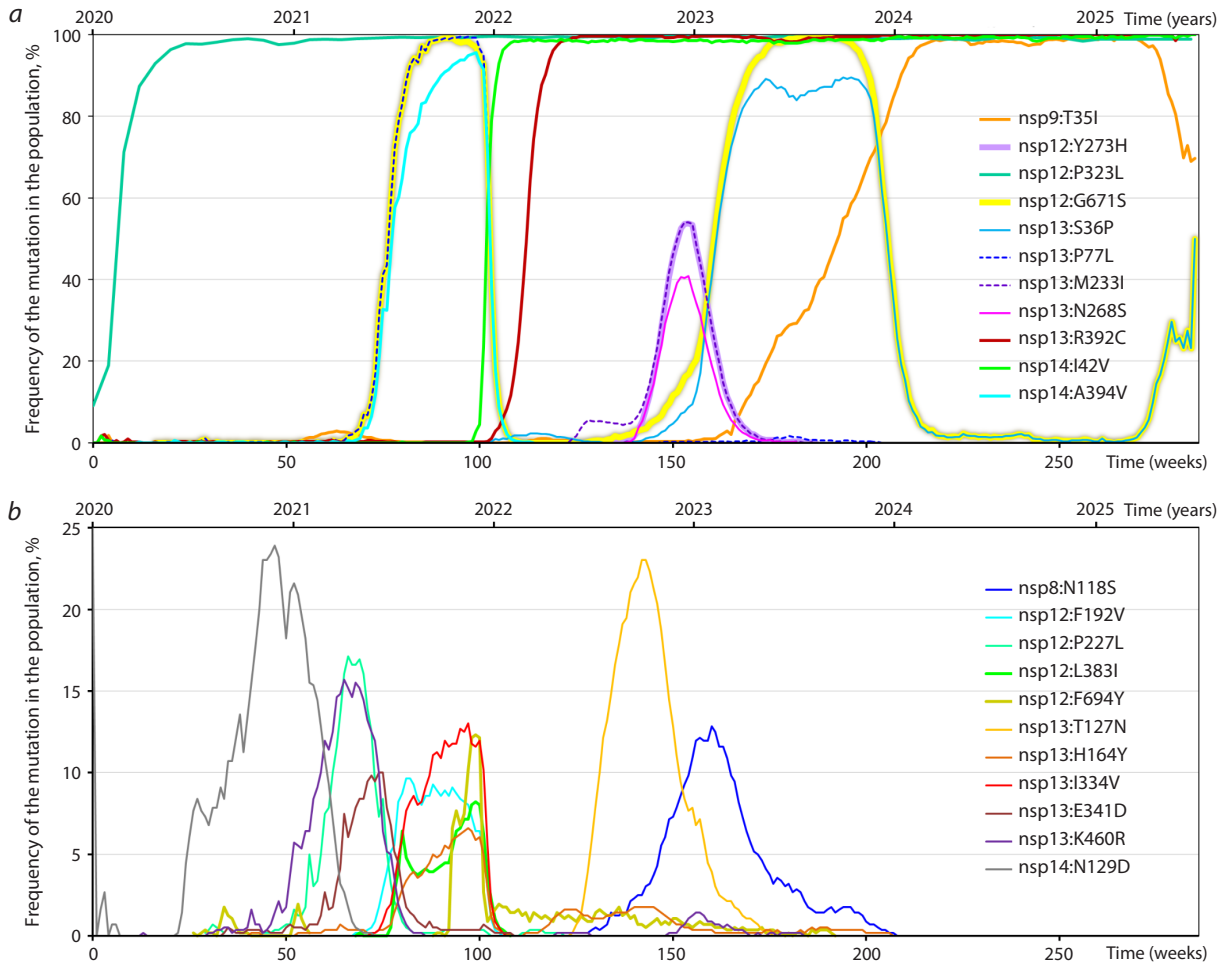


Fig. 1. Global frequency dynamics of mutations within RTC.

Weekly frequencies of 22 substitutions in the RTC proteins (listed in Table 1) from January 2020 to July 2025. *a*, Trajectories of mutations that reached a peak frequency between 25 and 100 %. *b*, Trajectories of mutations with a peak frequency below 25 %. Frequencies were calculated as the proportion of all GISAID genomes per week containing a specific mutation.

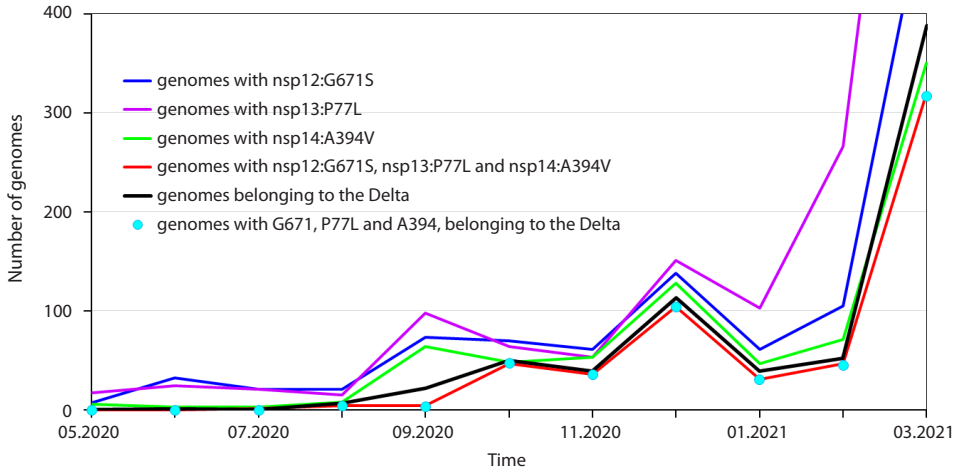


Fig. 2. Dynamics of the number of genomes with three key mutations in RTC genes related to the Delta variant.

Shortly after their simultaneous appearance in the Delta genomes, their number started to increase rapidly. Monthly numbers of genomes (GISAID) carrying only one of the mutations (nsp12:G671S, nsp13:P77L or nsp14:A394V), genomes with all three mutations and genomes belonging to Delta variant during the latent period (less than 10 cases per month) before the apparent rise of Delta in May 2021 are shown.

tions in question led us to an accidental finding: according to these data the Delta variant may have emerged 7–8 months earlier than commonly reported in the literature, and in a different geographical location. This observation is based on the GISAID data, which includes hundreds of genome samples (filtered by *complete* + *low coverage excluded* + *collection date complete* for reliability). Detailed data providing the evidence and explaining this conclusion are presented in the Supplementary Materials, Table S2 and Figure S1. In brief, the earliest B.1.617.2 (Delta) genome was collected on April 26, 2020 in "Europe/Germany/Freiburg" (GISAID EPI_ISL=852796). This contradicts the statement that "The Delta (B.1.617.2) variant was first identified in Maharashtra, India, in late 2020" (Cherian et al., 2021), which is frequently cited in earlier (Mlcochova et al., 2021), later (Rahman et al., 2023), and very recent (Kandhasamy et al., 2025) publications.

According to GISAID data, during the period from April 26, 2020 to December 9, 2020 – that is, before the first Delta case reported in India (Maharashtra) in "late 2020" (December 10, 2020, GISAID EPI_ISL_2131509) – the Delta (B.1.617.2) variant was not only detected but also collected and fully sequenced in various locations across multiple continents, including North America, South America, Europe, Asia, Africa, and Oceania, with a total of 275 genome records. Notably, 254 of these contained the nsp12:G671S mutation. During the subsequent period from April 26, 2020 to April 26, 2021, a total of 12,157 Delta genomes were collected and sequenced, of which 11,135 (92 %) also carried the nsp12:G671S substitution. A pronounced and rapid increase in the number of Delta genomes began only in mid-May 2021, marking the onset of its global expansion. The accompanying details and contextual events are described in "The first tide of nsp12:G671S".

The second tide of the nsp12:G671S

The SARS-CoV-2 variants with the mutations nsp13:P77L, nsp14:A394V and nsp12:G671S, which were discovered almost simultaneously and dominated at the peak of the first nsp12:G671S tide, completely disappeared in early 2023 (Fig. 1*a*). However, in contrast to nsp13:P77L and nsp14:A394V, which have not been detected since then, nsp12:G671S reappeared about six months later and once again reached a 100 % prevalence. This time, its rise almost coincided with the emergence of a new mutation, nsp13:S36P, which peaked at 89 %. The details and illustrations are provided in Supplementary Material 3 (Fig. S2).

We analyzed two sets of SARS-CoV-2 genome samples collected worldwide. The first set (499 samples) was collected on January 1, 2023, during the period of nsp12:G671S frequency increase, and the second set (1,497 samples) was collected in August 2023, at the onset of its frequency decline. During the rise, almost all samples (91 %) belonged to the XBB lineage (from XBB.1 to XBB.9), with the largest fraction (39 %) corresponding to XBB.1.5. During the decline, eight months later, the set of genomes was represented by XBB.* (31 %), EG.* (30 %), FL.* (9.3 %), GK.* (3.2 %), and several other sublineages. However, almost all of them were descendants of earlier XBB, and XBB itself was derived from Omicron, not Delta, which was the carrier of the nsp12:G671S mutation during the tide 1. Thus, it appears that in the tide 2 the nsp12:G671S emerged *de novo*. The cause of the subsequent extinction of all these variants was the emergence of the JN.1(24A) lineage, which appeared at the end of 2023 and took over almost the entire population by early 2024.

Additionally, within the considered period, another group of three mutations (nsp12:Y273H, nsp13:M233I, and nsp13:N268S) emerged and began to spread rapidly, coinciding with the rise of the nsp12:G671S + nsp13:S36P combination (Fig. 1*a*, *b*). Globally, the mutations nsp12:Y273H (a total of 493,881 samples), nsp13:M233I (489,239) and nsp13:N268S (377,576) showed similar dynamics and in most cases appeared in genomes together. This trio therefore represents a somewhat less extensive, but still substantial process: at one point, over 40 % of all genomes carried all three mutations. Variants harboring these substitutions mainly belonged to the BQ.1.* lineage.

The third tide of the nsp12:G671S, in progress

In autumn 2024, the frequencies of the nsp12:G671S and nsp13:S36P mutations declined for the second time to only a few percent, dropping below 1 % by the end of the year. However, this line did not disappear completely and continued to persist at low levels through late 2024. At the beginning of 2025, for the third time, the proportion of nsp12:G671S and nsp13:S36P in the population began to grow again, reaching 50 % in July 2025. The dynamics of nsp13:S36P in 2025 completely coincides with that of nsp12:G671S, i. e. these mutations occurred mostly together during this period. This pattern suggests that, unlike the disappearance of the carriers of the nsp12:G671S mutation after the first tide, it persisted after the second tide, albeit in small numbers in individual variants and then, under favorable circumstances and/or after advantageous mutations, began to spread again.

The analysis of samples collected in early 2025 showed that the vast majority of them originated from the JN.1 variant, which gave rise to such lineages as NB.1(24B), NB.1.8.1(25B), LP.8.1(25A), KP.3.1.1(24E), XFG(25C), LF.7(24H), XEC(24F) and XDV (24D), as well as to the recombinants XDA and XEV. Among them, the largest number of genomes carrying the nsp12:G671S mutation (42 %) corresponded to the lineage NB.1.8.1 (clade 25B) and 35 %, to PQ.1–PQ.9 (also clade 25B). Since JN.1 is a descendant of XBB, the suggestion that the nsp12:G671S and nsp13:S36P mutations have been preserved and their proportion in the population has grown again appears quite plausible.

Overview of all three tides of nsp12:G671S

At the start of both tide 1 and tide 2, a rapid and significant change in the proportion of nsp12:G671S in the population occurred, accompanied by the emergence of new co-occurring mutations in the RTC genes (and sometimes in the rest of the virus genome) or noticeable changes in their mutation frequencies. This trend also holds for tide 3 due to nsp13:S36P, which co-occurred with nsp12:G671S in 2025 (Fig. 1*a*). Additionally, we identified a recent mutation, nsp12:D284Y, which accompanied the two mutations and exhibited the same frequency dynamics in 2025. It is not included in Table 1,

because its prevalence among all genomes in GISAID in the time span from December 2019 to mid-July 2025 is below 1 %. However, in 2025, the frequency of this mutation followed the same pattern as nsp12:G671S and nsp13:S36P, reaching 50 % by mid-July 2025.

In addition, we noticed that during tide 3 (from early 2025 to mid-July 2025), the frequency curves of nsp12:G671S and nsp9:T35I were in antiphase (Fig. 1*a*). As the frequency of nsp12:G671S increased, that of nsp9:T35I decreased, and together, they consistently accounted for ≥80 % of all genomes in the weekly global samples. Based on these observations, a general trend can be inferred: when the fraction of nsp12:G671S mutation in the population changes rapidly, an emergence of new mutations in the RTC genes or changes in the frequencies of existing ones becomes more likely.

In contrast to nsp12:G671S, there are mutations with no obvious relation to others. One of them, nsp9:T35I, which emerged shortly after the onset of tide 2, displayed a nearly linear increase in frequency, from 0 to 100 %, in the interval between early 2023 and early 2024. This rise was much slower than any other curve in Figure 1, and it showed no correlation with any of them, which is quite unusual. The analysis of the genome samples carrying this mutation revealed the following. At the beginning of the growth phase in early 2023, it was represented by 51 % of FL.*, 25 % of XBB.1.9*, 20 % of EG.*, and 4 % others, including XCC recombinants. At the beginning of the decline phase in early 2025, the distribution shifted to its descendants: 29 % LP.*, 21 % XEC.*, 12 % NY.*, 7 % LP.*, 7 % MC.*, and the remaining 23 % comprised other lineages, including various recombinants: XFL, XFJ, XFH, XFC, XFB, XEW, XER, XEQ, and XEK.

Specific properties of the considered mutations in the RTC nsp7–nsp14 proteins

Each single mutation in amino acid sequences encoded by nucleotide triplets can include one to three changes within the triplet. Due to the degeneracy of the genetic code, there may be synonymous substitutions in RNA/DNA that do not change the amino acid, but even such mutations can have a noticeable effect on the fitness of their carrier, as shown by Shen et al. (2022). They came to this conclusion by experimenting with yeast, but they found "no particular reason why their results would not generalize to other organisms".

From the list of the previously considered mutations, we selected those with peak proportions ≥25 % of the population. By comparing their nucleotide and amino acid sequences in the reference genome and after mutation, we determined which mutated triplets appeared in the initial states and how they changed after mutation (Table 2).

All 13 mutations examined involved only a single nucleotide substitution within the codon encoding the affected amino acid; no double or triple substitutions were detected. This is not surprising, since if p is the probability of a single rare random event, then the probability that another similar event with the same probability occurs within the two adjacent nucleotides is $\sim p^2$ or less. The probability of a single point nucleotide substitution in SARS-CoV-2 is quite low since its replication mechanism includes a proofreading system. According to data from (Amicone et al., 2022), during one cell infection cycle (i. e., from virus entry into a cell until the release of new virions), on average, $1.3 \cdot 10^{-6} \pm 0.2 \cdot 10^{-6}$ substitutions occur per nucleotide position. The actual number of observable cases of 2 substitutions at once will be less than p^2 because there will

Table 2. Properties of most prevalent mutations in RTC genes (peak frequency ≥ 25 %)

Mutation name	Change at amino acid level	Changes at RNA level (bold indicates changed triplet, underline – changed a.a.)	Change of AA physico-chemical properties
nsp9:T35I	Thr→Ile	(ACU,ACC,**A<u>C</u>A**,ACG) → (ATT,ATC,**A<u>T</u>A**)	polar, hydroxylic → non-polar, aliphatic
nsp12:Y273H	Tyr→His	(**<u>T</u>AT**,TAC) → (**<u>C</u>AT**,CAC)	polar, aromatic → polar(+), heterocyclic
nsp12:D284Y	Asp→Tyr	(GAT,**<u>G</u>AC**) → (TAT,**<u>T</u>AC**)	polar(–) → polar, aromatic
nsp12:P323L	Pro→Leu	(**<u>C</u>C<u>T</u>**,CCC,CCA,CCG) → (TTA,TTG,**C<u>T</u>T**,CTC,CTA,CTG)	non-polar, heterocyclic → non-polar, aliphatic
nsp12:G671S	Gly→Ser	(**<u>G</u>GT**,GGC,GGA,GGG) → (TCT,TCC,TCA,TCG,**<u>A</u>GT**,AGC)	non-polar, aliphatic → non-polar, oxymonoaminocarboxylic
nsp12:L838I	Leu→Ile	(TTA,TTG,CTT,CTC,**<u>C</u>TA**,CTG) → (ATT,ATC,**<u>A</u>TA**)	non-polar, aliphatic → non-polar, aliphatic
nsp13:S36P	Ser→Pro	(TCT,TCC,**<u>T</u>CA**,TCG,AGT,AGC) → (CCT,CCC,**<u>C</u>CA**,CCG)	non-polar, oxymonoaminocarboxylic → non-polar, heterocyclic
nsp13:P77L	Pro→Leu	(CCT,CCC,**<u>C</u>C<u>A</u>**,CCG) → (TTA,TTG,CTT,CTC,**C<u>T</u>A**,CTG)	non-polar, heterocyclic → non-polar, aliphatic
nsp13:M233I	Met→Ile	**ATG** → (ATT,ATC,**A<u>T</u>A**)	non-polar, sulfur-containing → non-polar, aliphatic
nsp13:N268S	Asn→Ser	(**A<u>A</u>T**,AAC) → (TCT,TCC,TCA,TCG,**A<u>G</u>T**,AGC)	polar, amide → non-polar, oxymonoaminocarboxylic
nsp13:R392C	Arg→Cys	(**<u>C</u>GT**,CGC,CGA,CGG,AGA,AGG) → (**<u>T</u>GT**,GTC)	polar(+) → polar, sulfur-containing
nsp14:I42V	Ile→Val	(ATT,ATC,**A<u>T</u>A**) → (GTT,GTC,**<u>G</u>TA**,GTG)	non-polar, aliphatic → non-polar, aliphatic
nsp14:A394V	Ala→Val	(**G<u>C</u>T**,GCC,GCA,GCG) → (**G<u>T</u>T**,GTC,GTA,GTG)	non-polar, aliphatic → non-polar, aliphatic

Note. Substitutions of amino acids, changes in their physico-chemical properties and changes in RNA triplets encoding amino acids before and after mutations, are presented. In RNA triplets, the original combinations (among possible variants due to degeneracy of genetic code) are shown in boldface; changed points are underscored.

be only those variants which leave the virus viable despite caused changes in the replication-transcription complex, and the probability of this appears to be so small that it is practically unlikely to occur even over years. Nevertheless, it was not obvious in advance that all mutations would consist only of single-nucleotide substitutions, so this observation also brings some new knowledge of mutation patterns. However, changes can accumulate sequentially over multiple generations, experiencing only up to one nucleotide substitution per triplet in each replication cycle.

Additionally, the amino acid substitutions derived from the mutations examined frequently result in pronounced alterations of physicochemical properties.

The groups of co-existing mutations in the RTC nsp7–nsp14 proteins

Since we noticed co-occurring mutations in RTC proteins, we decided to identify all such groups in an attempt to reveal the underlying processes and possible connections between them. Using GISAID web interface queries (as described in "Data sources", section Materials and Methods) for the time interval from December 24, 2019 to July 15, 2025, we obtained the numbers of genome sequences containing all possible pairwise combinations of the mutations from Table 2, with the exception of nsp12:P323L (because it is present in 99.3 % of all sequences in GISAID). The results are presented in Table 3.

For a pair of mutations, for example, i = nsp12:G671S (34.9 % of total genomes) and j = nsp13:P77L (27.3 % of all genomes), the value in cell [i][j] of the colored matrix corresponds to 98.6 % of the smaller of the two proportions (34.9 and 27.3 %). Thus, the number of genomes carrying both mutations simultaneously is approximately $0.273 \cdot 0.986 \cdot 15.5$ million ≈ 4.2 million. In the global dataset, the most frequent pairs include:

– nsp13:R292C + nsp14:I42V (41.2 %),
– nsp12:G671S + nsp13:P77L (26.9 %),
– nsp13:P77L + nsp14:A394V (24.7 %),
– nsp12:G671S + nsp14:A394V (24.4 %).

Extending to triplets, representative high-frequency sets (among the 15.5 million GISAID genomes analyzed) are:
– nsp12:G671S + nsp13:P77L + nsp14:A394V (24.3 %),
– nsp13:R392C + nsp14:I42V + nsp13:M233I (3.6 %),
– nsp13:R392C + nsp14:I42V + nsp12:Y273H (2.9 %).

A notable triplet here is nsp12:G671S + nsp13:S36P + nsp12:D284Y (0.08 %), as nsp12:D284Y emerged only in 2025, and the number of genome samples carrying this mutation continues to increase, with its peak frequency reaching 50 % (see "The second tide of the nsp12:G671S", section Results and Discussion).

Finally, there are groups of genome samples with four or five mutations occurring simultaneously. Illustrative sets include:
– nsp13:R392C + nsp14:I42V + nsp13:M233I + nsp12:Y273H: (2.89 %),
– nsp13:R392C + nsp14:I42V + nsp13:M233I + nsp12:Y273H + nsp13:N268S: (2.23 %).

Three-dimensional visualization of the selected mutations

We rendered the examined mutations listed in Table 2 in the 3D structure of the RTC to learn about their spatial location, specific features, and distances between mutated amino acids, as well as between amino acids and RNA threads (Fig. 3).

According to Figure 3, the highest concentration of mutations is observed in nsp9 and nsp13 (helicase): 0.88 and 0.83 substitutions per 100 amino acids, respectively. Half as many mutations, 0.43 and 0.38 per 100 aa, occur in nsp12 (RdRp) and nsp14 (ExonN), correspondingly. In the remaining proteins, no mutations with a peak frequency exceeding 25 % of the population size were found. Three of four shown mutations in nsp12

Table 3. Pairs and larger groups of simultaneous mutations in RTC proteins

	mutation i →	nsp9: T35I	nsp13: S36P	nsp12: D284Y	nsp12: G671S	nsp12: L838I	nsp13: P77L	nsp14: A394V	nsp12: Y273H	nsp13: M233I	nsp13: N268S	nsp14: I42V	nsp13: R392C
mutation j ↓	% in the population → ↓	7.1	6.7	0.1	34.9	2.0	27.3	24.7	3.0	3.7	2.3	55.5	41.8
nsp9:T35I	7.1	100.0	35.1	0.7	34.3	0.3	0.8	0.7	0.3	0.3	0.2	95.7	96.3
nsp13:S36P	6.7	35.1	100.0	88.3	92.4	0.0	0.0	0.0	0.1	0.1	0.0	98.9	99.8
nsp12:D284Y	0.1	0.7	88.3	100.0	95.4	0.2	7.2	6.8	0.3	0.3	0.2	89.9	89.8
nsp12:G671S	34.9	34.3	92.4	95.4	100.0	98.9	98.6	98.7	0.2	1.1	0.2	22.4	22.7
nsp12:L838I	2.0	0.3	0.0	0.2	98.9	100.0	99.8	99.5	0.0	0.1	0.0	0.0	0.3
nsp13:P77L	27.3	0.8	0.0	7.2	98.6	99.8	100.0	99.8	0.1	1.0	0.1	0.0	0.2
nsp14:A394V	24.7	0.7	0.0	6.8	98.7	99.5	99.8	100.0	0.1	0.8	0.1	0.0	0.2
nsp12:Y273H	3.0	0.3	0.1	0.3	0.2	0.0	0.1	0.1	100.0	99.3	99.3	98.4	99.7
nsp13:M233I	3.7	0.3	0.1	0.3	1.1	0.1	1.0	0.8	99.3	100.0	99.4	97.2	98.5
nsp13:N268S	2.3	0.2	0.0	0.2	0.2	0.0	0.1	0.1	99.3	99.4	100.0	98.2	99.6
nsp14:I42V	55.5	95.7	98.9	89.9	22.4	0.0	0.0	0.0	98.4	97.2	98.2	100.0	98.5
nsp13:R392C	41.8	96.3	99.8	89.8	22.7	0.3	0.2	0.2	99.7	98.5	99.6	98.5	100.0

Note. The percentage of genomes in the GISAID (collected from December 2019 to July 15, 2025) that simultaneously contain mutation i and mutation j, relative to the minimum value between the percentages of genomes with mutation i and genomes with mutation j.

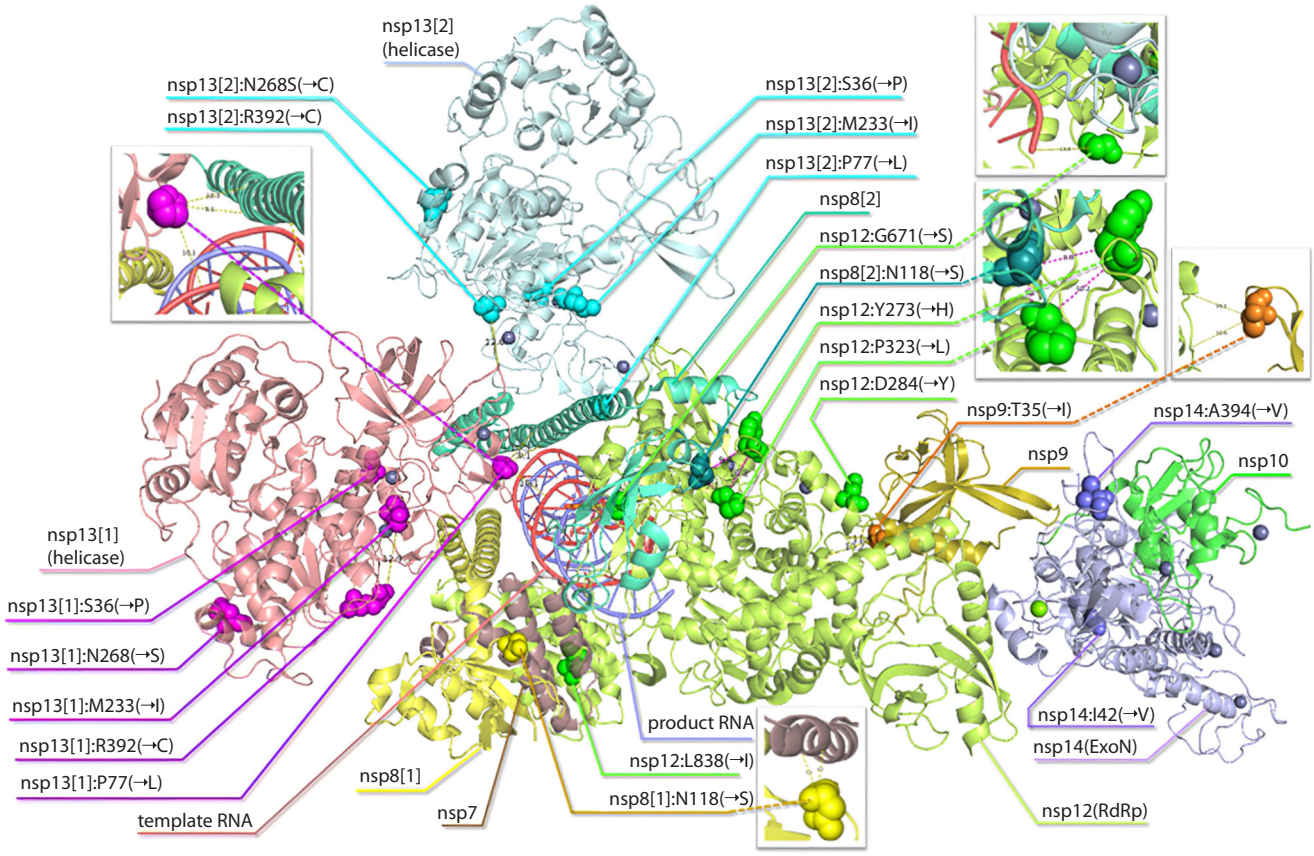


Fig. 3. 3D structure of the RTC, and its amino acids with mutations under consideration shown.

Distribution of the mutations from Table 3 over the replication-transcription complex (PDB:7EIZ) including nsp7, 2×nsp8, nsp9, nsp10, RdRp(nsp12), 2×nsp13, and nsp14 together with template RNA and product RNA. Cases with specific features are shown in the insets. The generated PyMOL files based on PDB RTC 3D structures and examined mutations, are available upon request.

(RdRp) are located on the surface of the polymerase, on the side closer to RNA. There are several proximity-based (within ~6–14 Å) potential relationships that may merit functional follow-up. The mutated amino acid closest to the template RNA (9.1 Å) is nsp13:M233I (Fig. 3, inset). In nsp13, P77L is positioned 10.1 Å from the RNA strand coursing through the complex and 9.1 Å from an nsp8 α-helix, placing it in a location where effects on replication kinetics and/or fidelity are plausible. Next, nsp8:N118S resides 6.1 Å from nsp7, which may affect the interaction between them.

RTC includes two copies of nsp8 and two copies of nsp13, with different neighborhoods for each protein of a couple. In one nsp8, the amino acid N118 (mutation N→S) lies near two nsp12 amino acids corresponding to the P323L and Y273H mutations: the distance between N118 and P323 is 5.1 Å, between N118 and Y273 is 9.8 Å, and between P323 and Y273 is 10.2 Å. This triangle (Fig. 3, inset) constitutes a cluster of three amino acids involved in mutations, close to each other, and we flag this cluster as a priority for subsequent functional investigation.

Search for literature on SARS-CoV-2 RTC mutation effects

To complement our mutational analysis, we performed a literature search for all substitutions listed in Table 2. The goal was to identify any previously reported structural, biochemical, or phenotypic effects associated with these mutations. Brief summaries of the available evidence are provided in Supplementary Material 4. Overall, the collected publications describe diverse potential impacts, including changes in RTC stability, replication efficiency, drug response, and enzymatic activity, although for many mutations experimental validation remains limited.

Conclusions

We surveyed the evolution of the SARS-CoV-2 RTC (nsp7–nsp14) over a 5.5-year horizon (December 2019 – July 15, 2025), identified 22 amino acid substitutions with a global average frequency >1 % and quantified their weekly dynamics. Compared to earlier global overviews that used higher prevalence thresholds (Rodriguez et al., 2023, 2025), this extended window and finer temporal resolution revealed three characteristic dynamic patterns: persistent, transient, and recurrent. Interestingly, one notable mutation, nsp12:G671S, exhibited a unique multi-phase trajectory (“three tides”) marked by a rise, near-disappearance, and re-emergence.

Re-analysis of early records indicated that SARS-CoV-2 genomes deposited to GISAID database and assigned to Delta (B.1.617.2) variant have submission dates that are 7–8 months earlier than commonly reported date of the first appearance of Delta. GISAID’s first samples of Delta appear for the first

time on April 26, 2020, they spanned multiple continents by the end of 2020 (and frequently carrying the trio of mutations nsp12:G671S + nsp13:P77L + nsp14:A394V). This observation motivates a reassessment of the widely cited timeline stating that "Delta was first identified in Maharashtra, India, in late 2020" (Cherian et al., 2021) and emphasizes the value of stringent filtering of public surveillance datasets.

In the course of systematically scanning RTC variation, we revealed a robust linkage between Delta and the nsp12:G671S + nsp13:P77L + nsp14:A394V trio. Tracing these substitutions back showed that each appeared separately before Delta, and that early genomes labeled as Delta appeared sporadically for several months but had little immediate effect on global frequency among sequenced genomes. The subsequent co-occurrence of all three substitutions around August 2020 coincided with a marked shift to rapid expansion and, soon after, global dominance. This perspective emerges specifically from an RTC-centered analysis (traditionally viewed as a conserved, lower-visibility target) and, in our view, illuminates otherwise overlooked contingencies in the evolutionary dynamics of the virus.

Across waves, nsp12:G671S re-emerged with different companions: first with nsp13:P77L and nsp14:A394V, later with nsp13:S36P, and most recently with nsp12:D284Y. In contrast, nsp9:T35I followed a slow, largely independent, monotonic rise, longer period of 97–100 % prevalence, and a slow decline that began in 2025. A codon-level inspection of the 13 highest-peak substitutions (≥25 %) showed only single-nucleotide changes (no double/triple codon substitutions), yet many corresponding amino acid replacements entail substantial physicochemical shifts. Co-occurrence analysis highlighted frequent groups of co-existing mutations, ranging from two to five. Structural mapping onto RTC assemblies showed a relatively even spatial distribution without dense clusters; only nsp13:M233I had a notable feature – it lay directly adjacent to the template RNA. The repeated rise and fall in the number of genome samples with the nsp12:G671S mutation, given the accompanying events, probably represents an example of natural selection in action. We can propose as a hypothesis for discussion that it may be not a neutral change, but a key player in the evolutionary "strategy" that provides a short-term fitness advantage (compensating for proofreading errors) at the expense of long-term stability (increasing mutation load). This creates a predictable cycle of emergence and extinction – a pattern that is the hallmark of selection, not random drift.

Independent population genetic analyses have shown that the evolutionary rate of SARS-CoV-2 varies among lineages and can undergo episodic accelerations, up to fourfold exceeding the baseline phylogenetic rate, with new variants emerging within weeks rather than months (Tay et al., 2022). The recurrent G671S tides with changing co-mutation partners documented in this study are consistent with such episodic phases. Given that several high-prevalence co-mutations lie in nsp14 (ExoN), these patterns raise the question of whether fluctuations in effective error correction could accompany periods of rapid diversification. Targeted assays will be required to test this hypothesis.

To contextualize these findings, we reviewed the available literature on all substitutions from Table 3 (Supplementary Material 4). Reported effects encompass altered polymerase stability/transmissibility for nsp12:P323L/G671S (Kim et al., 2023), lineage-linked occurrences and additional RTC changes documented through mid-2023 (Rodriguez et al., 2023, 2025), and mixed evidence regarding drug response and enzymatic activity for several sites. While suggestive, many substitutions still lack a definitive experimental validation of their functional impact. Protein-protein and protein-RNA interactions are complicated, and they involve many factors even if a single protein is considered (for example, (St Laurent et al., 2012, Fig. 5)). Our study is focused on one of the most intricate protein complexes with a critically important function, RNA replication.

To sum up, our results refine the timeline of early Delta emergence and dissemination, document repeated large-scale shifts centered around nsp12:G671S, and provide a reproducible catalog of RTC mutation dynamics, co-occurrence, coding changes, and structural context. The continued integration of high-quality genomic surveillance with targeted biochemical and virological assays will be essential to resolve the functional consequences of these substitutions and to anticipate future shifts in the RTC mutational landscape.

References

Abbasian M.H., Mahmanzar M., Rahimian K., Mahdavi B., Tokhanbigli S., Moradi B., Sisakht M.M., Deng Y. Global landscape of SARS-CoV-2 mutations and con-served regions. *J Transl Med.* 2023;21(1):152. doi 10.1186/s12967-023-03996-w

Ahmadi A.S., Shafiei-Jandaghi N.Z., Sadeghi K., Nejati A., Zadheidar S., Mokhtari-Azad T., Yavarian J. Comparison of circulating variants during the beginning, middle and the end of the 4th wave of COVID-19 in Tehran Province, Iran in 2021. *Iran J Public Health*. 2023;52(12):2621-2629. doi 10.18502/ijph.v52i12.14323

Aksamentov I., Roemer C., Hodcroft E.B., Neher R.A. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J Open Source Softw.* 2021;6(67):3773. doi 10.21105/joss.03773

Amicone M., Borges V., Alves M.J., Isidro J., Ze-Ze L., Duarte S., Vieira L., Guiomar R., Gomes P.J., Gordo I. Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. *Evol Med Public Health*. 2022;10(1):142-155. doi 10.1093/emph/eoac010

Bai C., Zhong Q., Gao G.F. Overview of SARS-CoV-2 genome-encoded proteins. *Sci China Life Sci.* 2022;65(2):280-294. doi 10.1007/s11427-021-1964-4

Bokolia N.P., Gadepalli R. Identification of genomic signatures and multiple lineage markers from the second and third wave samples of COVID-19 in Western Rajasthan, India. *bioRxiv*. 2022. doi 10.1101/2022.12.10.518819

Brant A.C., Tian W., Majerciak V., Yang W., Zheng Z.-M. SARS-CoV-2: from its discovery to genome structure, transcription, and replication. *Cell Biosci.* 2021;11(1):136. doi 10.1186/s13578-021-00643-z

Carabelli A.M., Peacock T.P., Thorne L.G., Harvey W.T., Hughes J.; COVID-19 Genomics UK Consortium; Peacock S.J., Barclay W.S., de Silva T.I., Towers G.J., Robertson D.L. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol.* 2023;21(3):162-177. doi 10.1038/s41579-022-00841-7

Cherian S., Potdar V., Jadhav S., Gupta N., Das M., Rakshit P., Singh S., Abraham P., Panda S., Team N. SARS-CoV-2 spike mutations, L452R, T478K, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. *Microorganisms*. 2021;9(7):1542. doi 10.3390/microorganisms9071542

Danda M., Klimešová A., Kušková K., Dostálková A., Pagáčová A., Prchal J., Kapisheva M., Ruml T., Rumlová M. Biochemical characterization of naturally occurring mutations in SARS-CoV-2 RNA-dependent RNA polymerase. *Protein Sci.* 2024;33(9):e5103. doi 10.1002/pro.5103

Elbe S., Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall.* 2017;1(1):33-46. doi 10.1002/gch2.1018

Eriani G., Martin F. Viral and cellular translation during SARS-CoV-2 infection. *FEBS Open Bio.* 2022;12(9):1584-1601. doi 10.1002/2211-5463.13413

Feng S., O'Brien A., Chen D., Saeed M., Baker S.C. SARS-CoV-2 nonstructural protein 6 from Alpha to Omicron: evolution of a transmembrane protein. *mBio.* 2023;14(4):e00688-23. doi 10.1128/mbio.00688-23

Ferrer-Orta C., Vázquez-Monteagudo S., Ferrero D.S., Verdaguer N. Point mutations at specific sites of the nsp12–nsp8 interface dramatically affect the RNA polymerization activity of SARS-CoV-2. *Proc Natl Acad Sci USA.* 2024;121(29):e2317977121. doi 10.1073/pnas.2317977121

Kandhasamy V., Pillai A.B., Mariappan V., Ramalingam M., Raganadin P., Ramadoss R., Moovarkumudalvan B., Easow J.M., Vasudevan M., Rao S.R. A study of SARS-CoV-2 genomic profiles, evolutionary changes, and transmission dynamics in Southeastern India during three pandemic waves. *JIOMICS.* 2025;15(2):239. doi 10.5584/jiomics.v15i2.239

Khare S., Gurry C., Freitas L., Schultz M.B., Bach G., Diallo A., Akite N., Ho J., Lee R.T., Yeo W., Curation Team G.C., Maurer-Stroh S. GISAID's role in pandemic response. *China CDC Wkly.* 2021;3(49):1049-1051. doi 10.46234/ccdcw2021.255

Kim S.M., Kim E.H., Casel M.A.B., Kim Y.I., Sun R., Kwak M.J., Yoo J.S., … Hwang J., Song M.S., Kim M.H., Jung J.U., Choi Y.K. SARS-CoV-2 variants with NSP12 P323L/G671S mutations display enhanced virus replication in ferret upper airways and higher transmissibility. *Cell Rep.* 2023;42(9):113077. doi 10.1016/j.celrep.2023.113077

Mack A.H., Menzies G., Southgate A., Jones D.D., Connor T.R. A proofreading mutation with an allosteric effect allows a cluster of SARS-CoV-2 viruses to rapidly evolve. *Mol Biol Evol.* 2023;40(10):msad209. doi 10.1093/molbev/msad209

Mlcochova P., Kemp S.A., Dhar M.S., Papa G., Meng B., Ferreira I.A.T.M., Datir R., … Piccoli L., Barclay W.S., Rakshit P., Agrawal A., Gupta R.K. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature.* 2021;599(7883):114-119. doi 10.1038/s41586-021-03944-y

Ou X., Xu G., Li P., Liu Y., Zan F., Liu P., Hu J., … Jin Q., Liu P., Lu J., Wang X., Qian Z. Host susceptibility and structural and immunological insight of S proteins of two SARS-CoV-2 closely related bat coronaviruses. *Cell Discov.* 2023;9(1):78. doi 10.1038/s41421-023-00581-9

Palyanov A.Y., Palyanova N.V. On the space of SARS-CoV-2 genetic sequence variants. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2023;27(7):839-850. doi 10.18699/VJGB-23-97

Palyanov A.Y., Palyanova N.V. A novel approach to analyzing the evolution of SARS-CoV-2 based on visualization and clustering of large genetic data compactly represented in operative memory. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2024;28(8):843-853. doi 10.18699/vjgb-24-92

Palyanova N.V., Sobolev I.A., Palyanov A.Y., Kurskaya O.G., Komissarov A.B., Danilenko D.M., Fadeev A.V., Shestopalov A.M. The development of the SARS-CoV-2 epidemic in different regions of Siberia in the 2020–2022 period. *Viruses.* 2023;15(10):2014. doi 10.3390/v15102014

Peng Q., Peng R., Yuan B., Zhao J., Wang M., Wang X., Wang Q., Sun Y., Fan Z., Qi J., Gao G.F., Shi Y. Structural and biochemical characterization of the nsp12-nsp7-nsp8 core polymerase complex from SARS-CoV-2. *Cell Rep.* 2020;31(11):107774. doi 10.1016/j.celrep.2020.107774

Rahalkar M.C., Bahulikar R.A. Lethal pneumonia cases in Mojiang miners (2012) and the mineshaft could provide important clues to the origin of SARS-CoV-2. *Front Public Health.* 2020;8:581569. doi 10.3389/fpubh.2020.581569

Rahman M.S., Hoque M.N., Chowdhury S.R., Siddique M.M., Islam O.K., Galib S.M., Islam M.T., Hossain M.A. Temporal dynamics and fatality of SARS-CoV-2 variants in Bangladesh. *Health Sci Rep.* 2023;6(4):e1209. doi 10.1002/hsr2.1209

Rodriguez L., Li J., Han D., Martin R., Moshiri J., Peinovich N., Bilello J.P., Perry J.K., Hedskog C. Remdesivir and obeldesivir retain potent activity against SARS-CoV-2 omicron variants. *Open Forum Infect Dis.* 2023;10(Suppl.2):ofad500.614. doi 10.1093/ofid/ofad500.614

Rodrigruez L., Zamora J.L.R., Han D., Moshiri J., Peinovich N., Martinez C., Ho P.Y., Li J., Aeschbacher T., Martin R., Pekosz A., Bilello J.P., Perry J.K., Hedskog C. Remdesivir and obeldesivir retain potent antiviral activity against SARS-CoV-2 omicron variants. *Viruses*. 2025;17(2):168. doi 10.3390/v17020168

Romano M., Ruggiero A., Squeglia F., Maga G., Berisio R. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells.* 2020;9(5):1267. doi 10.3390/cells9051267

Schubert K., Karousis E.D., Jomaa A., Scaiola A., Echeverria B., Gurzeler L.A., Leibundgut M., Thiel V., Mühlemann O., Ban N. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol.* 2020;27(10):959-966. doi 10.1038/s41594-020-0511-8

Shen X., Song S., Li C., Zhang J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature.* 2022;606(7915):725-731. doi 10.1038/s41586-022-04823-w

Singh D., Kushwaha T., Kulandaisamy R., Kumar V., Baswal K., Tiwari S.H., Ghorai A., … Gadde S., Côté M., Kayampeta S.R., Appaiahgari M.B., Inampudi K.K. Redefining *NSP12* activity in SARS-CoV-2 and its regulation by *NSP8* and *NSP7*. *Mol Ther Nucleic Acids.* 2025;36(1):102452. doi 10.1016/j.omtn.2025.102452

St Laurent G., Shtokalo D., Heydarian M., Palyanov A., Babiy D., Zhou J., Kumar A., Urcuqui-Inchima S. Insights from the HuR-interacting transcriptome: ncRNAs, ubiquitin pathways, and patterns of secondary structure dependent RNA interactions. *Mol Genet Genomics.* 2012;287(11-12):867-879. doi 10.1007/s00438-012-0722-8

Tay J.H., Porter A.F., Wirth W., Duchene S. The emergence of SARS-CoV-2 variants of concern is driven by acceleration of the substitution rate. *Mol Biol Evol.* 2022;39(2):msac013. doi 10.1093/molbev/msac013

Temmam S., Vongphayloth K., Baquero E., Munier S., Bonomi M., Regnault B., Douangboubpha B., … Nilges M., Rey F.A., van der Werf S., Brey P.T., Eloit M. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature.* 2022;604(7905):330-336. doi 10.1038/s41586-022-04532-4

Yan L., Yang Y., Li M., Zhang Y., Zheng L., Ge J., Huang Y.C., … Huang Y.Y., Guddat L.W., Gao Y., Rao Z., Lou Z. Coupling of N7-methyltransferase and 3′-5′ exoribonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading. *Cell.* 2021;184(13):3474-3485.e11. doi 10.1016/j.cell.2021.05.033

Yan W., Zheng Y., Zeng X., He B., Cheng W. Structural biology of SARS-CoV-2: open the door for novel therapies. *Signal Transduct Target Ther.* 2022;7(1):26. doi 10.1038/s41392-022-00884-5

Zhou P., Yang X.L., Wang X.G., Hu B., Zhang L., Zhang W., Si H.R., … Yan B., Zhan F.X., Wang Y.Y., Xiao G.F., Shi Z.L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020;579(7798):270-273. doi 10.1038/s41586-020-2012-7

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

Received November 22, 2025. Revised February 4, 2026. Accepted February 12, 2026.

BIOINFORMATICS AND SYSTEMS BIOLOGY
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(3):527-536

doi 10.18699/vjgb-26-55

Genetic analysis of wheat ear architecture in F_2 hybrid of tetraploid wheats *Triticum aethiopicum* and *T. carthlicum* and its computer phenotyping

Yu.V. Kruchinina [1], E.G. Komyshev [1, 2], M.A. Genaev [1, 2], V.S. Koval [1, 2], D.A. Afonnikov [1, 2, 3], N.P. Goncharov [1]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
kruchinina2023@yandex.ru

Abstract. A comprehensive description of plant phenotypes of certain taxa is an important task when describing genera and species, as well as when setting their natural taxonomies. The development of modern technologies of effective phenotyping makes it possible to obtain a large amount of data with a quantitative and/or qualitative description of various traits in plants, mainly based on the analysis of their digital images. The study compared the results of the F_2 hybrids assessment – visually and using machine learning methods – of two endemic tetraploid ($2n = 4x = 28$) wheat species which are Ethiopian wheat (*Triticum aethiopicum* Jakubz.) and Kartalian or Dika wheat (*T. carthlicum* Nevski). In the latter case, it is proposed to use the method of a mixture of Gaussian (normal) distributions in plant morphometry in order to identify groups that differ in character values. Most taxonomically important (species-specific) traits are controlled oligogenically and have a clear phenotypic manifestation, so hybridological analysis was an indispensable and basic type of analysis for subsequent detailed phenotyping of wheat spikes using machine-learning methods. According to a number of criteria, the estimates of patterns of inheritance obtained by different methods coincide. Based on the conducted research, we can state that the trait "tetraaristatum" (the presence of awns on both flower and spike glumes) is species-specific (taxonomically important) for *T. carthlicum* and it can be effectively used for taxonomic purposes both in carrying out hybridological analysis and in experiments using machine learning. Such a species-specific character is the "character (type) of awnedness" for *T. aethiopicum*. Our study demonstrates that a combination of automatic phenotyping methods and a model of a mixture of Gaussian distributions can, in principle, lead to an automatic analysis of the allocation of classes in F_2 hybrids. It allows, in turn, to detect the presence of genes associated with species-specific traits of wheat plants. Further, the improvement of the applied artificial intelligence (AI) algorithms is required.
Key words: tetraploid wheat; *Triticum aethiopicum*; *T. carthlicum*; genetic analysis; hybridization; cleavage; digital phenotyping; machine learning

For citation: Kruchinina Yu.V., Komyshev E.G., Genaev M.A., Koval V.S., Afonnikov D.A., Goncharov N.P. Genetic analysis of wheat ear architecture in F_2 hybrid of tetraploid wheats *Triticum aethiopicum* and *T. carthlicum* and its computer phenotyping. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(3):527-536. doi 10.18699/vjgb-26-55

Funding. The preparation of data, development, and verification of the algorithm were supported by the Russian Science Foundation, project No. 23-14-00150.

Acknowledgements. The plants were grown at the Collective Use Center for Plant Growing at the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences as part of the budget project FWNR-2022-0017.

Генетический анализ архитектуры колоса пшеницы и его компьютерное фенотипирование у F_2 гибридов тетраплоидных пшениц *Triticum aethiopicum* и T. *carthlicum*

Ю.В. Кручинина [1], Е.Г. Комышев [1, 2], М.А. Генаев [1, 2], В.С. Коваль [1, 2], Д.А. Афонников [1, 2, 3], Н.П. Гончаров [1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
kruchinina2023@yandex.ru

Аннотация. Всестороннее описание фенотипов растений конкретных таксонов является важной задачей при описаниях родов, видов и построениях их естественных классификаций. Развитие современных технологий эффективного фенотипирования позволяет получать большое количество данных с количественным и/или

качественным описанием различных признаков у растений, преимущественно на основе анализа их цифровых изображений. В исследовании проведено сравнение результатов оценки расщепления в F_2 гибридов двух эндемичных тетраплоидных ($2n = 4x = 28$) видов пшениц – пшеницы эфиопской (*Triticum aethiopicum* Jakubz.) и пшеницы карталинской (*T. carthlicum* Nevski), полученных визуально и с помощью методов машинного обучения. В последнем случае для анализа расщеплений предложено использовать метод смеси гауссовых (нормальных) распределений в морфометрии растений для того, чтобы выделить группы, которые различаются по значениям признаков. Ввиду того что большинство таксономически значимых (видообразующих) признаков у пшениц контролируется олигогенно и имеет четкое фенотипическое проявление, гибридологический анализ был незаменимым и основным видом анализа для последующего детального фенотипирования колосьев видов пшениц с использованием методов машинного обучения. По ряду признаков оценки характера наследования таких признаков, полученные разными методами, совпадают. На основании результатов проведенного исследования мы можем утверждать, что признак «тетраостость» (наличие остей одновременно на цветковой и колосковой чешуях) видоспецифический (таксономически значим) для *T. carthlicum* и он может быть эффективно использован исследователями в таксономических целях как при проведении гибридологического анализа, так и в экспериментах с использованием машинного обучениям. Для *T. aethiopicum* таким видоспецифическим признаком является «характер (тип) остистости». Наша работа демонстрирует, что комбинация методов автоматического фенотипирования, модели смеси гауссовых распределений в принципе позволит проводить автоматический анализ выделения классов в F_2 гибридов, различающихся по величинам фенотипических признаков. Это позволит, в свою очередь, выявлять наличие генов, ассоциированных с видоспецифическими признаками растений. Тем не менее требуется совершенствование применяемых алгоритмов искусственного интеллекта.

Ключевые слова: тетраплоидные пшеницы; *Triticum aethiopicum*; *T. carthlicum*; генетический анализ; гибридизация; расщепление; цифровое фенотипирование; машинное обучение

Introduction

A comprehensive description of the phenotypes of plants in specific taxa is an important task for effectively describing of species and constructing natural taxonomy (Hodač et al., 2023; Ran et al., 2024). However, this is a labor-consuming task. In some cases, there is very little information about a particular species, or the species is very rare (Goncharov, Adonina, 2024). In other cases, there is so much information that researchers disagree on the taxonomy of specific accession (Lyapunova, 2021). In addition, information on the species and volumes of genus varies significantly in a number of taxonomies that are still used today. Examples include *Solanum* L. by S.M. Bukasov (1971) and J. Hawkes (1963); *Aegilops* L. by P.M. Zhukovsky (1928) and A. Eig (1929); *Triticum* L. by N.P. Goncharov (2011) and K. Hammer et al. (2011). Therefore, researchers may experience inaccuracies in the taxonomy of species, even knowing which species-specific (taxonomically important) traits are characteristic of a certain taxon. The solution of this problem is to create a system of species-specific characters.

Using cereals (Poaceae Barnhart) as an experimental material, specifically wheat (genus *Triticum*), one can encounter all the difficulties described above. In addition, it should be borne in mind that wheat is a polymorphic crop, which implies significant inter- and intraspecific variability (Dorofeev et al., 1979). That is, in the process of studying it, one may encounter the fact that the plants inside the specific species will be phenotypically dissimilar and require conducting an experimental assessment of their species affiliation (Zuev et al., 2019). In such cases, we are talking about the unambiguous taxonomies of lower taxa, such as subspecies and subvarieties (Dorofeev et al., 1979; Goncharov, 2009).

Currently, there are a number of methods that allow to assign a specific wheat accessions to a particular species with a high degree of certainty. Using cytological methods, you can accurately assign an accessions to a specific ploidy group by counting the number of chromosomes (Dolezel et al., 2007), while molecular-biology methods can be used to assign accessions to a specific species (Golovnina et al., 2007). However, the use of molecular methods requires a large amount of information about each species, including a description of species-specific genes, their localization, and information about their nucleotide sequence. In wheat, species-specific genes have been described only in all hexaploid species ($2n = 6x = 42$) (Goncharov, 2011). This is the source of troubles for taxonomists. First, not each species of wheat that has been described so far has a species-specific gene(s). Second, even if a gene(s) is known, there may be difficulties in identifying its nucleotide sequence. Third, in some cases, a gene may be known, as well as its sequence, but its molecular function and phenotypic expression may not be studied. In addition, the information about the inheritance of a particular gene may vary among different authors (URL: http://shigen.nig.ac.jp/wheat/komugi/genes/symbolClassList.jsp Accepted October 20, 2025). Furthermore, it should be noted that molecular biological methods for analyzing large accessions lead to significant costs.

There are other methods that allow us to clarify information about species-specific genes. One of these methods is hybridological analysis. This method helps us to understand how species-specific traits are inherited in certain wheat species. By analyzing the patterns of inheritance in hybrid populations, we can gain insights into the way traits are passed down.

It should be noted that the development of modern technologies for effective phenotyping allows to obtain a large amount of data with quantitative or qualitative descriptions of various plant characteristics, primarily based on the analysis of digital images (Afonnikov et al., 2016; Awada et al., 2024). The data obtained through the analysis of digital images can be successfully utilized for the taxonomy and classification of plants (Chouhan et al., 2024; Mulugeta et al., 2024).

Approaches based on the analysis of large volumes of data and machine learning are also used to analyze the morphological characteristics of wheat. It has been shown that data obtained using image analysis and machine learning methods can successfully solve problems of the taxonomies of wheat species and their relatives (Martinek, Bednar, 2001; Pronozin et al., 2021; Artemenko et al., 2024; Komyshev et al., 2024). This is a promising area of research on hybrids using methods of bioinformatics, which is currently being developed. So the question remains open of how reliably it is to use characters, the characteristics of which are obtained on the basis of the analysis of digital images, for species taxonomies and how they are inherited.

The purpose of the investigation is to combine the possibilities of the hybridological method with the analysis of the splitting of the hybrid population manually with the analysis of digital characteristics of plants (spikes) to solve the problem of dividing plants into groups carrying different genes of plant traits control. The study is based on plant hybrids obtained by crossing two endemic wheat species, *T. aethiopicum* Jakubz. and *T. carthlicum* Nevski.

Materials and methods

Hybridological analysis. The work identified the genotypes of tetraploid ($2n = 4x = 28$) wheat and their F_2 hybrids based on species-specific (taxonomically important) traits (Supplementary Table S1)[1]. Since most of these traits are controlled by oligogenes and have a clear phenotypic manifestation, the hybridological method was an indispensable and primary type of analysis for subsequent detailed phenotyping spikes of wheat species using machine learning methods.

Plant material. The object of study was interspecific hybrids obtained by crossing two endemic tetraploid wheat species ♀*T. aethiopicum* Jakubz. (k-19301/2) with ♂*T. carthlicum* Nevski (k-32496). The experiment was produced in spring sowing in the greenhouses of the Breeding and Genetics Complex (BGC) of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (Novosibirsk). F_1 hybrids were planted and phenotyped in a greenhouse, and F_2 hybrids were grown in greenhouses in 2022. 185 F_2 hybrids were studied.

Phenotyping of spike. Image acquisition of spike was carried out according to the "on a clothespin" protocol described earlier (Genaev et al., 2018, 2019). The spike was fixed in an upright position with a clothespin on a blue background. The shooting was done with a Canon 350D digital camera and an EF-S 18–55 mm f/3.5–5.6 lens. An X-Rite Mini ColorChecker Classic marker (http://xritephoto.com/colorchecker-targets) was placed in the frame to assess the scale and color calibration. The spikes were shot in four projections. The first projection corresponded to the front (widest) side of the spike, and the remaining projections were obtained by rotating the spike on the clothespin 90°: the second and fourth projections were the side projections, and the third projection was the back projection.

Each spike image was segmented into background, color palette, spike body, and awns using a deep machine learning method developed by us earlier (Artemenko et al., 2024). The obtained masks were used by the WERecognizer program (Genaev et al., 2019) to calculate the morphometric characteristics of the spikes. We considered a reduced set of characters, which included a description of the spike based on a model of two quadrilaterals symmetrized with respect to the axis of the spike, characteristics of the contour of the spike (perimeter, area, roundness, etc.), and the area of the awns in the image (Komyshev et al., 2024). The diagram of the quadrilateral model and the parameters of the spike used are provided in the Supplementary Materials (Fig. S1). In total, 19 characters of the size and shape per spikes were analyzed independently for each projection (Table S2).

Statistical analysis of the signs of colossus. In our work, we assumed that the population of F_2 hybrid plants can be divided into classes not only based on qualitative (discrete) characteristics (Table 1), but also in quantitative terms. In the second case, it was assumed that each class of plants is characterized by a distribution of the magnitude of the trait with a different expectation and variance. Thus, the distribution of the character value for the entire accessions is a mixture of two Gaussian distributions with different means and variances (Merezhko, 2005; Rechkin, 2024). It should be noted that the method of Gaussian distribution mixture is used in plant morphometry to identify groups that differ in character values, as well as for taxonomy (Kim et al., 2024; Tiburtini et al., 2025).

We used the sklearn.mixture.GaussianMixture function from the scikit-learn v. 1.7 package (https://scikit-learn.org/stable/index.html) to identify mixtures of two Gaussian distributions. This function allows us to analyze a mixture model of Gaussian distributions using the maximum likelihood method, estimating the number and parameters of the distributions in the mixture, and classifying accessions in the dataset based on their belonging to these distributions. In this study, we assumed that the distribution of each character was a mixture of two distributions for each projection of the spike. Using the sklearn.mixture.GaussianMixture function, the mean values, variances, and the number of ears belonging to each distribution were estimated. After determining the number of spikes in each class, a Pearson

[1] Supplementary Tables S1–S3 and Figure S1 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Kruch_Engl_30_3.pdf

Table 1. The phenotypic description of the parent forms

Trait	Species/accession	
	♀ *T. aethiopicum* (k-19301/2)	♂ *T. carthlicum* (k-32496)
Tetraaristatum (the presence of awns on both the flower and spike glume), *ta*	–	+
Color of the spike	–	+
Character (type) of awedness	The awns are directed sideways	The awns are directed vertically
Teeth at the top of the spike glume	+	+
Hairy glume, *Hg*	–	+
Baldness in the trichomes on the spike glume, Hg^C	–	+/–

Note. Trait: – recessive, + dominant.

χ^2 test was performed to check for compliance with the theoretically expected ratio. It was assumed that the observed and AI-generated splits corresponded to the spike traits if the latter was reliable for all four spike projections. The Past v 5.3 package (Hammer Ø. et al., 2001) was used for statistical processing (tests for equality of means and analysis of variance) and data visualization.

Results

Examples of images of the spikes of the parent species, F_1, and F_2 hybrids (projection 1) are shown in Figure 1.

A principal component analysis (PCA) was performed for four projections to assess the diversity of the "size" and "shape" of the spikes. The results are presented in Figure 2. Of the 76 variables used for the analysis, the first two components accounted for 45 % of the variance, with 35 % accounted for by PC1 and 10 % accounted for by PC2 (Fig. 2). The main contributors to the variability were the traits related to the size of the ear (its length and area). The second principal component of variability is related to the shape of the spike and reflects its roundness. The diagram (Fig. 2) shows that the spikes of the mother, father, and F_1 hybrid plants are compactly arranged, occupying partially overlapping areas and demonstrating the similarity of the spike shape/size (Fig. 1*a–c*). The F_2 hybrid plants exhibit significantly greater variability in spike characteristics (primarily in size). It contains a large number of plants with a spike size larger than the average for the first three lines.

In the second stage of our research, we analyzed the F_2 hybrid segregation based on the species-specific characteristics of the spikes (Table 1). For each characteristics of

Fig. 1. The spikes of the parent species, F_1 and F_2 hybrids, presented in the same scale and projection.
The mother *T. aethiopicum* (k-19301/2) (*a*) and the father *T. carthlicum* (k-32496) (*b*) plants; the F_1 (*c*) and F_2 (*d*) hybrid plants.

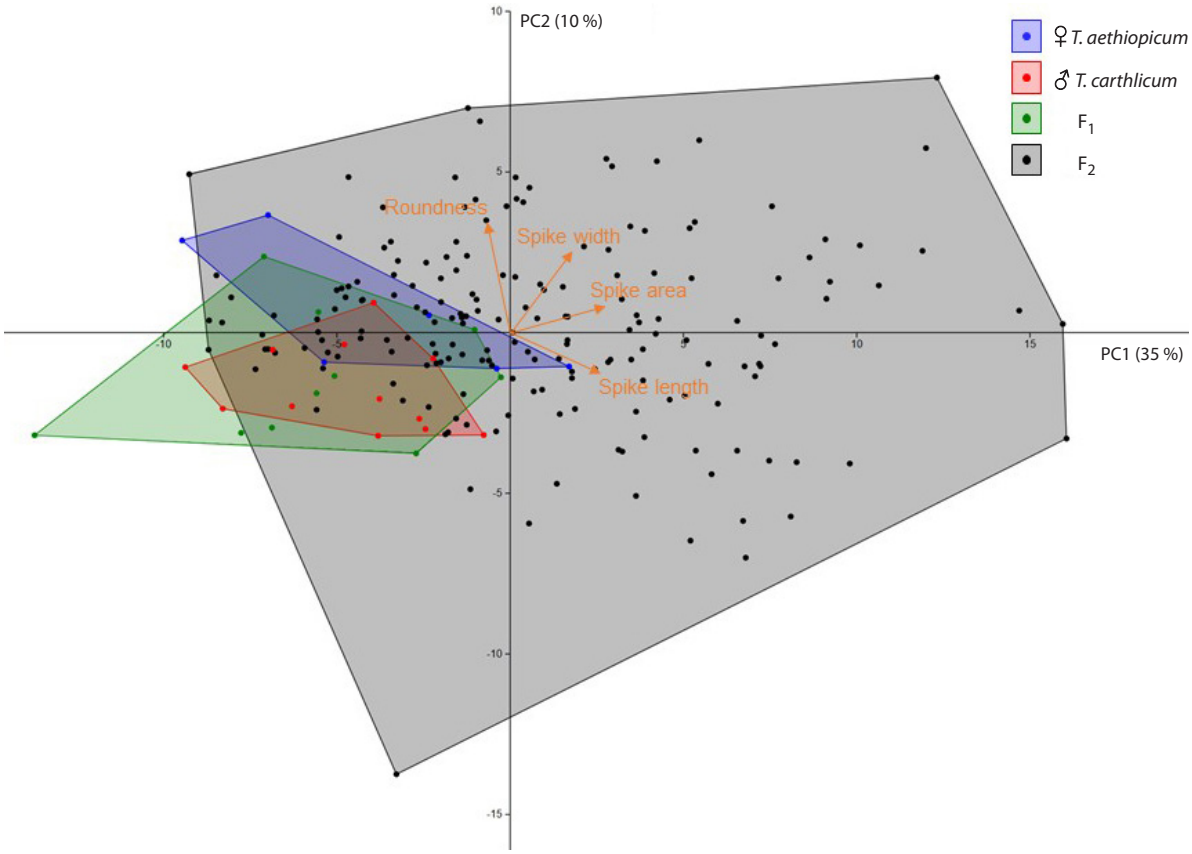


Fig. 2. Scatter plot in the space of the first two principal components (PC1, PC2) for the spikes, whose size and shape parameters were independently estimated for the four projections. The symbols are located in the upper right corner. The proportions of the variance accounted for by the first (PC1, horizontal axis) and second (PC2, vertical axis) principal components are given in parentheses. The averaged directions of the projection of the main groups of traits of the spike are indicated by orange arrows.

traits, we evaluated the reliability of the segregation in the population using proportions that reflect the type of genetic control (3:1, 13:3, 15:1, 61:3, and 63:1) the value of the Pearson χ^2 criterion was calculated (Table 2).

Based on the results presented in Table 2, it can be concluded that tetraaristatum and the character (type) of awness are split in F_2 hybrids in a 3:1 ratio (monogenic inheritance), the presence of spike colored in a 13:3 ratio (digenic inheritance with incomplete dominance), and the teeth of awn in a 15:1 ratio (digenic inheritance). For such traits as the hairy glume and the baldness in the trichomes on the spike glume, no significant similarity was found for the F_2 hybrids.

We evaluated 19 spike traits per each F_2 hybrid plants that are of interest for identifying those that may produce a split corresponding to specific types of inheritance. We used a mixture model of two Gaussian (normal) distributions to investigate this issue. Each projection was evaluated independently. Since a reliable estimate of the Gaussian distribution parameters depends on the accession size (the number of accessions studied), we decided to limit ourselves to testing the 3:1 ratio in this work. For a total samples of 187 spikes, the theoretically expected number of plant groups is 140 to 47. Table S3 shows the mean values and variances for the first and second group of ears for each character and projection (columns mean1, var1, mean2, var2), the estimated number of spikes in the first and second group (num1, num2), the ratio of the number of spikes (numratio_1vs2), the value of χ^2, and the significance level (chisq, chisq_p). It can be seen from the results presented in Table S3 that for two to seven traits in each projection, the ratio of the number of spikes in the two groups is close to 3:1. For example, for the first projection, these traits are the area of the awns (SAA) and the integrity (SSO) of the spike. For the second projection, the width of the spike model and its area (q_ym, q_S), the perimeter of the spike projection contour (SP), its area (SA), the area of the awns (SAA), and the roundness (SRO), etc. It is important to note that only for one character, namely, the "area of the awns" on the image (SAA), a significant split in the ratio of approximately 3:1 is observed for all four projections of the spike. The characteristics of the split for this trait for

Table 2. Values of the χ^2 statistic for different proportions of occurrence of species-specific spike traits in F_2 hybrids of *T. aethiopicum* and *T. carthlicum*

Traits with categories	Number of plants in categories 1, 2		Type of inheritance and corresponding them χ^2 values				
	1	2	3:1	13:3	15:1	61:3	63:1
Tetraaristatum, no (1), yes (2)	125	54	**2.549**	15.317	174.758	260.114	952.268
Colored spike, no (1), yes (2)	28	144	6.977	**0.689**	29.526	51.728	242.193
Character (type) of awness, *T. aethiopicum* (1), *T. carthlicum* (2)	46	133	**0.047**	5.673	115.549	176.868	677.948
Teeth of awn, no (1), yes (2)	10	175	37.883	21.626	**0.225**	**0.213**	17.763
Hairy glume, no (1), yes (2)	69	116	14.921	41.774	304.346	440.328	1,535.938
Baldness in the trichomes on the spike glume, no (1), yes (2)	120	65	10.135	32.602	263.432	383.873	1,355.695

Note. The left column lists the spike characters and the corresponding category numbers (in parentheses). $\chi^2_{0.05} = 3.84$. The values of χ^2 ratios that do not significantly differ from the corresponding proportions are highlighted in bold.

Table 3. Characteristics of spike splitting in F_2 hybrids based on the "area of awns" in the image (SAA)
For each spike projection, the estimates of the mean values (SAA_1 and SAA_2) and variances for the two groups of spikes (*Var*(SAA_1), *Var*(SAA_2)) identified based on the mixture model of two Gaussian distributions are given.

Projection of the spike image	SAA_1	*Var*(SAA_1)	SAA_2	*Var*(SAA_2)	N_1	N_2	χ^2	*p*
1	792.3	151,092.6	244.6	32,393.1	41	146	0.80	0.37
2	574.1	119,803.5	147.5	12,517.7	42	145	0.53	0.47
3	759.0	136,893.6	229.9	30,079.3	48	139	0.09	0.76
4	578.4	127,534.0	133.5	11,256.2	51	136	0.66	0.42

Note. N_1, N_2 are number of accessions of groups of spikes, and *p* is the probability.

four projections and estimates of the number of spikes for two groups, χ^2, and the probability of the 3:1 monogenic inheritance ratio hypothesis are shown in Table 3.

From the results presented in Table 3, it can be seen that the average values of the spike area are slightly larger for the projection of the front (1) and back (3) sides than for their two lateral projections (2, 4), for both the first and second studied groups. For the group of spikes with a larger area of spikes (and a large number of them, group 1), their number is smaller and varies between 41–51. For a group of plants with a smaller area of awness, the number of spikes varies from 136 to 146. The example of a histogram of the distribution of SAA values for the 4th projection, showing two Gaussian distributions, is shown in Figure 3*a*.

The histogram area is visually divided into two parts: the right part, with fewer plants and a higher SAA value, and the left part, with lower SAA values (more plants) (Fig. 3*a*). It is in good agreement with the data presented in Table 3. The data presented in Table 3 and in Figure 3*a* are in agreement with the results of the manual assessment of splitting of spikes in F_2 hybrids on the basis of "tetraaristatum" (Table 2). Although, based only on such a parameter as the area of the spikes, it is impossible to determine the tetraaristatum of the spike (the presence of spikes simultaneously on the flower and on spikelet glimes), tetraaristatum spikes generally have a larger number of spines, which leads to an increase in their SAA parameter. This is confirmed by the shift in the area of the awns in tetraploid spikes towards higher values compared to conventional awned spikes (Fig. 3*b*). We also calculated the average values of the SAA parameter for the four projections of the spikes in the images of the F_2 hybrids and performed a test for the equality of the means in three groups of plants: awnless, awned, and tetraaristatum spikes (Table 4).

Note that the average value of the spinous area is the smallest in spineless plants, the maximum – in tetraaristatum plants, and the intermediate value – in spinous plants, in accordance with the distribution results (Table 4), shown in Figure 3*b*. The analysis of variance demonstrates significant differences in the averages for the three groups of spikes ($p < 10^{-4}$). A pairwise comparison of averages using the Mann–Whitney paired test also shows significant dif-

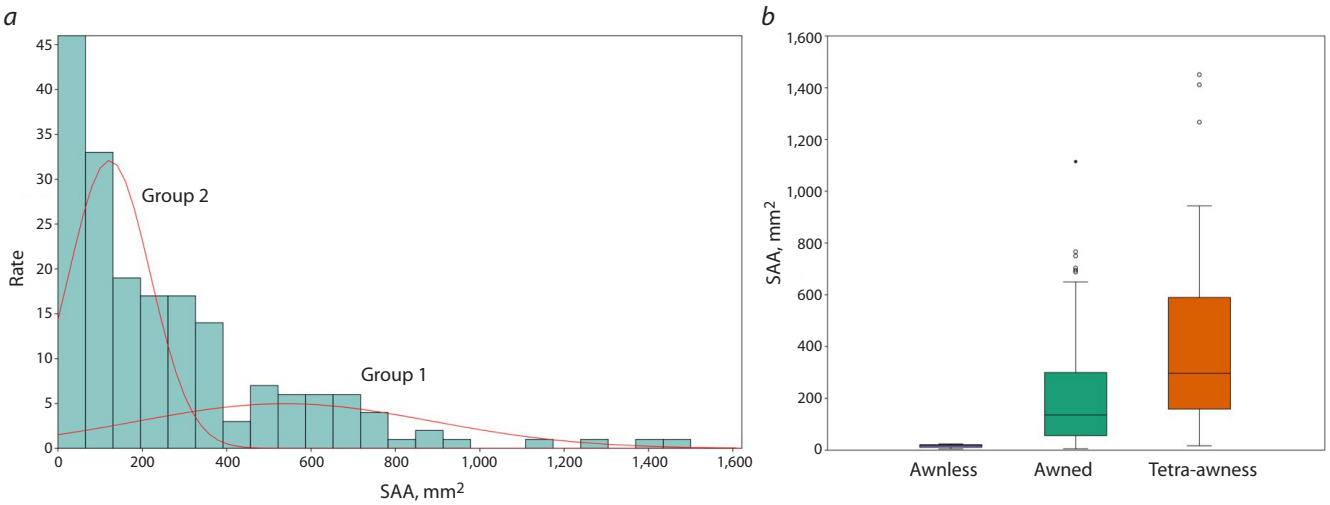


Fig. 3. Distribution of the area of the awns. *a* – histogram of the distribution of the area of the awns on the image (SAA) of the spikes in F_2 hybrids (projection No. 4). The X axis shows the value of the trait, and the Y axis shows the number of spikes. The red lines show the probability density distributions for two groups of spikes, namely, those with more (group 2) and fewer (group 1) awns. *b* – distribution of the area of the awns on the image of the spikes in the 4th projection for the awnless, awned, and tetraaristatum spikes, shown as diagrams. The X axis shows the types of awns, and the Y axis shows the area of the awns.

Table 4. Comparison of the statistics of the distributions of the spikes by the average value of the area of the awns (mm^2) in four projections

Statistics	Awnless spikes	Awned spikes	Tetraaristatum spikes
Number of ears	6	127	54
SAA	13.79	280.13	481.03
Var(SAA)	25.59	73,082.20	129,227.80
Std. deviation	5.06	270.34	359.48
Std. error	2.07	23.99	48.92

Note. SAA is the area of the spikes in the image; *Var* is the variance.

ferences between all three groups of spikes on the basis of "tip area" ($p<0.001$). Thus, the splitting of the ears in F_2 hybrids by the area of the awns is due to the fact that the tetraaristatum spikes have a larger area of the awns on average (Fig. 3*b*).

As for other characters of the spike, which demonstrate a monogenic 3:1 inheritance not for all projections, it is higher for the character associated with the area of the spike contour in the image (SA), which gives a split for three projections out of four. The perimeter of the spike contour (SP), its length (SL), the area of the quadrilateral model (q_S), and the length (q_L) demonstrate a close split to the 3:1 proportion for two projections of the spike out of four. It is possible that these characters are also controlled by a single gene, but the inaccuracy (ambiguity) of defining the parameters of the spike characters in the image may result in incomplete correspondence for different projections, making analysis more difficult.

Discussion

The most important result of this investigation is the answer to the question of how much the results obtained by different methods, namely, hybridological analysis and machine learning, coincide, and whether it is possible to judge the stage at which machine learning methods can be used in the context of working with species-specific characters.

Analysis of tetraaristatum. Among the studied hybrid plants, we obtain a split of 125 (normal) to 54 (tetraaristatum), χ^2 3:1 = 2.549, $p<0.05$. The results are consistent with the hypothesis of monogenic inheritance of the tetraaristatum in a recessive manner. Earlier, in the works of E.F. Migushova and P.M. Zhukovsky (1969), M.A. Haque et al. (2011), R.V. Rozhkov (2014), and O.B. Dobrovolskaya et al. (2020), the recessive nature of the inheritance of this trait has been shown. The gene controlling the tetraaristatum trait, *ta*, is located in the long arm of chromosome 5A

(Haque et al., 2011). Despite the fact that V.F. Dorofeev et al. (1979) noted the presence of tetraploid spike forms, in addition to *T. carthlicum*, in a number of subvarieties of the tetraploid species *T. aethiopicum*, most authors agree that the trait "tetraaristatum" is species-specific only for one tetraploid wheat species, namely for *T. carthlicum* (Haque et al., 2011; Goncharov, 2012; Dobrovolskaya et al., 2020).

In addition to tetraploid species, a trait phenotypically similar to tetraaristatum is characteristic of the hexaploid wheat species *T. aestivum* ssp. *petropavlovskyi* (Udacz. et Migusch.), however, in this case, it is a slightly elongated awn-like appendages on the spike glume (Goncharov, 2009). Nevertheless, we can argue that "tetraaristatum" is a species-specific trait only for *T. carthlicum* and can be effectively used for taxonomic purposes.

Color analysis of the spike. Among the studied hybrid plants, we obtain a cleavage of 144 (black + red spike) to 28 (white spike), χ^2 13:3 = 0.689, $p < 0.05$. These results illustrate the phenomenon of dominant epistasis. The genes controlling spike coloration, *Bg* and *Rg1*, are located in chromosome 1AS, while *Rg2* is located in chromosome 2BS. The digenic control of the red spike in wheat has been previously shown by a number of authors (Sobko, Sozinov, 1993; Kudryavtsev, Popova, 1994; among others).

Analysis of the type (character) of the awnedness. In the studied combination of F_2 hybrids, there was a split of 46 plants with "*T. aethiopicum*-type" awns to 133 plants with "*T. carthlicum*-type" awns, which is consistent with the hypothesis of monogenic inheritance by the recessive type ($\chi^2 = 0.047$). The "awnedness type" trait is inherited monogenic, and the gene controlling it is located on chromosome 3A (Goncharov et al., 2003). The variation observed in this investigation in the severity of the "awnedness" trait (Fig. 3) is due to the presence of a significant number of modifier genes that control its severity (Sourdille et al., 2002; Wang et al., 2019). It is known that modifiers lead to "partial awnedness" (semi-awnedness, etc.) or changes in the length of the awns (from short to very long, exceeding the length of the spike) in some wheat varieties.

Analysis of the shape of the teeth on the spike glime. The presence/absence of a teeth on the spike glume is characteristic of all tetraploid wheat species, and the trait is not species-specific. However, it may have different degrees of expression in different varieties. According to the data obtained, the 175 (teeth present) to 10 (teeth absent) ratio corresponds to the 15:1 dihybrid inheritance hypothesis ($\chi^2 = 0.225$). The "Wheat manual book" (1980) states that both *T. aethiopicum* and *T. carthlicum* have a teeth, and that the shape of the teeth varies from short to long, and from blunt to sharp. However, only *T. aethiopicum* has a teeth that can be so long that it becomes an awn-like appendages.

Analysis of the hairy glume and the baldness in the trichomes on the spike glume. The hairy glume is an important taxonomic trait (Goncharov et al., 2007). It is controlled in tetraploid wheat species by the *Hg* gene, which is located in chromosome 1AS (Goncharov, 2012). In addition, for *T. carthlicum*, the absence of continuous hairy glume is shown, namely, the baldness in the trichomes on the spike glume in its lower part (Gandilian, 1972). Due to the different types of hairy glume, for the species *T. carthlicum*, trichomes with a baldness is a species-specific trait, and for their distinction, we have assigned the symbol "*C*" to the *Hg* gene according to the species symbols (Hg^C). Considering these traits, we did not obtain reliable results for the 3:1 inheritance hypothesis, so additional research is required, as the severity is difficult to assess both visually and using AI.

Nevertheless, digital methods allowed us to evaluate the diversity of ears by 19 traits. And a much greater diversity of them in F_2 hybrid plants was demonstrated than in the parental and F_1 ones. It may be related to the presence of modifier genes (Bersimbaev, Shulembaeva, 2014). With the species-specific traits studied by us (Table 1), tetraaristatum, only one of the quantitative characters can be directly associated – the area of the awns in the image (SAA). The remaining 18 traits characterize the shape of the spike and its size and are not related to the color and glume traits. Our analysis of the mixture of Gaussian distributions showed that the F_2 hybrid plants exhibit a 3:1 split into two classes based on the area of the awns, which is consistent with the tetraaristatum trait. Similar effects of population splitting into classes based on grain mass distribution have been described in triticale ×*Triticosecale* Wittm. ex A. Camus. (= syn. ×*Triticale* Tscherm.-Seys. ex Müntzing) (Kim et al., 2024). It should be noted that triticale is the result of intergeneric hybridization between wheat *Triticum* spp. and rye *Secale cereale* L. The splitting of different triticale varieties by grain weight was observed on different days of maturity (2–5 days after earing) for different seeding densities (150–300 kg/ha). However, the authors did not check the ratio of the number of plants in the grain weight classes. Our investigation demonstrates that a combination of automatic phenotyping methods and a mixture types of Gaussian distributions can, in principle, allow for automatic analysis of the separation of classes in F_2 hybrids that differ in the values of phenotypes and exhibit a split depending on the inheritance model. This, in turn, can help identify genes associated with species-specific wheat plant traits.

Conclusion

Based on the produced research, we can state that "tetraaristatum" is a species-specific traits only for *T. carthlicum* and can be effectively used for taxonomic, as well as in experiments using machine learning. The same applies to the nature of awnedness for *T. aethiopicum*.

The search for algorithms to solve the problem of automated classification of phenotypes in hybrid populations

into classes continues (Merezhko, 2005; Rechkin, 2024). The method proposed in this investigation of using two Gaussian (normal) characters is effective for class distribution in oligogenic (mono- and di-genic) inheritance of traits.

References

Afonnikov D.A., Genaev M.A., Doroshkov A.V., Komyshev E.G., Pshenichnikova T.A. Methods of high-throughput plant phenotyping for large-scale breeding and genetic experiments. *Russ J Genet*. 2016;52(7):688-701. doi 10.1134/S1022795416070024

Artemenko N.V., Genaev M.A., Epifanov R.U., Komyshev E.G., Kruchinina Y.V., Koval V.S., Goncharov N.P., Afonnikov D.A. Image-based classification of wheat spikes by glume pubescence using convolutional neural networks. *Front Plant Sci.* 2024;14:1336192. doi 10.3389/fpls.2023.1336192

Awada L., Phillips P.W.B., Bodan A.M. The evolution of plant phenomics: global insights, trends, and collaborations (2000-2021). *Front Plant Sci*. 2024;15:1410738. doi 10.3389/fpls.2024.1410738

Bersimbaev R.I., Shulembaeva K.K. Cytogenetic studies of common wheat in Kazakhstan. *Informatsionnyy Vestnik VOGiS*. 2005; 9(3):317-323 (in Russian)

Bukasov S.M. Potato species system of section *Tuberarium* (Dum.) Buk. genus *Solanim* L. *Proc Appl Bot Genet Breed.* 1971;46(1):3-44 (in Russian)

Chouhan S.S., Singh U.P., Sharma U., Jain S. Classification of different plant species using deep learning and machine learning algorithms. *Wireless Pers Commun*. 2024;136(4):2275-2298. doi 10.1007/s11277-024-11374-y

Dobrovolskaya O.B., Dresvyannikova A.E., Badaeva E.D., Popova K.I., Travnickova M., Martinek P. The study of genetic factors that determine the awned glume trait in bread wheat. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2020;24(6):568-574. doi 10.18699/VJ20.650 (in Russian)

Dolezel J., Greilhuber J., Suda J. Estimation of nuclear DNA content in plants using flow cytometry. *Nat Protoc*. 2007;2(9):2233-2244. doi 10.1038/nprot.2007.310

Dorofeev V.F. Cultural flora of the USSR. Vol. 1. Wheat. Leningrad, Kolos, 1979 (in Russian)

Eig A. Monographisch-kritische Ubersicht der Gattung *Aegilops*. Repertorium specierum novarum regni vegetabilis. Beihefte, Berlin, 1929

Gandilian P.A. Spontaneous hybridization, mutations, and phylogeny issues in wheat. *Genetika* (*Moscow*). 1972;8(8):5-19 (in Russian)

Genaev M.A., Komyshev E.G., Fu Hao, Koval V.S., Goncharov N.P., Afonnikov D.A. SpikeDroidDB: an information system for annotation of morphometric characteristics of wheat spike. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2018; 22(1):132-140. doi 10.18699/VJ18.340 (in Russian)

Genaev M.A., Komyshev E.G., Kruchinina Y.V., Goncharov N.P., Afonnikov D.A., Smirnov N.V. Morphometry of the wheat spike by analyzing 2D images. *Agronomy*. 2019;9(7):390. doi 10.3390/agronomy9070390

Golovnina K.A., Glushkov S.A., Blinov A.G., Mayorov V.I., Adkison L.R., Goncharov N.P. Molecular phylogeny of the genus *Triticum* L. *Plant Syst Evol*. 2007;264:195-216. doi 10.1007/s00606-006-0478-x

Goncharov N.P., Mitina R.L., Anfilova N.A. Inheritance of awnlessness in tetraploid wheat species. *Russ J Genet*. 2003;39(4):463-466. doi 10.1023/A:1023326202320

Goncharov N.P., Bannikova S.V., Kawahara T. Wheat artificial amphiploids involving *Triticum timopheevii* genome: their studies, preservation and reproduction. *Genet Resour Crop Evol*. 2007;54(7): 1507-1516. doi 10.1007/s10722-006-9141-1

Goncharov N.P. Manual book of common and durum wheat. Novosibirsk: RAS SB Publ. House, 2009 (in Russian)

Goncharov N.P. Genus *Triticum* L. taxonomy: The present and the future. *Plant Syst Evol*. 2011;295(1):1-11. doi 10.1007/s00606-011-0480-9

Goncharov N.P. Comparative genetics of wheat and its relatives. Novosibirsk: Geo Publ., 2012 (in Russian)

Goncharov N.P., Adonina I.G. A new hexaploid wheat species *Triticum aminovii*. *Cereal Res Comm*. 2025;53:2159-2165. doi 10.1007/s42976-025-00677-w

Hammer K., Filatenko A.A., Pistrick K. Taxonomic remarks on *Triticum* L. and ×*Triticosecale* Wittm. *Genet Resour Crop Evol*. 2011; 58:3-10. doi 10.1007/s10722-010-9590-4

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

Haque M.A., Takayama A., Watanabe N., Kuboyama T. Cytological and genetic mapping of the gene for four-awned phenotype in *Triticum carthlicum* Nevski. *Genet Resour Crop Evol*. 2011;58:1087-1093. doi 10.1007/s10722-010-9644-7

Hawkes J. A revision of the tuber-bearing Solanums. In: Record Scottish Plant Breeding Station. 1963;76-181

Hodač L., Karbstein K., Tomasello S., Wäldchen J., Bradican J.P., Hörandl E. Geometric morphometric versus genomic patterns in a large polyploid plant species complex. *Biology*. 2023;12(3):418. doi 10.3390/biology12030418

Kim B.H., Kwon H., Kim W. Deciphering individual triticale grain weight patterns: A gaussian mixture model approach. *PloS One*. 2024;19(11):e0313942. doi 10.1371/journal.pone.0313942

Komyshev E.G., Genaev M.A., Kruchinina Yu.V., Koval V.S., Goncharov N.P., Afonnikov D.A. Evaluation of the spike diversity of seven hexaploid wheat species and an artificial amphidiploid using a quadrangle model obtained from 2D images. *Plants*. 2024;13(19): 2736. doi 10.3390/plants13192736

Kudryavtsev A.M., Popova T.A. Genetic linkage between gliadin-coding genes and genes for color and pubescence of the ear in spring durum wheat (*Triticum durum* Desf.). *Genetika* (*Moscow*). 1994; 30(12):1587-1592 (in Russian)

Lyapunova O.A. Intraspecific diversity of durum wheat (*Triticum durum* Desf.): a unified classification. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2021;25(3):260-268. doi 10.18699/VJ21.029

Martinek P., Bednar J. Changes of spike morphology (multirow spike – MRS, long glumes – LG) in wheat (*Triticum aestivum* L.) and their importance for breeding. In: Proc. Intern. Conf. "Genetic collections, isogenic and alloplasmic lines". Novosibirsk, 2001;192-194

Merezhko A.F. The use of Mendelian principles in computer analysis of the inheritance of varying traits. In: Ecological genetics of cultivated plants. Krasnodar: All-Russian Research Institute of Rice Publ., 2005;107-117 (in Russian)

Migushova E.F., Zhukovsky P.M. Towards the knowledge of wheat *T. ispahanicum* Heslot. *Proc Appl Bot Genet Breed*. 1969;39(3): 71-90 (in Russian)

Mulugeta A.K., Sharma D.P., Mesfin A.H. Deep learning for medicinal plant species classification and recognition: a systematic review. *Front Plant Sci*. 2024;14:1286088. doi 10.3389/fpls.2023.1286088

Pronozin A.Yu., Paulish A.A., Zavarzin E.A., Prikhodko A.Yu., Prokhoshin N.M., Kruchinina Y.V., Goncharov N.P., Komyshev E.G., Genaev M.A. Automatic phenotyping of ear morphology of tetra- and hexaploid wheat species by computer vision methods. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2021;25(1): 71-81. doi 10.18699/VJ21.009

Ran Z., Xiao X., Zhou L., Yan C., Bai X., Ou J., Li Z. Phenotypic diversity analysis in the Sect. *Tuberculate* (*Camellia* L.) population, an endemic taxon in China. *Plants*. 2024;13(22):3210. doi 10.3390/plants13223210

Rechkin D.V. Analysis of quantitative traits segregation based on fundamental properties of the normal distribution. *Pisma v Vavilovskii Zhurnal Genetiki i Selektsii = Lett Vavilov J Genet Breed*. 2024; 10(4):199-203. doi 10.18699/letvjgb-2024-10-25 (in Russian)

Rozhkov R.V., Krivoruchenko R.V., Kovalenko I.V. Genetic control of the awned glume trait in wheat. *Vestnik Khar'kovskogo Natsional'nogo Agrarnogo Universiteta = Herald of the Kharkiv National Agrarian University*. 2014;2(32):70-76 (in Russian)

Sobko T.A., Sozinov A.A. Genetic control of morphological features of the ear and the relationship of allelic variability of marker loci of chromosomes 1A and 1B of winter soft wheat. *Tsitologiya i Genetika = Cytology and Genetics*. 1993;27(5):15-22 (in Russian)

Sourdille P., Cadalen T., Gay G., Bernard M. Molecular and physical mapping of genes affecting awning in wheat. *Plant Breeding*. 2002; 121(4):320-324. doi 10.1046/j.1439-0523.2002.728336.x

Tiburtini M., Scrucca L., Peruzzi L. Using Gaussian Mixture Models in plant morphometrics. *Perspect Plant Ecol Evol Syst*. 2025;69:125902. doi 10.1016/j.ppees.2025.125902

Wang D., Yu K., Jin D., Sun L., Chu J., Wu W., Xin P., Gregová E., Li X., Sun J., Yang W., Zhan K., Zhang A., Liu D. Natural variations in the promoter of Awn Length Inhibitor 1 (ALI-1) are associated with awn elongation and grain length in common wheat. *Plant J*. 2019;101(5):1075-1090. doi 10.1111/tpj.14575

Wheat manual book / Eds V.F. Dorofeev, A.A. Filatenko, E.F. Migushova. Leningrad: VIR, 1980 (in Russian)

Zhukovsky P.M. A critical and systematic review of the species of the genus *Aegilops*. *Proc Appl Bot Genet Breed*. 1928;18(1):417-609 (in Russian)

Zuev E.V., Brykova A.N., Kudryavtseva E.Yu. Results of analyzing the passport database 'Spring bread wheat landraces in the VIR collection'. *Proc Appl Bot Genet Breed*. 2019;180(1):7-11. doi 10.30901/2227-8834-2019-1-7-11 (in Russian)

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

Received November 24, 2025. Revised December 22, 2025. Accepted December 22, 2025.

Прием статей через электронную редакцию на сайте http://vavilov.elpub.ru/index.php/jour
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» до 2011 г. выходил под названием «Информационный вестник ВОГиС»/ "The Herald of Vavilov Society for Geneticists and Breeding Scientists".

Сетевое издание «Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» – реестровая запись СМИ Эл № ФС77-85772, зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 14 августа 2023 г.

Издание включено ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Белый список (уровень 1), Russian Science Citation Index, Российский индекс научного цитирования, ВИНИТИ, Web of Science CC, Scopus, PubMed Central, DOAJ, ROAD, Ulrich's Periodicals Directory, Google Scholar.

Открытый доступ к полным текстам:
русскоязычная версия – на сайте https://vavilovj-icg.ru/
и платформе Научной электронной библиотеки, elibrary.ru/title_about.asp?id=32440

англоязычная версия – на сайте vavilov.elpub.ru/index.php/jour
и платформе PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/journals/3805/

✉ email: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.
Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218.
Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.
Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.

Дата выхода в свет 21.05.2026. Формат 60 × 84 1/8. Уч.-изд. л. 26.6.