

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-26-56***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

Главный редактор

А.В. Кочетов – академик РАН, д-р биол. наук, профессор РАН (Россия)

Заместители главного редактора

Н.А. Колчанов – академик РАН, д-р биол. наук, профессор (Россия)

И.Н. Леонова – д-р биол. наук (Россия)

Н.Б. Рубцов – д-р биол. наук, профессор (Россия)

Ответственный секретарь

Г.В. Орлова – канд. биол. наук (Россия)

Редакционная коллегия

Е.Е. Андронов – д-р биол. наук (Россия)

Ю.С. Аульченко – д-р биол. наук (Нидерланды)

О.С. Афанасенко – академик РАН, д-р биол. наук (Россия)

Д.А. Афонников – д-р биол. наук, доцент (Россия)

Л.И. Афтанас – академик РАН, д-р мед. наук (Россия)

Л.А. Беспалова – академик РАН, д-р с.-х. наук (Россия)

А. Бёрнер – д-р наук (Германия)

Н.П. Бондарь – канд. биол. наук (Россия)

С.А. Боринская – д-р биол. наук (Россия)

П.М. Бородин – д-р биол. наук, проф. (Россия)

А.В. Васильев – чл.-кор. РАН, д-р биол. наук (Россия)

М.И. Воевода – академик РАН, д-р мед. наук (Россия)

Т.А. Гавриленко – д-р биол. наук (Россия)

Н.Е. Грунтенко – д-р биол. наук (Россия)

С.А. Демаков – д-р биол. наук (Россия)

И.К. Захаров – д-р биол. наук, проф. (Россия)

И.А. Захаров-Гезехус – чл.-кор. РАН, д-р биол. наук (Россия)

С.Г. Инге-Вечтомов – академик РАН, д-р биол. наук (Россия)

А.В. Кильчевский – чл.-кор. НАНБ, д-р биол. наук (Беларусь)

А.М. Кудрявцев – чл.-кор. РАН, д-р биол. наук (Россия)

Д.М. Ларкин – канд. биол. наук (Великобритания)

С.А. Лашин – д-р биол. наук (Россия)

Ж. Ле Гуи – д-р наук (Франция)

И.Н. Лебедев – чл.-кор. РАН, д-р биол. наук, проф. (Россия)

Л.А. Лутова – д-р биол. наук, проф. (Россия)

Б. Люгтенберг – д-р наук, проф. (Нидерланды)

В.Ю. Макеев – чл.-кор. РАН, д-р физ.-мат. наук (Россия)

И.В. Максимов – д-р биол. наук (Россия)

Б.А. Малярчук – д-р биол. наук (Россия)

Ю.Г. Матушкин – канд. биол. наук (Россия)

В.И. Молодин – академик РАН, д-р ист. наук (Россия)

М.П. Мошкин – д-р биол. наук, проф. (Россия)

Л.Ю. Новикова – д-р с.-х. наук (Россия)

Е.К. Потокина – д-р биол. наук (Россия)

В.П. Пузырев – академик РАН, д-р мед. наук (Россия)

Д.В. Пышный – чл.-кор. РАН, д-р хим. наук (Россия)

Е.Ю. Рыкова – д-р биол. наук (Россия)

Е.А. Салина – чл.-кор. РАН, д-р биол. наук, проф. (Россия)

В.А. Степанов – академик РАН, д-р биол. наук (Россия)

И.А. Тихонович – академик РАН, д-р биол. наук (Россия)

Е.К. Хлесткина – чл.-кор. РАН, д-р биол. наук, проф. РАН (Россия)

Э.К. Хуснутдинова – д-р биол. наук, проф. (Россия)

М. Чен – д-р биол. наук (Китайская Народная Республика)

Е.В. Шахтшнейдер – д-р мед. наук (Россия)

С.В. Шестаков – академик РАН, д-р биол. наук (Россия)

С.В. Шеховцов – д-р биол. наук (Россия)

Н.С. Юдин – канд. биол. наук (Россия)

Н.К. Янковский – академик РАН, д-р биол. наук (Россия)

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-26-56

Founders

Siberian Branch of the Russian Academy of Sciences

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

The Vavilov Society of Geneticists and Breeders

Editor-in-Chief

A.V. Kochetov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Professor of the RAS, Russia

Deputy Editor-in-Chief

N.A. Kolchanov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

I.N. Leonova, Dr. Sci. (Biology), Russia

N.B. Rubtsov, Professor, Dr. Sci. (Biology), Russia

Executive Secretary

G.V. Orlova, Cand. Sci. (Biology), Russia

Editorial board

O.S. Afanasenko, Full Member of the RAS, Dr. Sci. (Biology), Russia

D.A. Afonnikov, Associate Professor, Dr. Sci. (Biology), Russia

L.I. Aftanas, Full Member of the RAS, Dr. Sci. (Medicine), Russia

E.E. Andronov, Dr. Sci. (Biology), Russia

Yu.S. Aulchenko, Dr. Sci. (Biology), The Netherlands

L.A. Bepalova, Full Member of the RAS, Dr. Sci. (Agricul.), Russia

N.P. Bondar, Cand. Sci. (Biology), Russia

S.A. Borinskaya, Dr. Sci. (Biology), Russia

P.M. Borodin, Professor, Dr. Sci. (Biology), Russia

A. Börner, Dr. Sci., Germany

M. Chen, Dr. Sci. (Biology), People's Republic of China

S.A. Demakov, Dr. Sci. (Biology), Russia

T.A. Gavrilenko, Dr. Sci. (Biology), Russia

N.E. Gruntenko, Dr. Sci. (Biology), Russia

S.G. Inge-Vechtomov, Full Member of the RAS, Dr. Sci. (Biology), Russia

E.K. Khlestkina, Corr. Member of the RAS, Professor of the RAS,
Dr. Sci. (Biology), Russia

E.K. Khusnutdinova, Professor, Dr. Sci. (Biology), Russia

A.V. Kilchevsky, Corr. Member of the NAS of Belarus, Dr. Sci. (Biology),
Belarus

A.M. Kudryavtsev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

D.M. Larkin, Cand. Sci. (Biology), Great Britain

S.A. Lashin, Dr. Sci. (Biology), Russia

J. Le Gouis, Dr. Sci., France

I.N. Lebedev, Corr. Member of the RAS, Professor, Dr. Sci. (Biology), Russia

B. Lugtenberg, Professor, Dr. Sci., Netherlands

L.A. Lutova, Professor, Dr. Sci. (Biology), Russia

V.Yu. Makeev, Corr. Member of the RAS, Dr. Sci. (Physics and Mathem.),
Russia

I.V. Maksimov, Dr. Sci. (Biology), Russia

B.A. Malyarchuk, Dr. Sci. (Biology), Russia

Yu.G. Matushkin, Cand. Sci. (Biology), Russia

V.I. Molodin, Full Member of the RAS, Dr. Sci. (History), Russia

M.P. Moshkin, Professor, Dr. Sci. (Biology), Russia

L.Yu. Novikova, Dr. Sci. (Agricul.), Russia

E.K. Potokina, Dr. Sci. (Biology), Russia

V.P. Puzyrev, Full Member of the RAS, Dr. Sci. (Medicine),
RussiaD.V. Pyshnyi, Corr. Member of the RAS, Dr. Sci. (Chemistry),
Russia

E.Y. Rykova, Dr. Sci. (Biology), Russia

E.A. Salina, Corr. Member of the RAS, Professor,
Dr. Sci. (Biology), Russia

E.V. Shakhtshneider, Dr. Sci. (Medicine), Russia

S.V. Shekhovtsov, Dr. Sci. (Biology), Russia

S.V. Shestakov, Full Member of the RAS, Dr. Sci. (Biology),
RussiaV.A. Stepanov, Full Member of the RAS, Dr. Sci. (Biology),
RussiaI.A. Tikhonovich, Full Member of the RAS, Dr. Sci. (Biology),
Russia

A.V. Vasiliev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

M.I. Voevoda, Full Member of the RAS, Dr. Sci. (Medicine),
RussiaN.K. Yankovsky, Full Member of the RAS, Dr. Sci. (Biology),
Russia

N.S. Yudin, Cand. Sci. (Biology), Russia

I.K. Zakharov, Professor, Dr. Sci. (Biology), Russia

I.A. Zakharov-Gezekhus, Corr. Member of the RAS,
Dr. Sci. (Biology), Russia

Молекулярная и клеточная биология

- 541 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Концепция природной реконструкции генома. Часть 6. Анализ изменения частот встречаемости SNP в экзонах генов, ассоциированных с клональным гемопоэзом неопределенного потенциала у пациента с нейроэндокринной опухолью тонкой кишки в терминальной стадии прогрессии после модификации генома гемопоэтических стволовых клеток препаратом hDNA^{gr} (клинический случай).
Д.Ю. Ощепков, В.С. Рузанова, С.Г. Ошихмина, А.А. Траспов, Е.В. Долгова, Г.С. Риттер, С.С. Кирикович, Е.В. Левитес, Я.Р. Ефремов, О.С. Таранов, Л.Ю. Гривцова, С.В. Сидоров, С.Д. Никонов, О.Ю. Леплина, А.А. Останин, Е.Р. Черных, Н.А. Колчанов, А.С. Брюховецкий, С.С. Богачев
- 557 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Эффекты хронического психосоциального стресса на развитие ооцитов и преимплантационных эмбрионов мышей.
Т.Н. Иголина, И.Д. Чекашов, А.Л. Левинсон, С.Я. Амстиславский

Генетика растений

- 569 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Некоторые аспекты межвидовой гибридизации *Solanum lycopersicum* L. и *Solanum sisymbriifolium* Lam. А.В. Вишнякова, А.П. Федотов, Е.В. Осминина, А.З. Мартиросян, А.Д. Кобяшова, А.А. Александрова, А.А. Миронов, С.Г. Монахос
- 581 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Дифференциация субгеномов среди StY-геномных видов рода *Elymus* (Triticeae, Poaceae) с территории России по данным секвенирования ядерного гена *GBSS1* (*waxy*). А.В. Агафонов, Е.В. Шабанова (Кобозева), А.А. Бондарь, О.В. Дорогина
- 593 **ОБЗОР**
Исследования генетического разнообразия и геномов мискантуса для оптимизации их биотехнологического потенциала.
И.В. Чадаева, Д.И. Каретников, А.Ю. Пронозин, Э.С. Стародубцева, Н.А. Омелянчук, А.В. Кочетов, Д.А. Афонников
- 601 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Создание растений картофеля (*Solanum tuberosum* L.) с инактивированным геном *StPain-1* с помощью CRISPR/Cas9. В.Д. Карлов, М.К. Волков, А.Д. Антипов, Ю.С. Монахова, А.С. Трофимов, Л.Н. Коновалова, А.В. Бабаков, Р.А. Комахин, В.В. Таранов

Селекция растений на иммунитет и продуктивность

- 612 **ОБЗОР**
Грибные патогены растений: расшифровка механизмов фузариозного увядания.
М.П. Банкин, А.А. Самсонова, Т.А. Рожмина, М.Г. Самсонова
- 625 **ОБЗОР**
Использование геномных методов скрининга в создании селекционного материала риса с высоким качеством зерна.
Н.Г. Туманьян, Ж.М. Мухина
- 636 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Создание тритикале с фиолетовой окраской зерна. Н.В. Петраш, П.И. Стёпочкин
- 645 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Получение голозерных гибридов *Triticum durum* с *Triticum dicoccum* с антоциановой окраской перикарпа зерновки. Е.И. Гордеева, А.Ю. Новосельская-Драгович, О.Ю. Шоева, Е.К. Хлесткина

Генетика человека

- 658 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Генетическая характеристика и метаанализ населения Северного Причерноморья I тысячелетия н. э. по данным древней ДНК.
Э.Д. Айтуганова, А.С. Коньков, Т.В. Андреева, Е.В. Рождественских, А.Д. Манахов, Е.И. Рогаев

Медицинская генетика

- 669 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Отбор и анализ DNase I-гиперчувствительных сайтов для пренатального определения трисомии 21 у плода. А.М. Мазур, А.С. Старшин, Н.В. Богуш, Е.Б. Прохорчук
- 676 **ОБЗОР**
Парадигма соматического мозаицизма при многофакторных заболеваниях.
А.А. Слепцов, М.С. Назаренко, В.П. Пузырев
- 685 **ОБЗОР**
Мобильные генетические элементы в патогенезе репродуктивных нарушений человека. А.А. Бабовская, Е.А. Трифонова, В.А. Степанов

Эволюция видов

- 696 **ОБЗОР**
Эволюция вида *Ovis aries*.
Т.Н. Хамируев

Molecular and cell biology

- 541 **ORIGINAL ARTICLE**
The concept of natural genome reconstruction. Part 6. Analysis of changes in the frequency of occurrence of SNPs in exons of the genes associated with clonal hematopoiesis of uncertain potential in a patient with a neuro-endocrine tumor of the small intestine in the terminal stage of progression after modification of the hematopoietic stem cell genome with hDNAST (a clinical case report).
D.Y. Oshchepkov, V.S. Ruzanova, S.G. Oshikhmina, A.A. Traspov, E.V. Dolgova, G.S. Ritter, S.S. Kirikovich, E.V. Levites, Y.R. Efremov, O.S. Taranov, L.U. Grivtsova, S.V. Sidorov, S.D. Nikonov, O.Y. Leplina, A.A. Ostanin, E.R. Chernykh, N.A. Kolchanov, A.S. Bryukhovetskiy, S.S. Bogachev

- 557 **ORIGINAL ARTICLE**
Effects of chronic psychosocial stress on the development of mouse oocytes and preimplantation embryos.
T.N. Igonina, I.D. Chekashov, A.L. Levinson, S.Ya. Amstislavsky

Plant genetics

- 569 **ORIGINAL ARTICLE**
Issues of interspecific hybridization between *Solanum lycopersicum* L. and *Solanum sisymbriifolium* Lam. A.V. Vishnyakova, A.P. Fedotov, E.V. Osmolina, A.Z. Martirosyan, A.D. Kobyashova, A.A. Alexandrova, A.A. Mironov, S.G. Monakhos
- 581 **ORIGINAL ARTICLE**
Differentiation of subgenomes in *StY*-genomic species of the genus *Elymus* (Triticeae, Poaceae) from the territory of Russia according to sequencing data of the nuclear gene *GBSS1* (*waxy*). A.V. Agafonov, E.V. Shabanova (Kobozeva), A.A. Bondar, O.V. Dorogina
- 593 **REVIEW**
Research on genetic diversity and genomes of *Miscanthus* for optimizing their biotechnological potential. I.V. Chadaeva, D.I. Karetnikov, A.Yu. Pronozin, E.S. Starodubtseva, N.A. Omeljanchuk, A.V. Kochetov, D.A. Afonnikov
- 601 **ORIGINAL ARTICLE**
Development of potato (*Solanum tuberosum* L.) plants with inactivated *StPain-1* gene using CRISPR/Cas9. V.D. Karlov, M.K. Volkov, A.D. Antipov, Yu.S. Monahova, A.S. Trofimov, L.N. Konovalova, A.V. Babakov, R.A. Komakhin, V.V. Taranov

Plant breeding for immunity and performance

- 612 **REVIEW**
Fungal pathogens of plants: deciphering the mechanisms of Fusarium wilt. M.P. Bankin, A.A. Samsonova, T.A. Rozhmina, M.G. Samsonova
- 625 **REVIEW**
Using genomic screening methods to create high-quality rice breeding material. N.G. Tumanyan, Zh.M. Mukhina
- 636 **ORIGINAL ARTICLE**
Development of purple-grain triticale. N.V. Petrash, P.I. Stepochkin
- 645 **ORIGINAL ARTICLE**
Development of naked *Triticum durum* × *Triticum dicoccum* hybrids with anthocyanin color of the grain pericarp. E.I. Gordeeva, A.Yu. Novoselskaya-Dragovich, O.Yu. Shoeva, E.K. Khlestkina

Human genetics

- 658 **ORIGINAL ARTICLE**
Genetic characterization and meta-analysis of the population of the Northern Black Sea region in the 1st millennium CE based on ancient DNA data. E.D. Aituganova, A.S. Kon'kov, T.V. Andreeva, E.V. Rozhdestvenskikh, A.D. Manakhov, E.I. Rogaeva

Medical genetics

- 669 **ORIGINAL ARTICLE**
Selection and evaluation of DNase I hypersensitive sites for prenatal screening of trisomy 21 in the fetus. A.M. Mazur, A.S. Starshin, N.V. Bogush, E.B. Prokhortchouk
- 676 **REVIEW**
The paradigm of somatic mosaicism in complex diseases. A.A. Sleptcov, M.S. Nazarenko, V.P. Puzyrev
- 685 **REVIEW**
Mobile genetic elements in the pathogenesis of human reproductive disorders. A.A. Babovskaya, E.A. Trifonova, V.A. Stepanov

Evolution of species

- 696 **REVIEW**
Evolution of the species *Ovis aries*. T.N. Khamiruev

doi 10.18699/vjgb-26-57

The concept of natural genome reconstruction.

Part 6. Analysis of changes in the frequency of occurrence of SNPs in exons of the genes associated with clonal hematopoiesis of uncertain potential in a patient with a neuroendocrine tumor of the small intestine in the terminal stage of progression after modification of the hematopoietic stem cell genome with hDNA^{gr} (a clinical case report)

D.Y. Oshchepkov ^{1&}, V.S. Ruzanova ^{1&}, S.G. Oshikhmina ^{1, 2}, A.A. Traspov³, E.V. Dolgova ¹, G.S. Ritter ¹, S.S. Kirikovich ¹, E.V. Levites ¹, Y.R. Efremov ¹, O.S. Taranov ⁴, L.U. Grivtsova³, S.V. Sidorov ^{2, 5}, S.D. Nikonov⁶, O.Y. Leplina ⁷, A.A. Ostanin ⁷, E.R. Chernykh ⁷, N.A. Kolchanov ¹, A.S. Bryukhovetskiy^{8#}, S.S. Bogachev ^{1#} 

¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ National Medical Research Radiological Centre of the Ministry of Health of the Russian Federation, Obninsk, Russia

⁴ State Scientific Centre of Virology and Biotechnology "Vector" of Rospotrebnadzor, Koltsovo, Novosibirsk region, Russia

⁵ City Clinical Hospital No. 1, Novosibirsk, Russia

⁶ State Regional Novosibirsk Clinical Tuberculosis Hospital, Novosibirsk, Russia

⁷ Research Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia

⁸ Clinical Hospital "NeuroVita", Moscow, Russia

 labmolbiol@mail.ru

Abstract. The final part of our series of studies reports the results of a pilot study of a clinical case (patient K.), indicating that therapy enabling reconstruction of the genome of hematopoietic stem cells (HSCs) reduces the frequency of SNPs in exons of the genes associated with clonal hematopoiesis of uncertain potential in a patient with a neuroendocrine tumor of the small intestine in the terminal stage of progression. The fundamental principles of the therapy were as follows. Control samples of peripheral blood mononuclear cells (baseline) were collected from the patient. CD34+ HSCs were then mobilized and collected. The HSC genome was modified using a technology enabling *ex vivo* correction of the nucleotide sequences of DNA chromosomes of poorly differentiated hematopoietic progenitor cells. For this purpose, the collected leukocyte suspension enriched with CD34+ HSCs was treated with fragmented deproteinized genomic DNA obtained from approximately 100 young healthy women in labor (hDNA^{gr}). The patient was reinfused intravenously with the treated cells. Five days before reinfusion, the patient received immunosuppressive therapy. Samples of peripheral blood mononuclear cells were also collected 4, 8, 12, and 27 months after the therapeutic intervention. Full-exome sequencing of DNA isolated from selected cell samples was performed. A group of commonly accepted clonal hematopoiesis genes was selected as a criterion demonstrating changes in the genome. The following multiplex panels were used at the initial time point as well as time points after 4, 8, and 12 months: MGI Easy Exome Capture V5 Probe set, Roche KAPA HyperExome, Nanodigmbio NEXome Plus Panel v1.0, covering the entire exome and non-coding regions of the analyzed genes adjacent to exons with ~100× coverage. A ~1000× coated panel (Nanodigmbio NanOnco Plus Panel v3.0), including the genes with mutant alleles identified at the initial time point, was used to verify the identified mutations and more accurately determine their proportions in readings at the last stage of the analysis (27 months after the therapy). Whole-exome sequencing of hDNA^{gr} (Roche KAPA HyperExome), which is used to reconstruct the genome of HSCs whose fragments collectively constitute the complete human genome, was also performed. Although the analysis of readings of whole-exome sequencing of DNA isolated from blood samples after treatment did not reveal any currently known signs of clonal hematopoiesis with uncertain potential in the patient K., we found a significant, within the framework of the selected criterion, decrease in the frequency range of SNPs to 18 % by 27 months of follow-up in 15 % cases of the heterozygous alleles (*DNMT1*, *SF3B1*(1–2)). Subsequent analysis of the frequency of occurrence of controlled SNPs and the corresponding data on the depth of their sequencing, performed using the All-FIT algorithm, indicates that by the 27th month of follow-up, the proportion of cells carrying the basic set of SNPs significantly decreases by 12 % (confidence interval (CI) = 5–16 %). In other words, in a mixture of cells heterozygous for this SNP, a population of cells carrying a homozygote without an SNP appears. Our findings may indicate that a significant proportion of cells underwent correction of the SNP to an alternative allele, and this correction was associated with the treatment of the original HSCs with hDNA^{gr}. The detected changes in the group of clonal hematopoiesis genes suggest the possibility of correcting other loci throughout the HSCs genome. Therefore, the clinical manifestations in the development of the disease show a pronounced positive trend, which has persisted for four years, until present. It is assumed that the positive clinical outcome after therapy with reconstructed hematopoietic stem cells is associated with an increase in the regenerative potential of HSCs, which resulted from the genetic correction of unfavorable mutations in the HSCs genome.

Key words: clonal hematopoiesis; neuroendocrine tumor; genome reconstruction technology; whole-genome sequencing; hematopoietic stem cells

For citation: Oshchepkov D.Y., Ruzanova V.S., Oshikhmina S.G., Traspov A.A., Dolgova E.V., Ritter G.S., Kirikovitch S.S., Levites E.V., Efremov Y.R., Taranov O.S., Grivtsova L.U., Sidorov S.V., Nikonov S.D., Leplina O.Y., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bryukhovetskiy A.S., Bogachev S.S. The concept of natural genome reconstruction. Part 6. Analysis of changes in the frequency of occurrence of SNPs in exons of the genes associated with clonal hematopoiesis of uncertain potential in a patient with a neuroendocrine tumor of the small intestine in the terminal stage of progression after modification of the hematopoietic stem cell genome with hDNA^{gr} (a clinical case report). *Vavilovskii Zhurnal Genetiki i Selektii* = *Vavilov J Genet Breed.* 2026;30(4):541-556. doi 10.18699/vjgb-26-57

Funding. This work was conducted at the Institute of Cytology and Genetics and supported by the Ministry of Science and Higher Education of the Russian Federation (state budget-funded project No. FWNR-2026-0025), I.N. Zaitseva and Clinical Hospital "NeuroVita" JSC.

Концепция природной реконструкции генома.

Часть 6. Анализ изменения частот встречаемости SNP в экзонах генов, ассоциированных с клональным гемопоэзом неопределенного потенциала у пациента с нейроэндокринной опухолью тонкой кишки в терминальной стадии прогрессии после модификации генома гемопоэтических стволовых клеток препаратом hDNA^{gr} (клинический случай)

Д.Ю. Ощепков ^{1&}, В.С. Рузанова ^{1&}, С.Г. Ошихмина ^{1, 2}, А.А. Траспов³, Е.В. Долгова ¹, Г.С. Риттер ¹, С.С. Кирикович ¹, Е.В. Левитес ¹, Я.Р. Ефремов ¹, О.С. Таранов ⁴, А.Ю. Гривцова³, С.В. Сидоров ^{2, 5}, С.Д. Никонов⁶, О.Ю. Леплина ⁷, А.А. Останин ⁷, Е.Р. Черных ⁷, Н.А. Колчанов ¹, А.С. Брюховецкий^{8#}, С.С. Богачев ^{1#} 

¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

² Новосибирский национальный государственный исследовательский университет, Новосибирск, Россия

³ Национальный медицинский исследовательский центр радиологии Министерства здравоохранения Российской Федерации, Обнинск, Россия

⁴ Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора, р. п. Кольцово, Новосибирская область, Россия

⁵ Городская клиническая больница № 1, Новосибирск, Россия

⁶ Государственная областная Новосибирская клиническая туберкулезная больница, Новосибирск, Россия

⁷ Научно-исследовательский институт фундаментальной и клинической иммунологии, Новосибирск, Россия

⁸ АО Клинический госпиталь «НейроВита», Москва, Россия

 labmolbiol@mail.ru

Аннотация. В последней части цикла работ приводятся результаты пилотного исследования клинического случая (пациент К.), свидетельствующие о снижении частоты встречаемости SNP в экзонах генов, ассоциированных с клональным гемопоэзом неопределенного потенциала у пациента с нейроэндокринной опухолью тонкой кишки в терминальной стадии прогрессии при применении терапии, позволяющей реконструировать геном гемопоэтических стволовых клеток (ГСК). Основные принципы терапии состояли в следующем. У пациента отбирали контрольные образцы мононуклеаров периферической крови (нулевая точка). Затем проводились мобилизация и сбор CD34+ ГСК. Осуществляли модификацию генома ГСК с использованием технологии, позволяющей провести *ex vivo* коррекцию нуклеотидных последовательностей ДНК хромосом низкодифференцированных предшественников гемопоэза. Для этого собранную лейкозвесь, обогащенную CD34+ ГСК, обрабатывали фрагментированной депротеинизированной геномной ДНК, полученной от ~100 молодых здоровых рожениц (hDNA^{gr}). Пациенту внутривенно реинфузировали обработанные клетки. За 5 дней до реинфузии пациент получал иммуносупрессивную терапию. Через 4, 8, 12 и 27 мес. также забирались образцы мононуклеаров периферической крови. Было проведено полноэкзомное секвенирование ДНК, выделенной из отобранных образцов клеток. В качестве критерия, отражающего изменения в геноме, была выбрана группа общепринятых генов клонального гемопоэза. В нулевой точке, в точках через 4, 8, 12 мес. использовались мультиплексные панели: MGI Easy Exome Capture V5 Probe set, Roche KAPA HyperExome, Nanodigmbio NEXome Plus Panel v1.0, охватывающие весь экзом и некодирующие области анализируемых генов, прилегающие к экзонам с покрытием ~100x. Для верификации выявленных мутаций и более точного определения их долей в прочтениях на последнем этапе анализа, через 27 мес. после терапии, использовалась панель с покрытием ~1000x (Nanodigmbio NanOnco Plus Panel v3.0), включающая выявленные в нулевой точке гены с мутантными аллелями. Также было выполнено полноэкзомное секвенирование препарата hDNA^{gr} (Roche KAPA HyperExome), используемого для реконструкции генома ГСК, фрагменты которого в совокупности составляют полный геном человека. Несмотря на то что анализ прочтений полноэкзомного секвенирования ДНК, выделенной из мононуклеаров крови после проведенного лечения, не позволил выявить у пациента К. известных на данный момент признаков клонального гемопоэза с неопределенным потенциалом, нами обнаружено достоверное в рамках выбранного критерия снижение диапазона частот SNP до 18 % к 27-му месяцу наблюдения в 15 % гетерозиготных аллелей (*DNMT1*, *SF3B1*(1-2)). Последующий анализ частоты встречаемости контролируемых SNP и соответствующих данных по

глубине их секвенирования, выполненный с использованием алгоритма All-FIT, свидетельствует, что к 27-му месяцу наблюдения доля клеток, несущих базовый набор SNP, значимо снижается на 12 % (доверительный интервал 5–16 %). Иными словами, в смеси клеток, гетерозиготных по этим SNP, появляется популяция клеток, несущая гомозиготу без SNP. Полученные результаты могут свидетельствовать о том, что для значимой доли клеток произошла коррекция SNP на альтернативный аллель и эта коррекция связана с обработкой исходных ГСК препаратом hDNA^{gr}, причем обнаруженные изменения в группе генов клонального гемопоэза предполагают возможность коррекции других локусов по всему геному ГСК. Как следствие, клинические проявления в развитии патологии болезни имеют ярко выраженную положительную динамику, сохраняющуюся на протяжении четырех лет, до настоящего времени. Предполагается, что положительный клинический эффект после проведенной терапии реконструированными стволовыми клетками крови связан с усилением регенеративного потенциала ГСК, произошедшего вследствие генетической коррекции неблагоприятных мутаций в геноме ГСК.

Ключевые слова: клональный гемопоэз; нейроэндокринная опухоль; технология реконструирования генома; полногеномное секвенирование; гемопоэтические стволовые клетки

Introduction

Regenerative medicine is currently an actively developing medical field. Its objective is to restore the damaged or disease-affected tissues and organs using stem cells. Transplantation of mature stem cells is the key method employed in regenerative medicine. Two types of stem cells are used most widely: mesenchymal stem cells (MSCs) and hematopoietic stem cells (HSCs).

Mesenchymal stem cells repair damage via various mechanisms. They are activated by pro-inflammatory cytokines such as interferon gamma (IFN- γ), tumor necrosis factor alpha (TNF α), interleukin-1 beta (IL-1 β), as well as by different ligands of toll-like receptors 3 (TLR3) and 4 (TLR4) (Waterman et al., 2010; English, 2013). Activated MSCs secrete a variety of cytokines and chemokines, such as fibroblast growth factor 2 (FGF2), insulin-like growth factor 1 (IGF1), transforming growth factor beta (TGF β), prostaglandin E2 (PGE2), granulocyte-macrophage colony-stimulating factor (GM-CSF), interleukin 6 (IL-6) and interleukin 13 (IL-13), and indoleamine 2,3-dioxygenase (IDO). Through the factors secreted by them, MSCs affect many cells of the innate and adaptive immunity, including macrophages, dendritic cells, natural killer cells, T-helper and regulatory T-cells, as well as B-cells, shifting the local balance toward the anti-inflammatory signaling to promote tissue regeneration (Yin, Heit, 2021).

Due to their paracrine activity, MSCs inhibit oxidative stress and exhibit anti-apoptotic properties (Brown et al., 2019). *In vitro* studies have demonstrated that MSCs can migrate toward specific cytokines and chemokines, such as stromal cell-derived factor 1 (SDF-1), platelet-derived growth factor AB (PDGF-AB), IGF1, TNF α , and epidermal growth factor (EGF), which are released into the extracellular space in response to tissue injury or ischemia, and elicit anti-inflammatory responses. MSCs also release extracellular vesicles containing various growth factors promoting angiogenesis and cell proliferation, as well as microRNAs, tRNAs, and lipids, which inhibit apoptosis, facilitate cell repair, maintain the multipotency of stem cells, and contribute to angiogenesis (Maacha et al., 2020).

Another regeneration mechanism involves direct mitochondrial transfer from MSCs to damaged tissue cells via tunneling

nanotubes, thus enhancing the regenerative potential of target cells and stimulating the antimicrobial activity of immune cells (Islam et al., 2012; Li X. et al., 2014; Jackson et al., 2016). MSCs are capable of differentiating into various types of connective tissue cells such as osteoblasts, chondrocytes, adipocytes, and myocytes. MSCs are found in the bone marrow, umbilical cord blood, placenta, adipose tissue, etc.; they can repair damaged cells and tissues and accelerate their regeneration (Spees et al., 2016). MSCs transplantation has shown promising potential in preclinical and clinical studies for treating numerous degenerative pathologies, including Alzheimer's disease (Wang et al., 2018), Parkinson's disease (Mendes Filho et al., 2018), and amyotrophic lateral sclerosis (Gugliandolo et al., 2019). Along with transdifferentiation, MSCs were shown to be able to fuse with recipient cells, including brain, muscle, intestinal, and hepatic cells, thus contributing to tissue repair in these organs (Fan et al., 2020).

The contribution of hematopoietic stem cells (HSCs) to the body's repair potential is no less significant than that of mesenchymal stem cells. HSCs are involved in repair of tissue damage, and primarily the damage related to activity of blood cells. HSCs are capable of differentiating into all types of blood and immune cells, thus ensuring their continuous turnover in the body. There is evidence that HSCs can transdifferentiate into somatic cells, including hepatocytes, cardiomyocytes, as well as neural, epithelial, and endothelial cells, thereby contributing to regeneration of respective tissues (Lee, Hong, 2020; Li L. et al., 2021; Yuan et al., 2023). Nevertheless, there is also a viewpoint that the reparative potential of HSCs in tissue systems other than lymphohematopoietic ones is more likely to be associated with paracrine activity of HSCs rather than true transdifferentiation. It is believed that the developmental potential of HSCs, like for other adult stem cells, is limited by the innate type of stem cells and tissue (Müller et al., 2016). Similar to MSCs, HSCs actively affect cells at injury sites by secreting various molecules, including cytokines, chemokines, and growth factors, thus modulating pro- and anti-inflammatory responses and activating repair mechanisms in target cells (Schwartz et al., 2008; Li N. et al., 2010; Wright et al., 2011; Liao et al., 2018; Fast et al., 2021; Biermann, Reya, 2022).

It is known that the repair potential of hematopoietic stem cells is primarily related to their allelic diversity, underpinning normal hematopoiesis. This characteristic of HSCs is lost throughout life, and hematopoiesis in the aging organism acquires clonal features.

The classical model of hematopoiesis, positing that differentiation of HSCs into mature blood cells is a strictly hierarchical process, is evolving. The current understanding of hematopoiesis involves a concept of hematopoiesis as functioning of a heterogeneous pool of HSCs and committed progenitor cells, with numerous differentiation pathways and gradual transitions between states (Jagannathan-Bogdan, Zon, 2013; Velten et al., 2017; Watcham et al., 2019). Hematopoietic homeostasis is ensured by an intricate network of interactions between the mechanisms of transcriptional regulation, epigenetic modification, and metabolic adaptation, which are influenced by external factors, as well as humoral and local signals from the bone marrow microenvironment (Pinho, Frenette, 2019). Aging is accompanied by quantitative and qualitative changes in the hematopoietic system. The proliferative activity of hematopoietic stem cells decreases; immune activity is shifted toward pro-inflammatory responses (Weiskopf et al., 2009; Ainciburu et al., 2023). It is believed that the development of clonal hematopoiesis is one of the factors contributing to these phenotypic changes.

Normal hematopoiesis is maintained by numerous long- and short-lived hematopoietic cell clones sequentially replacing each other. The bone marrow is a tissue proliferating most rapidly in the human body. The human body produces approximately one trillion mature blood cells daily (Doulatov et al., 2012). Hematopoietic stem cells capable of differentiating into all mature blood cell lineages underlie the hematopoietic system (Phillips, 1991; Bonnet, 2003). Clone is the progeny derived from a single hematopoietic stem cell (Siminovitch et al., 1963). The system is polyclonal upon steady-state hematopoiesis, meaning that numerous hematopoietic stem cells are simultaneously involved in maintaining normal parameters of peripheral blood cells (Drize et al., 1996; Goyal, Zandstra, 2015; Carrelha et al., 2018). Fewer than 1.3 million hematopoietic stem cells support the formation of mature peripheral blood cells throughout human life (Watson et al., 2020; Petinati, Drize, 2021). Clonal hematopoiesis is the disproportionate expansion of the progeny of several HSCs harboring specific mutations with respect to the remaining pool of clones. Clonal hematopoiesis is generally considered a biological state of the hematopoietic system rather than a disease per se.

The concept of clonal hematopoiesis has emerged after typical somatic mutations in peripheral blood cells of large patient cohorts had been detected (Genovese et al., 2014; Jaiswal et al., 2014). Mutations in a similar set of genes have been identified across all the studied groups; their prevalence was age-dependent. Over 70 different genes carrying specific mutations have been identified (Sleptsov et al., 2023). These include epigenetic regulators (*DNMT3A*, *TET2*, and *ASXL1*), DNA damage repair (DDR) genes (*TP53*, *PPM1D*, *CHEK2*, and *ATM*), signaling genes of cell growth (*JAK2*), and splicing factors (*SF3B1* and *SRSF2*). *DNMT3A*, *TET2*, and *ASXL1* were found to be the most commonly mutated genes (80–90 % of

all the mutations detected) (Genovese et al., 2014; Jaiswal et al., 2014; Heyde et al., 2021; Joo et al., 2023). It is currently known that somatic mutations that have arisen and been fixed in HSCs genes, primarily those involved in regulation of epigenetic modifications and cell cycle control, are the key drivers of clonal expansion.

Somatic mutations in the genes of clonal hematopoiesis result from stochastic events; however, the accelerated pace of their accumulation is largely associated with age-related changes in HSCs, which are characterized by a decline in the overall regenerative potential (Geiger et al., 2013). Somatic mutations were found to be present in 10 % of individuals over 65 years of age vs. only 1 % in those aged under 50. The pace of mutation accumulation depends directly on the specific gene. *DNMT3A*-mutant clones appear at an early age and are characterized by a steady rate of emergence of new clones (approximately 5 % per year). Clones harboring mutations in splicing factor genes (*SF3B1* and *SRSF2*) emerge at a later age; the rate of emergence of new ones is ~50 % per year. *TET2*-mutant clones are continuously formed and detected throughout life (Fabre et al., 2022).

The formation of mutant HSCs occurs as a result of the accumulation of mutations, and the cell loses its regenerative potential. As a result of this process, the mutated clone acquires a selective proliferative advantage, leading to clonal expansion. Clonal hematopoiesis develops, and the hematopoietic system loses the ability to repair damage in the body. Elimination of mutant HSCs clones that have accumulated somatic mutations making the cell unviable is an additional mechanism promoting clonal expansion. This defensive mechanism reduces the pool of stem and progenitor cells, thus increasing the pace of mutation accumulation through higher proliferative burden on the remaining HSCs (Sleptsov et al., 2023).

Age-related telomere shortening, caused by the semi-conservative mechanism of chromosomal DNA replication, increases the likelihood of DNA damage, thus also contributing to mutation accumulation (D'Adda Di Fagagna et al., 2003).

Expansion of mutant clones is not necessarily accompanied by abnormal blood parameters or clinical signs. In other words, the number of cells harboring somatic mutations in peripheral blood increases even before malignant transformation is observed. This state is known as clonal hematopoiesis of uncertain potential (CHUP) (Heuser et al., 2016; Shlush, 2018; Steensma, 2018). The key criterion for CHUP is determining the variant allele frequency (VAF) (i. e., the mutational burden, which is proportional to the number of abnormal cells). In the context of clonal hematopoiesis of uncertain potential, this value is $\geq 2\%$ (Duncavage, Tandon, 2015).

There are external factors causing somatic mutagenesis: age, genetic predisposition, pharmacotherapy, and environmental exposure risks (Joo et al., 2023). Nevertheless, the proposed variants fail to fully elucidate the molecular and cellular mechanisms of the emergence of somatic mutations in HSCs, which underlie the development of clonality in hematopoietic progenitor cells.

Based on the previously identified mechanism of extracellular DNA internalization by stem cells of different lineage, it is

fair to hypothesize the following pathway for the accumulation of unfavorable SNPs in HSCs.

In peripheral tissues, mutations continuously emerge and are accumulated within cells (Shamal, 2017; Cagan et al., 2022). After apoptotic or necrotic death, cells harboring mutations release fragments containing genetic abnormalities into the bloodstream (Anker et al., 1999; Jahr et al., 2001; Laktionov et al., 2004). These fragments reach HSCs (apparently, as well as other types of stem cells) and are internalized by them via a natural mechanism due to the characteristic structure of the glycocalyx of HSCs, which is positively charged (Dolgova et al., 2014; Petrova et al., 2022; Ritter et al., 2022; Potter et al., 2024). The extracellular fragments delivered into HSCs (and potentially other stem cells) initiate terminal differentiation of HSCs, which is characterized by the emergence of pangenomic single-stranded DNA breaks (Vatolin et al., 1997; Potter et al., 2024; Ruzanova et al., 2024). This event triggers activation of the recombinational repair system of the cell and induces the so-called recombinogenic situation (Likhacheva et al., 2008). It is hypothesized that during the development of the recombinogenic situation, extracellular DNA fragments can be engaged in recombinational interactions with genomic DNA and be integrated into the genome at homologous loci by replacing native genomic sequences (Yakubov et al., 2007; García-Olmo et al., 2012). Since recombination under conditions of pangenomic single-stranded breaks is highly precise (Xu, 2015; Vriend, Krawczyk, 2017; Maizels, Davis, 2018; Zilio, Ulrich, 2021), this insertion is functional and will not disrupt the genomic sequence, except for inserting the mutation. The mutations accumulated over time can result in the development of clonal features in HSCs, with all subsequent implications. This mechanism becomes more likely in the case of activation of apoptosis or the appearance of foci of necrotic destruction in any of the body's tissues due to pathologies that have arisen.

Therefore, analyzing the genes associated with clonal hematopoiesis has naturally become the primary focus of research aiming to elucidate the consequences of targeted therapy. If the proposed concept is valid, it will be possible to detect respective alterations in biomarker genes associated with clonal hematopoiesis in a direct experiment using treatment of HSCs with a preparation based on fragmented DNA (hDNA^{gr}) as well as the modern whole-genome sequencing techniques. In the cases when clonal hematopoiesis of uncertain potential (CHUP) develops, the commonly accepted genes from this group, carrying driver mutations specific to this condition, can be an adequate model for detecting allele substitutions in the nucleotide sequences of these genes. If such changes are detected, the analyzed group of genes can become a marker platform indicative of *ex vivo* correction, which can then be extrapolated to the entire genome. However, if none of the known marker events (in the known genes) of clonal hematopoiesis are revealed, the detected changes in SNP frequencies can actually attest to alterations in the clonal composition of HSCs. In turn, this would suggest potential correction of unfavorable alleles in other genes and the emergence of clones carrying a different set of SNP alleles responsible for the regenerative potential of HSCs. In the clinical context, replacing the mutant alleles (related either to the clonal hematopoiesis genes or to other

diverse genes involved in maintaining the regenerative properties of hematopoietic stem cells) with alternative favorable alleles will indicate that the original regenerative potential of HSCs has been restored.

This part of the study series reports the results of a pilot study of a clinical case, demonstrating that therapy with blood mononuclear cells enriched with CD34+ hematopoietic stem cells, which had been treated with a double-stranded human DNA preparation (hDNA^{gr}), reduced the occurrence frequency of a certain set of SNPs marking the common genes associated with CHUP.

Under the selected criteria, we detected changes in the occurrence frequencies of certain heterozygous germline mutant alleles of the genes associated with clonal hematopoiesis: *ASXL1*, *DNMT1*, *RAD21*, and *SF3B1*(1–3). The nucleotide substitution frequency for them was reduced to 18 % of the normal distribution expected for pure heterozygous variants (50 %). It attests to the significant deviation from the expected frequency distribution for a single clonal cell lineage with the set of SNPs characteristic of the heterozygous state and the emergence of a mixture of clonal lineages harboring the investigated SNPs in different proportions.

Based on the data on allele frequencies of SNPs and the respective sequencing depth obtained using the All-FIT platform (Loh et al., 2020), we assessed the proportion of cells belonging to different clonal lineages harboring a specific SNP set and revealed statistically significant evidence for the emergence of 12 % (confidence interval (CI) ranging from 5 to 16 %) of cells carrying an alternative SNP set. Therefore, it is hypothesized that the conducted therapy led to replacement of mutant alleles with non-mutant ones. The regenerative potential of HSCs, and possibly other types of stem cells such as MSCs, was restored, which was related not only to correction in the clonal hematopoiesis genes but also involved the restoration of other mutant loci within a significant proportion of HSCs. The genetic abnormalities in these loci had caused weakening of regenerative potential. Prominent positive clinical improvement in the patient's condition was observed.

Materials and methods

The legal framework for conducting clinical trials. In study conduct, the authors adhered to the principles of the Declaration of Helsinki developed by the World Medical Association. The clinical part of the study was regulated by relevant treatment protocols; informed consent was obtained from the patient (Supplementary Material 1)¹.

The hDNA^{gr} preparation. The cells were treated using fragmented human genomic DNA (hDNA^{gr} preparation, technical specifications TU 20.42.15-001-58179831-2023, Patent No. 2855434 dated February 2, 2026) obtained from ~100 healthy women in labor. The preparation had undergone real-time PCR verification and several assays characterizing its epidemiological purity in compliance with the regulatory documents.

Patient. Medical history. Cancer patient K., 67 years old, presenting with severe cardiovascular disorders. Medical record No. 2022/0025. The patient stayed in hospital (Clinical

¹ Supplementary Materials 1–4 are available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Oschep_Engl_30_4.zip

Hospital “NeuroVita” JSC.) from July 13, 2022 until July 29, 2022 (a total of 16 inpatient days).

Diagnosis of the underlying medical condition: C17. A neuroendocrine tumor of the small intestine (T3N0M0). The patient had undergone small bowel resection with ileostomy. The histological report dated October 29, 2018 confirmed a G2 neuroendocrine tumor with involvement of the small intestine wall, with signs of lymphovascular and perineural invasion, but without evidence of metastases. The bowel resection margins were free of tumor infiltration. Comorbidities: ischemic heart disease, post-infarction cardiosclerosis, obliterating atherosclerosis of coronary arteries.

October 29, 2018: 1 course of polychemotherapy according to the XELOX scheme. Pharmacotherapy with Octreotide Depot (until May, 2019).

October 26, 2021: disease progression, growth of foci in the liver (possibly, metastases). Pharmacotherapy with Somatuline.

November 11, 2021: chemoembolization of the segmental branches of the right hepatic artery with irinotecan. November 19, 2021: Transluminal Balloon Coronary Angioplasty (TBCA) with stenting of the right coronary artery (RCA), left anterior descending artery (LAD), and the aorta (AO).

Hypertension grade 3, risk level 4. Cardiac rhythm and conduction abnormalities: paroxysmal atrial fibrillation and paroxysmal supraventricular tachycardia. January 31, 2022: radiofrequency ablation of the slow pathway of the AV junction and the cavotricuspid isthmus. First-degree atrioventricular block. Supraventricular extrasystoles. Chronic heart failure (CHF), stage 2a, NYHA functional class 2, in the subcompensation phase.

February 8, 2022: therapy with Somatuline 120 mg every 28 days. Stabilization.

Disease complications: disseminated intravascular coagulation (DIC) (May 15, 2022). Systemic inflammatory response syndrome (May 15, 2022), without pathogen identified. Liver abscess (February 16, 2022). Right lower lobe pneumonia (May 16, 2022), resolving. Oral cavity candidiasis (February 20, 2022). The patient was declared incurable by a council of oncologists; hospice care for cancer patients was recommended.

Since May 11, 2022, the patient was experiencing a persistent fever ($>38.0^{\circ}\text{C}$). At his own discretion, the patient received Amoxiclav and NSAIDs, without any effect. He was experiencing febrile episodes ($>38.0^{\circ}\text{C}$) two or three times daily, which persisted until he had been hospitalized. He was admitted to the “NeuroVita” Clinic on July 13, 2022, having signs of fever of unknown origin. Complaints upon admission: right upper quadrant abdominal pain, general weakness, malaise, chest pain episodes, and arthralgia. The objective of hospitalization was to elucidate the etiology of the fever of unknown origin and to perform corrective therapy.

The patient had an advanced malignant neoplasm exhibiting signs of resistance to treatment. He met the inclusion criteria for the “Protocol of Personalized Adoptive Gene Therapy for Treating Malignant Tumors” and was enrolled in the treatment protocol based on the histologically confirmed diagnosis of malignancy; evidence for disease progression despite stan-

dard therapy or resistance to therapy; and age between 18 and 85 years.

The key phases of personalized adoptive gene therapy.

All the therapeutic procedures were carried out in compliance with the protocol described in refs. (Bryukhovetskiy, Bogachev, 2023; Bryukhovetskiy, Shurdov, 2024) and the Clinical Treatment Protocol (Supplementary Material 1). At patient’s discretion and with his voluntary consent, hematopoietic blood and bone marrow cells were harvested under the program in compliance with the aforementioned protocol. The initial step involved mobilization and collection of CD34⁺ HSCs. Starting June 22, 2022, the patient received eight subcutaneous injections of granulocyte-colony stimulating factor (G-CSF) at 10–12-h intervals during four days following the routine hematopoietic stimulation protocol aiming to mobilize HSCs into the peripheral blood. The treatment regimen was as follows: G-CSF (figrastim) injected on days 1 and 2 at a dose of $4.4\text{ }\mu\text{g/kg}$ per day; on days 3 and 4, $8.8\text{ }\mu\text{g/kg}$ per day. On the 5th day, leukapheresis was performed, and hematopoietic stem cells from blood mononuclear cells were collected, followed by trephine biopsy and bone marrow sampling. The cellularity of the harvested material was sufficient for further manipulations. The proportion of CD34⁺ HSCs in the fraction of mononuclear cells (MNCs) was $\sim 2\%$. The autologous MNCs concentrate was then purified to remove red blood cells, platelets, and granulocytes using ficoll density gradient centrifugation (BioloT Ltd., Russia). The purified MNCs concentrate (leukocyte concentrate) was cryopreserved according to the standard cryopreservation procedure. Dimethyl sulfoxide (Beijing Solarbio Science & Technology Co., Ltd, China) at a final concentration of 6–10 % was added to the cell suspension. The mixture was gradually frozen to -80°C at a rate of 1°C/min and stored in liquid nitrogen vapor.

The clinical part of the therapy. Starting July 13, 2022: immunosuppression with methylprednisolone (N2) was initiated; starting July 14, 2022: sirolimus (N4). On July 19, 2022, *ex vivo* gene therapy was performed via reinfusion of autologous HSCs restored using the medicinal substance hDNA^{gr}. The sample of cryopreserved MNCs enriched with CD34⁺ HSCs was thawed. Immediately prior to reinfusion, the genome of the HSCs was modified using a technology enabling *ex vivo* correction of nucleotide sequences in chromosomal DNA of poorly differentiated hematopoietic progenitor cells. For this purpose, the collected leukocyte suspension enriched with CD34⁺ HSCs was washed to remove the cryoprotectant, resuspended in normal saline, and coincubated (60 min, 37°C) with fragmented deproteinized genomic DNA (hDNA^{gr}) obtained from ~ 100 young healthy women in labor. The leukocyte suspension was then washed to remove the hDNA^{gr}. With the informed consent provided by patient K., an autologous cell-based gene therapy preparation containing 3.6×10^{10} autologous peripheral blood mononuclear cells in 100 mL of Sol. Polyglukini was administered intravenously via a central venous catheter under aseptic conditions at the intensive care unit. The transfusion was conducted after intramuscular premedication with Sol. Suprastini (2.0 mL). The hemodynamic parameters were monitored during the reinfusion (BP 170/70 mmHg, HR 62 bpm, SpO₂ 97 %). Sol. Dexametasoni 4.0 was infused

intravenously after the procedure to prevent hypersensitivity reactions. Hemodynamic monitoring was performed one hour post-infusion (BP 115/60 mmHg, HR 67 bpm, SpO₂ 97–96 %). The patient underwent the procedure satisfactorily. No complications occurred during the procedure. Copies of the documents accompanying the “Protocol for Personalized Adoptive Immunotherapy in the Treatment of Advanced Cancer and Other Malignant Neoplasms” and the Patient Informed Consent Form are provided in Supplementary Material 1.

Patient. Follow-up of October 28, 2022. The patient feels satisfactorily and reports having hiked in the mountains and swum in a mountain lake. Contrast-enhanced CT revealed no findings indicative of cancer progression. Retroperitoneal lymph nodes were enlarged by 1–2 mm. The patient’s overall condition was interpreted as cancer stabilization.

Patient. Follow-up of April 10, 2023. On April 10, 2023, the patient was admitted to hospital for inpatient examination. His condition was satisfactory. Contrast-enhanced CT/PET imaging of the thoracic, abdominal, and pelvic organs revealed no evidence of disease progression. No abnormal tumor markers were identified. Cardiac condition had improved. Although having heart failure, ischemic heart disease, and stents in coronary arteries, the patient had experienced no chest pain over the past six months (whereas prior to therapy, he was seeking emergency care 3–4 times per week). Gout has resolved (the patient was having painful gout attacks over the past five years); hemorrhoids have become compensated. Patient’s hair color has recovered; hair has become thick and silky. The patient’s facial complexion has normalized. He is having a normal everyday life. The patient would feel absolutely healthy if he had not undergone colostomy.

Whole-genome sequencing. Cell material was collected to analyze alterations in hematopoietic stem cells after the therapy (reinfusion of mononuclear cells enriched with autologous HSCs restored using the hDNA^{gr} preparation). Control bone marrow and peripheral blood samples were collected prior to HSCs mobilization and MNCs harvesting. Bone marrow and peripheral blood samples were also collected 4, 8, 12, and 27 months post-intervention. Whole-exome sequencing (WES) of DNA extracted from the collected samples of MNCs was performed. Furthermore, whole-exome sequencing was conducted for the hDNA^{gr} preparation used to reconstruct the HSCs genome, whose fragments collectively represent the complete human genome sequence.

Genomic DNA was extracted from all the collected biological specimens; alterations in the genes associated with clonal hematopoiesis were analyzed by next-generation sequencing (NGS).

Targeted paired-end sequencing was conducted using an MGISEQ-G400 sequencing platform (MGI, China) utilizing multiplex panels MGIEasy Exome Capture V5 Probe Set, Roche KAPA HyperExome, Nanodigmbio NEXome Plus Panel v1.0, or Nanodigmbio NanOnco Plus Panel v3.0 (as specified in the “Test kit” column in the tables provided in Supplementary Material 2). The sequencing aimed to identify germline mutations and clonal somatic mutations in the genes associated with clonal hematopoiesis (*AKT1*, *ASXL1*, *ASXL2*, *BCOR*, *BCORL1*, *BRCC3**, *CBL*, *CTCF*, *CUX1**, *DNMT1*,

DNMT3A, *DNMT3b*, *ETNK1**, *GNAS*, *GNB1**, *IDH1*, *IDH2*, *JAK2*, *KIT*, *KRAS*, *MPL*, *MTOR*, *MYD88*, *NLRP3**, *NRAS*, *PPM1D*, *PTEN*, *PTPN11*, *RAD21*, *RUNX1*, *SETBP1**, *SF3B1*, *SOX2*, *SRSF2*, *TET2*, *TP53*, *U2AF1*, *UBA1**). The genes marked with an asterisk are not included in the Nanodigmbio NanOnco Plus Panel v3.0; in the table, the “ND” (no data) index is given for them. The targeted coverage depth is $\times 100$ for exome panels (MGIEasy Exome Capture V5 Probe Set, Roche KAPA HyperExome, and Nanodigmbio NEXome Plus Panel v1.0) (Schema for Exome Probesets – Exome Capture Probesets and Targeted Region, n. d.), and $\sim \times 1000$ or higher for the Nanodigmbio NanOnco Plus Panel v3.0 (NanOnco Plus Panel v3.0, n. d.).

Two bioinformatics pipelines were employed for analyzing the sequencing data: the first one was optimized for germline mutation detection; the second one, for somatic mutation detection.

The following software tools were used for the germline pipeline: BWA-MEM2 v2.2.1 for aligning to the human reference genome (GRCh37); StreamMD v4.3.0 for marking and removing PCR duplicates; Sambamba v1.0.1 for filtering mapped variants; GATK Picard for quality control and mapping metrics assessment; Google DeepVariant v1.6.1 for nucleotide variant identification; BCFtools v1.20 for variant filtering; and Ensembl Variant Effect Predictor (VEP) release 112 for variant annotation.

The software used for the somatic pipeline was as follows: BWA-MEM2 v2.2.1 for aligning to the human reference genome (GRCh37); StreamMD v4.3.0 for marking and removing PCR duplicates; Sambamba v1.0.1 for filtering mapped variants; GATK v4.4.0.0 for quality control and mapping metrics assessment; GATK Mutect 2 for nucleotide variant identification; FINGS v1.7.2 for variant filtering; and Ensembl Variant Effect Predictor (VEP) release 112 for variant annotation.

The coding and non-coding genetic variants are listed in individual tables in Supplementary Material 2. The coding genetic variants include those residing within the protein-coding region of the gene, including synonymous variants, as well as the variants of donor and acceptor splice sites. The variants residing outside the protein-coding regions of the gene, excluding variants of donor and acceptor splice sites, were classified as non-coding genetic variants.

The risk of pathogenicity was assessed in compliance with the ACMG (SF v2.0)/Medical and Genetic Research Center guidelines, including the following criteria: low population frequency (variant frequency in 1,000 genomes and/or ExAC C NM_015338.6 < 0.01); a damaging effect (nonsense, frameshift); information about the variant is contained in locus-specific databases, etc.

The data on variant allele frequency (VAF) are reported in Supplementary Material 2 and constitute the official report on the analysis performed at the P.A. Herzen Moscow Cancer Research Institute, a branch of the National Medical Research Radiological Centre of the Ministry of Health of the Russian Federation. The data are provided in tabulated values. The data on variant allele frequencies in exons and adjacent regions listed in the summarized table are visualized using a green gradient scale.

Analysis of the frequency of occurrence of SNPs using the All-FIT software (Loh et al., 2020). The analysis was carried out using default program settings. The input data were the SNPs identified at the previous stage, along with sequencing depth data (Supplementary Material 2, Table 1). The total number of SNPs was 15. Data without removal of germline SNPs, LOH (Loss of Heterozygosity), or subclonal SNPs were included in the analysis, since the clonality set of HSCs was not supposed to contain malignant tissue variants. Only the coding SNPs for which the VAF data were available at both the initial (pre-treatment) and final (27 months post-treatment) points were used for the computation using the All-FIT software. Therefore, the sets of SNPs selected for computation did not differ, and any differences in the results were attributable exclusively to variations in frequencies, but not the set of SNPs.

Statistical analysis was conducted using the Statistica 8 software (StatSoft, Tulsa, USA). Statistical significance was assessed using the Mann–Whitney U test; the differences were considered statistically significant at $P_v < 0.05$.

Results

The summarized whole-genome sequencing data. Choosing the genome sequencing platform to analyze changes in the frequency of occurrence of SNPs

Patient K. with a neuroendocrine tumor of the small intestine in the terminal stage of progression (stage IV, metastatic spread to the liver, lymph nodes, and the left kidney) underwent therapy with autologous MNCs enriched with HSCs treated with the hDNA^{gr} preparation. Patient K. presenting with signs of fever of unknown origin was admitted to the “NeuroVita” clinic. The patient was declared incurable by a council of oncologists; hospice care for cancer patients was recommended. The objective of hospitalization was refining the etiology of fever of unknown origin and performing corrective therapy. The patient had an advanced malignant neoplasm exhibiting signs of resistance to treatment. During the corrective therapy, he was found to meet the inclusion criteria for the “Protocol of Personalized Adoptive Gene Therapy for Treating Advanced Cancer and other Malignant Tumors” and was enrolled in the treatment program.

After the therapy, patient K. experienced a rapid and steady improvement of his condition. Remission of the underlying disease has been observed for four years to date. It was assumed that the favorable changes in clinical manifestations of the underlying disease and comorbidities in patient K. were related to the therapeutic effect on hematopoietic stem cells present within MNCs. We have put forward a hypothesis that this impact had resulted in genetic correction of the mutant loci in hematopoietic progenitor cells (Ruzanova et al., 2024, 2025), restoring the regenerative repair potential of HSCs, which was the reason behind the prominent clinical effect.

It is known that MNCs therapy, whether administered as monotherapy or in combination with physical (e. g., ultraviolet irradiation) (Andreu et al., 1994; Greinix et al., 2000) or humoral interventions (e. g., interleukin-2) (Dummer et al., 1993; Mendelenko et al., 1997; Pal'tsev et al., 2000), or through

using MNCs conditioned medium (Beer et al., 2016), elicits an anti-inflammatory response at multiple levels. This effect is attributed to factors released during MNCs apoptosis and to activation of anti-inflammatory properties in macrophages residing in various tissues (Beer et al., 2016). Furthermore, MNCs transfusion stimulates angiogenesis, which is currently extensively studied in the context of tissue tropism impairment in patients with type 1 diabetes mellitus, cardiovascular diseases (including ischemic myocardial infarction), and stroke. This effect is believed to be related to activation of endothelial progenitor cells present within the leukocyte concentrate (Zhang, Huang, 2012; Vahidy et al., 2016; Gurusamy et al., 2018; Yunir et al., 2021). However, we have failed to find any additional cases of long-lasting and diverse clinical effects of MNCs therapy aside the one observed for patient K. in the academic literature available through PubMed. As previously noted, we attribute the restoration of the regenerative potential of HSCs to correction of somatic mutations fixed in the HSCs genome. The recovered ability of HSCs to effectively repair various types of injury in different tissues and organs is additive and synergistic with respect to the regenerative potential reported for the mononuclear cell fraction, where this tissue regeneration effect may also be related to activated (but not reconstructed) HSCs, explaining the limited scope of the observable regenerative effects of MNCs therapy per se.

The fundamental elements of adoptive cell therapy (gene therapy) involve the following manipulations. As mentioned previously, the initial step includes mobilization of hematopoietic progenitor cells using granulocyte colony-stimulating factor. The therapeutic procedure entails treating MNCs with the hDNA^{gr} preparation, exhaustive elimination of extraneous extracellular DNA from the cells, followed by reinfusion of the treated mononuclear cells into the patient's bloodstream. The reinfused mononuclear cell fraction largely remains in its original state, except for the fact that HSCs present in the leukocyte concentrate have internalized extracellular dsDNA fragments, and processes within the cells leading to genetic reconstruction have been initiated.

The DNA fragments in the hDNA^{gr} preparation collectively constitute a complete human genome, encompassing a broad allelic representation of healthy variants across all the genomic sequences. Following treatment, these dsDNA fragments of the hDNA^{gr} preparation are naturally internalized by HSCs, thereby inducing the development of a recombinogenic situation (Likhacheva et al., 2008; Ruzanova et al., 2024). According to the concept of natural genome reconstruction, an exchange of homologous sequences between extrachromosomal fragments and the genomic DNA takes place via the mechanism known as single-strand assimilation (ss assimilation) (Leung et al., 1997; Langston, Symington, 2005; Symington, 2014; Yakubov et al., 2024). This exchange results in substitution of mutant alleles in the genome with alleles derived from the fragments. If the alleles within these fragments are normal, healthy variants, corrective restoration of the mutant loci will take place. These considerations suggest that the prominent clinical effect of the therapy was primarily attributable to the genetic correction of HSCs rather than to activation of undefined factors within the integral population of mononuclear blood cells.

In this study, we have put forward a hypothesis that as a result of stochastic genome reconstruction, hematopoietic stem cells have restored the regenerative potential lost throughout their life and because of the developed pathology, eventually expanding this potential to the entire body.

We conducted whole-exome sequencing (WES) of DNA extracted from the patient's MNCs to test the hypothesis that treatment with HSCs within MNCs will be accompanied by SNP correction and the emergence of HSCs clones with non-mutant alleles, ultimately restoring the regenerative potential of HSCs and contributing to the observed clinical effects. Only the results of sequencing blood-derived MNCs samples were analyzed. Control samples were collected prior to the therapeutic intervention. MNCs samples were collected again 4, 8, 12, and 27 months post-treatment. DNA was extracted from these samples, and WES was conducted. Multiplex panels (MGIEasy Exome Capture V5 Probe Set, Roche KAPA HyperExome, and Nanodigmbio NEXome Plus Panel v1.0), targeting the entire exome with $\sim 100\times$ coverage (Schema for Exome Probesets – Exome Capture Probesets and Targeted Region, n. d.), were used at the initial time point and 4, 8, and 12 months post-treatment. A targeted panel with $\sim 1,000\times$ coverage (Nanodigmbio NanOnco Plus Panel v3.0) (NanOnco Plus Panel v3.0, n. d.), which included the genes harboring mutant alleles detected at the initial point, was employed to verify the identified mutations and more accurately quantify their proportions at the final stage of follow-up 27 months post-therapy. Non-coding regions adjacent to the exons were additionally sequenced. Furthermore, whole-exome sequencing of hDNA^{gr} preparation, Roche KAPA HyperExome, which is used for reconstructing the HSCs genome whose fragments constitute the complete human genome, was also performed (Supplementary Material 2, Fig. 1). The utilization of multiple multiplex panels was a technical solution; their specificity for the selected coverage sufficiently aligned with the criteria required for SNP analysis (target, off-target, and duplications), making this approach adequate for assessing SNP frequencies (Schema for Exome Probesets – Exome Capture Probesets and Targeted Region, n. d.; Belova et al., 2025).

A set of genes associated with clonal hematopoiesis of uncertain potential (CHUP) was selected for analyzing genomic alterations after the treatment. This genetic platform was chosen because changes affecting the specified genes constitute the molecular foundations for the mechanism of clonal expansion of mutated HSCs clone. The resulting dominance of such clones is a key factor in the development of clonal hematopoiesis of uncertain potential compromising the body's defense function.

It was hypothesized that dominant clones had been formed in the hematopoietic system of patient K., which could have been a confounding factor for the underlying pathology.

It was assumed that the assessment of SNP frequency reduction within the selected gene group could had been one of the criteria attesting to the potential changes in the HSCs genome after treatment. The advantages of this approach are related to the ease of detecting changes in the exome by next-generation sequencing (NGS), the relative cost-effectiveness of the technique, and the availability of bioinformatics tools

enabling targeted analysis of this specific gene set. If genetic alterations are found in the analyzed genes, it would suggest that there are high chances that similar events have occurred across the entire genome. Importantly, restoration of the regenerative potential of HSCs is more likely to be related to the recovery of other mutated loci (in which the mutations had initially weakened the regenerative potential) rather than solely to correction of mutations within the genes associated with clonal hematopoiesis. The detection of these genetic changes will be a marker of enhanced regenerative potential of the hematopoietic system and simultaneously, a marker of *ex vivo* correction per se. Furthermore, it is hypothesized that similar processes can occur for other types of stem cells (MSCs and endothelial progenitor cells) that are present in the leukocyte concentrate and internalize extracellular fragments (Dolgova et al., 2014), which will also be a factor significantly enhancing the overall regenerative potential of the body.

Whole-exome sequencing revealed no known driver SNPs included in the established list of genes associated with clonal hematopoiesis and responsible for its development (Genovese et al., 2014; Jaiswal et al., 2014; Heyde et al., 2021; Joo et al., 2023). It means that the patient had no systemic changes in hematopoiesis that would be detectable by genomic analysis. According to the available literature, no dominant clones of HSCs with disrupted mechanisms responsible for their regenerative potential had been formed. In other words, the clonal hematopoiesis in its classical form was not detected in this patient. Simultaneously, this fact suggests that the prominent positive changes in the patient's condition observed after treatment are associated with genetic alterations in other mutant alleles of the HSCs genes directly involved in the aforementioned regenerative mechanisms.

Nevertheless, the analysis identified SNPs associated with the common clonal hematopoiesis genes, but these genes do not correspond to driver alleles and appear to have a germline origin (100 % occurrence frequency for a homozygote; ~ 50 % occurrence frequency for a heterozygote) (Supplementary Material 2, Table 1).

Primary analysis demonstrated that the frequency of some of the detected germline SNPs at the final follow-up point was significantly lower compared that at the initial point. This could mean that during the interaction between extracellular DNA fragments and hematopoietic stem cells, some SNPs affecting the regenerative potential of HSCs in the genome were replaced with an alternative allele; the hDNA^{gr} preparation was the only possible source. In other words, a somatic correction occurred. Due to correction of mutant SNPs, a new clonal HSCs lineage could have emerged, differing from the pre-treatment one.

An important implication of this observation is that alleles of the HSCs genes harboring germline SNPs were affected. The presence of germline SNPs suggests that all the cells in the body have only this genotype, and there are no other hematopoietic stem cell populations harboring alternative alleles. This fact efficiently eliminates the questions related to the potential activation of HSCs that had existed in the quiescent state since the embryonal development, which could be triggered by the treatment. In other words, if new SNPs have

appeared or existing ones have disappeared, these changes could only have occurred during the postgerminal period and resulted from somatic processes.

We needed to find approaches that would demonstrate the validity of these changes; assessment of the variant allele frequencies of the identified SNPs could be the main one.

Analysis of the frequency of occurrence of SNPs

Two approaches were selected to assess the frequencies of SNP occurrence: (1) comparative analysis of changes in SNP frequencies in selected genes at the initial and final follow-up points (27 months), and (2) the All-FIT algorithm (Loh et al., 2020) (Supplementary Material 2, Table 2).

Analysis of changes in the frequencies of SNP occurrence in selected genes at the initial and final follow-up points (27 months). A specific approach was used in this part of the study to prove that replacement of SNPs with alternative alleles had occurred. This approach consisted in the analysis of changes in allele frequencies within the exons of heterozygous genes. The results were initially normalized to the existing literature data, which incorporated findings obtained in different laboratories, at different time points, and employing different NGS platforms. The average error in estimating the distribution of allele frequencies within exons and adjacent non-coding regions of heterozygous genes was determined at the second stage, based solely on the findings of the present study (the “internal” experimental error). These analyses employed four different NGS platforms (Supplementary Material 2, Table 3).

Six genes (*ASXL1*, *RAD21*, *DNMT1*, *SF3B1-1*, *SF3B1-2*, and *SF3B1-3*) were selected for the comparative analysis; the frequency of SNP occurrence in these genes was ~40–60 %, suggesting that these alleles were heterozygous. The occurrence frequencies of homozygous SNPs were close to 100 % at all the analysis points. The comparison revealed a statistically significant reduction in the frequency of SNPs in three of the selected genes (*ASXL1*, *DNMT1*, and *SF3B1-2*) at the final follow-up point according to the selected criterion (Fig. 1). The logic and the main steps of the analysis are outlined in

Supplementary Material 3. It means that the observed reduction in SNP frequency is probably associated with genetic correction of SNP that had resulted from homologous exchange of mutant genomic sequence with a non-mutant allele present in extracellular DNA fragments within the hDNA^{ex} preparation.

The All-FIT (allele-frequency-based imputation of tumor purity) algorithm. The second approach employed the All-FIT bioinformatics platform. According to the algorithm proposed by the authors, this method allows one to determine the proportion (the “purity” according to ref. (Loh et al., 2020)) of cancer cells and accompanying somatic tissue (cells) in the resected tumor sample with a respective confidence interval (CI) using the weighted least squares method through iterative steps based on variant allele frequencies (VAF) of SNPs detected in target clinical high-throughput sequencing data and the sequencing depth for each selected allele. The simulated and clinical data were used to demonstrate the accuracy and improved performance of the All-FIT approach compared to other leading computational methods. The program was validated for specific clinical samples and showed high predictive effectiveness (Loh et al., 2020).

Following the core concept of the platform, the proposed algorithm can be used to determine the ratio between the proportion of cells harboring different SNP sets by analyzing changes in variant allele frequencies of a limited number of SNPs available for analysis as well as the sequencing depth, which shows the reliability of frequency estimates. The reduced frequency of the analyzed alleles compared to the marker (anchor) frequencies ascertained by the algorithm would attest to a decline in the number of cells carrying the generalized set of selected SNPs, being fully consistent with the objectives of our study.

The study analyzed the frequencies of SNPs in the genes associated with clonal hematopoiesis based on the initial hypothesis regarding the presence of CHUP. The findings demonstrated that the patient had no clonal hematopoiesis. However, the limited number of SNPs available for VAF analysis (together with sequencing depth data), related only to the genes of this group, allows one to conduct

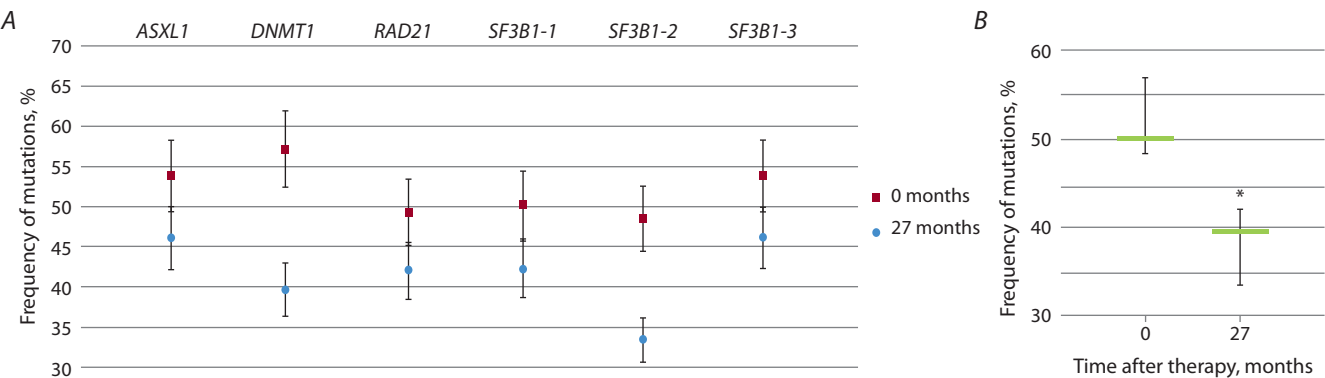


Fig. 1. The frequency of SNPs of the analyzed genes at the final follow-up point 27 months post-intervention. A, The diagrams illustrating the overlapping of the total percentage error values for the SNP allele frequencies of three heterozygous genes at the initial and final follow-up points. B, Changes in the frequencies of SNPs in the selected heterozygous genes at the initial and final follow-up points; frequencies at the initial point fluctuate around 50 % (three SNP alleles). The median allele frequencies are compared with allowance for the total percentage error of the values obtained using different multiplex panels. * Statistically significant differences (Pv < 0.05) as determined using the Mann–Whitney U test.

All-FIT analysis, which will reliably reveal any significant change in the occurrence frequency of controlled SNPs if it is detected.

HSCs within peripheral blood mononuclear cells were treated with the hDNA^{gr} preparation containing a pool of ~100 different DNA samples encompassing all the sequences in the human genome (100×). According to the concept of natural genome reconstruction and the results of its experimental validation, DNA fragments internalized by HSCs undergo stochastic and homologous recombination with the accessible homology regions of genomic DNA. An assumption was made that a detected reduction in frequency of identified SNPs will imply that some SNPs, including those outside the analyzed SNP set, had been replaced with alternative alleles via the aforementioned homologous recombination. This genetic correction has altered the ratio between VAFs of the analyzed SNPs, which can potentially be detected by computations using the All-FIT algorithm. Extracellular dsDNA fragments internalized by HSCs within the MNCs are the best-substantiated source of this correction. Figure 2 shows the results of the comparison.

The analysis revealed that the purity (the proportion of cells carrying the initial SNP set) at the initial point was 0.99 (CI, 92–99 %), which actually means that there is a single “clone”. After 27-month follow-up, the purity decreased to 0.88 (CI, 84–95 %). It means that ~12 % of cells having a different set of SNP alleles (CI, 5–16 %) have emerged. A homozygous SNP-free clonal cell lineage appears in the pool of heterozygous cells with analyzed SNPs. The most plausible explanation to these values is that some SNP alleles have been replaced with alternative alleles present in the hDNA^{gr} preparation, which have become fixed in some HSCs lineages and are characterized by an enhanced regenerative potential.

The data obtained using the two independent approaches reinforce the conclusion that the proposed mechanism of interaction between extracellular dsDNA fragments and chromosomal DNA does actually exist. In other words, the hypothesized concept of natural genome reconstruction is viable.

The hypothetical scenario of events taking place during reinfusion of reconstructed HSCs to patient K. is provided in Supplementary Material 4.

Discussion

Overall, our findings are consistent with the proposed concept of genome reconstruction via a natural mechanism (Ruzanova et al., 2024, 2025; Yakubov et al., 2024). The key questions of the entire study address the feasibility of non-homologous integration into the genome and the accuracy of homologous integration. The clinical applicability of the concept depends on the answer to this question.

Undoubtedly, certain events take place within the nucleus, leading to potential recombinational repair interactions between the extrachromosomal fragments of extracellular dsDNA and chromosomal DNA. It is supported by the data reported below.

We have previously demonstrated that telomeric DNA content increases upon internalization of dsDNA fragments by HSCs, which can be associated with both true integration

of telomeric repeats into the genome and with alternative lengthening of telomeres (Ruzanova et al., 2025). Furthermore, circular DNA is formed, consisting of extrachromosomal extracellular fragments looping into circles and giving rise to multiple associates with chromosomal DNA strands, which persist until the metaphase and can be detected by FISH or by direct analysis of the distribution of fluorescent-labeled DNA probes (Ruzanova et al., 2025).

The results of this study suggest that mutant alleles of the genes of hematopoietic stem cells (HSCs) are apparently restored, which is interpreted as projection of the core idea of the concept: stochastic substitution of various chromosomal DNA segments with homologous sequences derived from extrachromosomal fragments internalized by HSCs. An important factor in this context is that according to the findings of whole-genome sequencing conducted in (Oshikhmina et al., 2026), there is no evidence for non-homologous probe integration, indicating that homologous allele replacement is more likely to occur.

Vital tests demonstrate that syngeneic reinfusion of mouse and rat HSCs treated with heterologous human DNA (hDNA^{gr}) causes neither immediate nor long-term (one year) pathological sequelae. They also suggest that there is no illegitimate integration of foreign extrachromosomal DNA into active genomic loci; therefore, there are no genetic abnormalities manifesting themselves as an immediate functional response (Ruzanova et al., 2025).

The most important result of our study is the reported clinical case involving patient K. Long-lasting stabilization of cancer progression was achieved, accompanied by positive clinical manifestations. Along with stabilization of disease progression, there were explicit favorable effects on the cardiovascular system. The phenomenon of new hair growth demonstrates that stem cells migrated to hair follicles and were fixed in them. However, the key aspect of this clinical case is that genetic alterations in the mutant heterozygous germline alleles have been detected. ***In other words, our study revealed the phenomenon of somatic correction in hematopoietic stem cells, which occurred naturally through internalization of extracellular dsDNA fragments by hematopoietic progenitor cells.***

The primary critical question concerning the entire technology, including the reported clinical case, is whether extracellular DNA fragments that had originally been internalized by HSCs can be preserved in their progeny for a long time (up to 4, 9, 12, and 27 months) and whether this DNA (SNPs) can be detected in leukocyte concentrate cells by whole-exome sequencing after time passes. We believe that this outcome (namely, long-term preservation of extrachromosomal extracellular DNA in the progeny of hematopoietic cells), which may give rise to the corresponding artifact, is highly unlikely. Our findings demonstrate that activated HSCs treated with dsDNA preparation continuously proliferate, at least during 15-day follow-up (Ruzanova et al., 2024). After eight months, blood cells potentially containing extracellular fragments inherited from the maternal HSCs, which is capable of internalizing up to 1 % of the haploid genome (Dolgova et al., 2013, 2016; Potter et al., 2024), will either be eliminated based on their

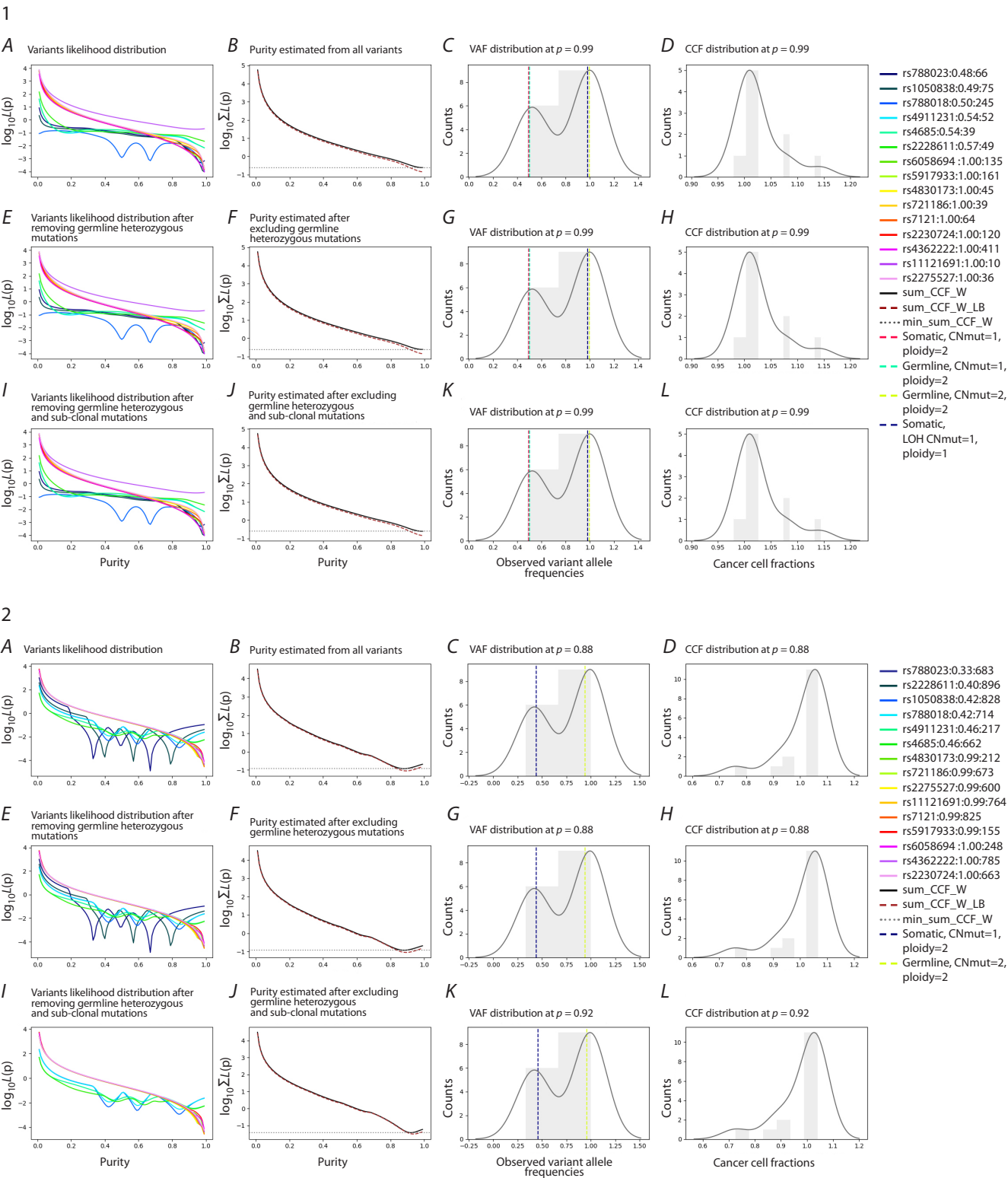


Fig. 2. Presentation of the results of All-FIT analysis for the detected clonal hematopoiesis gene variants (Loh et al., 2020). An analysis focused on the allele frequencies of SNPs within the exons of the clonal hematopoiesis genes evaluated at two time points: at the initial (point 1) and the final follow-up points (2) (27 months post-therapy).

A, The contribution of each variant to the sum of probabilities. **B,** The sum of probabilities across all the variants and the 2σ curve (the brown dashed line). The points of intersection between the 2σ curve and the tangent to the log of the sum of probabilities at its minimum (the gray dashed line) define the confidence interval (CI) around the calculated purity parameter p . **C,** The distribution of observed allele frequencies. The dashed lines are the expected allele frequencies in the included mutation models for the hypothesized p . **D,** The distribution of the proportion of clones (cancer cell fraction, CCF) for the hypothesized p . **E–H,** The results after unfiltered heterozygous germline variants had been excluded from analysis. **I–L,** The results after excluding the subclonal somatic variants. The legends to the figures show the mutated genes and the VAF of each detected variant.

lifespan or will constitute an extremely small proportion of the blood cell population, so that their detection will be infeasible from a quantitative perspective. Furthermore, our unpublished data indicates that extracellular DNA material is no longer detectable within cells ~22 days post-treatment. Therefore, the results of whole-exome sequencing of mononuclear cell samples derived from patient K. 4, 8, 12, and 27 months after the therapy reflect the real situation with genomic alterations and the emergence of genetically reconstructed renewed hematopoietic stem cell clones.

An interesting observation was that homozygous germline alleles of the genes have not been corrected, suggesting they are under selective pressure. Heterozygous alleles not being under selective pressure have undergone correction; the regenerative potential of the therapy has statistically significantly increased. The emergence of extracellular DNA fragments within the intracellular space of HSCs is accompanied by homologous correction and simultaneously initiates cell division (Potter et al., 2024; Ruzanova et al., 2024). The progenitor cells divide both symmetrically and asymmetrically (Potter et al., 2024; Ruzanova et al., 2024). Their progeny, resulting from both symmetric and asymmetric division, circulate in peripheral blood. It is fair to hypothesize that the genes responsible for tissue regeneration have undergone correction, thus restoring the original regenerative potential of the HSCs, and that these cells, which have acquired a new status, have caused the extensive systemic regeneration in patient K.

Conclusions

Overall, the technology of natural genome reconstruction using the therapeutic hDNA^{gr} (Patent No. 2855434 dated February 2, 2026) can be employed for treating patients with diseases of civilization and elderly patients without evident pathologies. Nonetheless, serious experimental studies of all subtleties of the molecular and cellular processes triggered upon interaction between a stem cell (and HSC in particular) and extracellular DNA fragments need to be conducted.

References

- Ainciburu M., Ezponda T., Berastegui N., Alfonso-Pierola A., Vilas-Zornoza A., Martin-Uriz P.S., Alignani D., ... Diez-Campelo M., Valcarcel D., Hernaez M., Romero J.P., Prosper F. Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single-cell resolution. *eLife*. 2023;12:e79363. doi 10.7554/eLife.79363
- Andreu G., Leon A., Heshmati F., Tod M., Menkes C.J., Baudelot J., Laroche L. Extracorporeal photochemotherapy: evaluation of two techniques and use in connective tissue disorders. *Transfus Sci*. 1994;15(4):443-454. doi 10.1016/0955-3886(94)90178-3
- Anker P., Mulcahy H., Chen X.Q., Stroun M. Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients. *Cancer Metastasis Rev*. 1999;18(1):65-73. doi 10.1023/A:1006260319913
- Beer L., Mildner M., Gyöngyösi M., Ankersmit H.J. Peripheral blood mononuclear cell secretome for tissue repair. *Apoptosis*. 2016;21(12):1336-1353. doi 10.1007/S10495-016-1292-8
- Belova V., Vasiliadis I., Repinskaia Z., Samitova A., Shmitko A., Ponikarovskaya N., Suchalko O., Cheranov V., Shatalov P., Shegai P., Kaprin A., Rebrikov D., Korostin D. Comparative evaluation of four exome enrichment solutions in 2024: Agilent, Roche, Vazyme and NanodigmBio. *BMC Genomics*. 2025;26(1):76. doi 10.1186/S12864-024-11196-Z
- Biermann M., Reya T. Hematopoietic stem cells and regeneration. *Cold Spring Harb Perspect Biol*. 2022;14(8):a040774. doi 10.1101/CSHPERSPECT.A040774
- Bonnet D. Biology of human bone marrow stem cells. *Clin Exp Med*. 2003;3(3):140-149. doi 10.1007/S10238-003-0017-9
- Brown C., McKee C., Bakshi S., Walker K., Hakman E., Halassy S., Svinarich D., Dodds R., Govind C.K., Chaudhry G.R. Mesenchymal stem cells: cell therapy and regeneration potential. *J Tissue Eng Regen Med*. 2019;13(9):1738-1755. doi 10.1002/TERM.2914
- Bryukhovetskiy A.S., Bogachev S.S. At the front edge of the fight against clonal hematopoiesis in diseases of civilization: from bone marrow transplantation to personalized gene-oriented and proteome-based restitution of bone marrow and hemopoietic stem cells. *National Association of Scientists*. 2023;97(2):26-45. doi 10.31618/nas.2413-5291.2023.2.97.845 (in Russian)
- Bryukhovetskiy A.S., Shurdov M.A. Aging and Anti-aging: Biomedical Approaches to Increasing Life Expectancy and Active Longevity. Ridero, 2024 (in Russian)
- Cagan A., Baez-Ortega A., Brzozowska N., Abascal F., Coorens T.H.H., Sanders M.A., Lawson A.R.J., ... Gerstung M., Campbell P.J., Murchison E.P., Stratton M.R., Martincorena I. Somatic mutation rates scale with lifespan across mammals. *Nature*. 2022;604(7906):517-524. doi 10.1038/s41586-022-04618-z
- Carrelha J., Meng Y., Kettyle L.M., Luis T.C., Norfo R., Alcolea V., Boukarabila H., ... Lord A.M., Sanjuan-Pla A., Woll P.S., Nerlov C., Jacobsen S.E.W. Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature*. 2018;554(7690):106-111. doi 10.1038/NATURE25455
- D'Adda Di Fagagna F., Reaper P.M., Clay-Farrace L., Fiegler H., Carr P., Von Zglinicki T., Saretzki G., Carter N.P., Jackson S.P. A DNA damage checkpoint response in telomere-initiated senescence. *Nature*. 2003;426(6963):194-198. doi 10.1038/NATURE02118
- Dolgova E.V., Efremov Y.R., Orishchenko K.E., Andrushkevich O.M., Alyamkina E.A., Proskurina A.S., Bayborodin S.I., ... Omigov V.V., Minkevich A.M., Rogachev V.A., Bogachev S.S., Shurdov M.A. Delivery and processing of exogenous double-stranded DNA in mouse CD34+ hematopoietic progenitor cells and their cell cycle changes upon combined treatment with cyclophosphamide and double-stranded DNA. *Gene*. 2013;528(2):74-83. doi 10.1016/j.gene.2013.06.058
- Dolgova E.V., Alyamkina E.A., Efremov Y.R., Nikolin V.P., Popova N.A., Tyrinova T.V., Kozel A.V., ... Mayorov V.I., Shurdov M.A., Ostanin A.A., Chernykh E.R., Bogachev S.S. Identification of cancer stem cells and a strategy for their elimination. *Cancer Biol Ther*. 2014;15(10):1378-1394. doi 10.4161/cbt.29854
- Dolgova E.V., Potter E.A., Proskurina A.S., Minkevich A.M., Chernykh E.R., Ostanin A.A., Efremov Y.R., Bayborodin S.I., Nikolin V.P., Popova N.A., Kolchanov N.A., Bogachev S.S. Properties of internalization factors contributing to the uptake of extracellular DNA into tumor-initiating stem cells of mouse Krebs-2 cell line. *Stem Cell Res Ther*. 2016;7(1):76. doi 10.1186/s13287-016-0338-8
- Doulavov S., Notta F., Laurenti E., Dick J.E. Hematopoiesis: a human perspective. *Cell Stem Cell*. 2012;10(2):120-136. doi 10.1016/j.stem.2012.01.006
- Drize N., Keller J., Chertko J. Local clonal analysis of the hematopoietic system shows that multiple small short-living clones maintain life-long hematopoiesis in reconstituted mice. *Blood*. 1996;88(8):2927-2938. doi 10.1182/blood.V88.8.2927.bloodjournal8882927
- Dummer R., Becker J.C., Eilles C., Schafer E., Borner W., Burg G. T cells migrate to tumour sites after extracorporeal interleukin 2 stimulation and reinfusion in a patient with metastatic melanoma. *Br J Dermatol*. 1993;128(4):399-403. doi 10.1111/J.1365-2133.1993.TB00198.X
- Duncavage E.J., Tandon B. The utility of next-generation sequencing in diagnosis and monitoring of acute myeloid leukemia and myelodysplastic syndromes. *Int J Lab Hematol*. 2015;37(S1):115-121. doi 10.1111/IJLH.12361

- English K. Mechanisms of mesenchymal stromal cell immunomodulation. *Immunol Cell Biol.* 2013;91(1):19-26. doi 10.1038/ICB.2012.56
- Fabre M.A., de Almeida J.G., Fiorillo E., Mitchell E., Damaskou A., Rak J., Orrù V., ... Campbell P.J., McKinney E.F., Cucca F., Gerstung M., Vassiliou G.S. The longitudinal dynamics and natural history of clonal haematopoiesis. *Nature.* 2022;606(7913):335-342. doi 10.1038/S41586-022-04785-Z
- Fan X.L., Zhang Y., Li X., Fu Q.L. Mechanisms underlying the protective effects of mesenchymal stem cell-based therapy. *Cell Mol Life Sci.* 2020;77(14):2771-2794. doi 10.1007/S00018-020-03454-6
- Fast E.M., Spörrij A., Manning M., Rocha E.L., Yang S., Zhou Y., Guo J., Baryawno N., Barkas N., Scadden D., Camargo F., Zon L.I. External signals regulate continuous transcriptional states in hematopoietic stem cells. *eLife.* 2021;10:e66512. doi 10.7554/eLife.66512
- García-Olmo D.C., Picazo M.G., García-Olmo D. Transformation of non-tumor host cells during tumor progression: theories and evidence. *Expert Opin Biol Ther.* 2012;12(Suppl.1):S199-207. doi 10.1517/14712598.2012.681370
- Geiger H., De Haan G., Florian M.C. The ageing haematopoietic stem cell compartment. *Nat Rev Immunol.* 2013;13(5):376-389. doi 10.1038/NRI3433
- Genovese G., Kähler A.K., Handsaker R.E., Lindberg J., Rose S.A., Bakhoum S.F., Chambert K., ... Sullivan P.F., Sklar P., Grönberg H., Hultman C.M., McCarroll S.A. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med.* 2014;371(26):2477-2487. doi 10.1056/NEJMoa1409405
- Goyal S., Zandstra P.W. Stem cells: chasing blood. *Nature.* 2015;518(7540):488-490. doi 10.1038/nature14203
- Greinix H., Volc-Platzer B., Kalhs P., Fischer G., Rosenmayr A., Keil F., Hönigsmann H., Knobler R. Extracorporeal photochemotherapy in the treatment of severe steroid-refractory acute graft-versus-host disease: a pilot study. *Blood.* 2000;96(7):2426-2431. doi 10.1182/blood.V96.7.2426
- Gugliandolo A., Bramanti P., Mazzon E. Mesenchymal stem cells: a potential therapeutic approach for amyotrophic lateral sclerosis? *Stem Cells Int.* 2019;2019:3675627. doi 10.1155/2019/3675627
- Gurusamy N., Alsayeri A., Rajasingh S., Rajasingh J. Adult stem cells for regenerative therapy. In: Teplow D.B. (Ed.) *Progress in Molecular Biology and Translational Science.* Vol. 160. Academic Press, 2018;160:1-22. doi 10.1016/bs.pmbts.2018.07.009
- Heuser M., Thol F., Ganser A. Clonal hematopoiesis of indeterminate potential. *Dtsch Arztebl Int.* 2016;113(18):317-322. doi 10.3238/arztebl.2016.0317
- Heyde A., Rohde D., McAlpine C.S., Zhang S., Hoyer F.F., Gerold J.M., Cheek D., ... Craig M., Swirski F.K., Nahrendorf M., Nowak M.A., Naxerova K. Increased stem cell proliferation in atherosclerosis accelerates clonal hematopoiesis. *Cell.* 2021;184(5):1348-1361.e22. doi 10.1016/j.cell.2021.01.049
- Islam M.N., Das S.R., Emin M.T., Wei M., Sun L., Westphalen K., Rowlands D.J., Quadri S.K., Bhattacharya S., Bhattacharya J. Mitochondrial transfer from bone-marrow-derived stromal cells to pulmonary alveoli protects against acute lung injury. *Nat Med.* 2012;18(5):759-765. doi 10.1038/NM.2736
- Jackson M.V., Morrison T.J., Doherty D.F., McAuley D.F., Matthay M.A., Kissenpfennig A., O'Kane C.M., Krasnodembskaya A.D. Mitochondrial transfer via tunneling nanotubes is an important mechanism by which mesenchymal stem cells enhance macrophage phagocytosis in the in vitro and in vivo models of ARDS. *Stem Cells.* 2016;34(8):2210-2223. doi 10.1002/stem.2372
- Jagannathan-Bogdan M., Zon L.I. Hematopoiesis. *Development.* 2013;140(12):2463-2467. doi 10.1242/dev.083147
- Jahr S., Hentze H., Englisch S., Hardt D., Fackelmayer F.O., Hesck R.D., Knippers R. DNA fragments in the blood plasma of cancer patients: quantifications and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* 2001;61(4):1659-1665
- Jaiswal S., Fontanillas P., Flannick J., Manning A., Grauman P.V., Mar B.G., Lindsley R.C., ... Atzmon G., Wilson J.G., Neuberg D., Altshuler D., Ebert B.L. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med.* 2014;371(26):2488-2498. doi 10.1056/NEJMoa1408617
- Joo L.J., Bradley C.C., Lin S.H., Scheet P.A., Nead K.T. Causes of clonal hematopoiesis: a review. *Curr Oncol Rep.* 2023;25(3):211-220. doi 10.1007/S11912-023-01362-Z
- Laktionov P.P., Tamkovich S.N., Rykova E.Y., Bryzgunova O.E., Starikov A.V., Kuznetsova N.P., Sumarokov S.V., Kolomiets S.A., Sevostianova N.V., Vlassov V.V. Extracellular circulating nucleic acids in human plasma in health and disease. *Nucleosides Nucleotides Nucleic Acids.* 2004;23(6-7):879-883. doi 10.1081/NCN-200026035
- Langston L.D., Symington L.S. Opposing roles for DNA structure-specific proteins Rad1, Msh2, Msh3, and Sgs1 in yeast gene targeting. *EMBO J.* 2005;24(12):2214-2223. doi 10.1038/sj.emboj.7600698
- Lee J.Y., Hong S.H. Hematopoietic stem cells and their roles in tissue regeneration. *Int J Stem Cells.* 2020;13(1):1-12. doi 10.15283/IJSC19127
- Leung W.Y., Malkova A., Haber J.E. Gene targeting by linear duplex DNA frequently occurs by assimilation of a single strand that is subject to preferential mismatch correction. *Proc Natl Acad Sci USA.* 1997;94(13):6851-6856. doi 10.1073/pnas.94.13.6851
- Li L., Zhang S., Ge C., Ji L., Lv Y., Zhao C., Xu L., Zhang J., Song C., Chen J., Wei W., Fang Y., Yuan N., Wang J. HSCs transdifferentiate primarily to pneumocytes in radiation-induced lung damage repair. *Aging (Albany NY).* 2021;13(6):8335-8354. doi 10.18632/aging.202644
- Li N., Zhang L., Li H., Fang B. Human CD34⁺ cells mobilized by granulocyte colony-stimulating factor ameliorate radiation-induced liver damage in mice. *Stem Cell Res Ther.* 2010;1(3):22. doi 10.1186/SCRT22
- Li X., Zhang Y., Yeung S.C., Liang Y., Liang X., Ding Y., Ip M.S.M., Tse H.F., Mak J.C.W., Lian Q. Mitochondrial transfer of induced pluripotent stem cell-derived mesenchymal stem cells to airway epithelial cells attenuates cigarette smoke-induced damage. *Am J Respir Cell Mol Biol.* 2014;51(3):455-465. doi 10.1165/rcmb.2013-0529OC
- Liao F.L., Tan L., Liu H., Wang J.J., Ma X.T., Zhao B., Chen Y., Bihl J., Yang Y., Chen R.L. Hematopoietic stem cell-derived exosomes promote hematopoietic differentiation of mouse embryonic stem cells in vitro via inhibiting the miR126/Notch1 pathway. *Acta Pharmacol Sin.* 2018;39(4):552-560. doi 10.1038/aps.2017.130
- Likhacheva A.S., Rogachev V.A., Nikolin V.P., Popova N.A., Shilov A.G., Sebeleva T.E., Strunkin D.N., Chernykh E.R., Gel'fgat E.L., Bogachev S.S., Shurdov M.A. Involvement of exogenous DNA in the molecular processes in somatic cell. *Informatsionnyy Vestnik VOGIS = The Herald of Vavilov Society for Geneticists and Breeders.* 2008;12(3):426-473 (in Russian)
- Loh J.W., Guccione C., Di Clemente F., Riedlinger G., Ganesan S., Khiabani H. All-FIT: allele-frequency-based imputation of tumor purity from high-depth sequencing data. *Bioinformatics.* 2020;36(7):2173-2180. doi 10.1093/bioinformatics/btz865
- Maacha S., Sidahmed H., Jacob S., Gentilecore G., Calzone R., Grivel J.C., Cugno C. Paracrine mechanisms of mesenchymal stromal cells in angiogenesis. *Stem Cells Int.* 2020;2020:4356359. doi 10.1155/2020/4356359
- Maizels N., Davis L. Initiation of homologous recombination at DNA nicks. *Nucleic Acids Res.* 2018;46(14):6962-6973. doi 10.1093/nar/gky588
- Mendelenko M.M., Kravtsov S.A., Rodionov E.P. Clinical effectiveness of immunotherapy with extracorporeally activated blood mononuclear cells for complex treatment of sepsis. *Russ J Immunol.* 1997;2(3-4):191-198
- Mendes Filho D., Ribeiro P.D.C., Oliveira L.F., De Paula D.R.M., Capuano V., De Assunção T.S.F., Da Silva V.J.D. Therapy with me-

- senchymal stem cells in Parkinson disease. *Neurologist*. 2018;23(4): 141-147. doi 10.1097/NRL.0000000000000188
- Müller A.M., Huppertz S., Henschler R. Hematopoietic stem cells in regenerative medicine: astray or on the path? *Transfus Med Hemother*. 2016;43(4):247-254. doi 10.1159/000447748
- NanOnco Plus Panel v3.0 n.d. <https://nanodigmbio.com/Solid-Tumor/819.html>
- Oshikhmina S.G., Ruzanova V.S., Ritter G.S., Dolgova E.V., Kirikovich S.S., Levites E.V., Efremov Y.R., ... Ostanin A.A., Chernykh E.R., Kolchanov N.A., Proskurina A.S., Bogachev S.S. Concept of natural genome reconstruction. Part 4. Integration of extracellular double-stranded DNA fragments into the genome of hematopoietic stem cells and the formation of extrachromosomal intermediates. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed*. 2026;30(2):163-180. doi 10.18699/vjgb-26-18
- Pal'tsev A., Ovechkin A., Zakharova N., Rovina A., Leplina O., Shevela E., Ostanin A., Chernykh E. Cytokines in the treatment of a generalized surgical infection. *Anesteziol Reanimatol*. 2000;2: 27-30 (in Russian)
- Patinati N.A., Drize N.J. Clonal hematopoiesis and its role in the development of hematological diseases. *Gematologiya i Transfuziologiya = Russ J Hematol Transfusiol*. 2021;66(4):580-592. doi 10.35754/0234-5730-2021-66-4-580-592 (in Russian)
- Petrova D.D., Dolgova E.V., Proskurina A.S., Ritter G.S., Ruzanova V.S., Efremov Y.R., Potter E.A., Kirikovich S.S., Levites E.V., Taranov O.S., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bogachev S.S. The new general biological property of stem-like tumor cells. Part II: Surface molecules, which belongs to distinctive groups with particular functions, form a unique pattern characteristic of a certain type of tumor stem-like cells. *Int J Mol Sci*. 2022;23(24): 15800. doi 10.3390/IJMS232415800/S1
- Phillips R. Hematopoietic stem cells: concepts, assays, and controversies. *Semin Immunol*. 1991;3(6):337-347
- Pinho S., Frenette P.S. Haematopoietic stem cell activity and interactions with the niche. *Nat Rev Mol Cell Biol*. 2019;20(5):303-320. doi 10.1038/s41580-019-0103-9
- Potter E.A., Dolgova E.V., Proskurina A.S., Ruzanova V.S., Efremov Y.R., Kirikovich S.S., Oshikhmina S.G., ... Grivtsova L.U., Kolchanov N.A., Ostanin A.A., Chernykh E.R., Bogachev S.S. Stimulation of mouse hematopoietic stem cells by angiogenin and DNA preparations. *Braz J Med Biol Res*. 2024;57:e13072. doi 10.1590/1414-431X2024e13072
- Ritter G.S., Dolgova E.V., Petrova D.D., Efremov Y.R., Proskurina A.S., Potter E.A., Ruzanova V.S., Kirikovich S.S., Levites E.V., Taranov O.S., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bogachev S.S. The new general biological property of stem-like tumor cells. Part I. Peculiarities of the process of the double-stranded DNA fragments internalization into stem-like tumor cells. *Front Genet*. 2022;13:954395. doi 10.3389/fgene.2022.954395
- Ruzanova V.S., Oshikhmina S.G., Proskurina A.S., Ritter G.S., Kirikovich S.S., Levites E.V., Efremov Y.R., ... Ostanin A.A., Chernykh E.R., Kolchanov N.A., Dolgova E.V., Bogachev S.S. A concept of natural genome reconstruction. Part 2. Effect of extracellular double-stranded DNA fragments on hematopoietic stem cells. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed*. 2024;28(8): 993-1007. doi 10.18699/vjgb-24-106
- Ruzanova V.S., Oshikhmina S.G., Ritter G.S., Dolgova E.V., Kirikovich S.S., Levites E.V., Efremov Y.R., ... 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 Selekcii = Vavilov J Genet Breed*. 2025;29(4):479-495. doi 10.18699/vjgb-25-52
- Schema for Exome Probesets – Exome Capture Probesets and Targeted Region [www document]. URL https://genome.ucsc.edu/cgi-bin/hgTables?db=hg38&hgta_group=map&hgta_track=exomeProbesets&hgta_table=KAPA_HyperExome_hg38_capture_targets&hgta_doSchema=describe+table+schema (Accessed 5.27.25)
- Schwarting S., Litwak S., Hao W., Bähr M., Weise J., Neumann H. Hematopoietic stem cells reduce postischemic inflammation and ameliorate ischemic brain injury. *Stroke*. 2008;39(10):2867-2875. doi 10.1161/STROKEAHA.108.513978
- Shamal N. Somatic Mutations in Nature [www document]. 2017. Available at: <https://www.irb.basnet.by/ru/somaticheskie-mutacii-v-prirode/> (in Russian)
- Shlush L.I. Age-related clonal hematopoiesis. *Blood*. 2018;131(5): 496-504. doi 10.1182/blood-2017-07-746453
- Siminovitch L., McCulloch E.A., Till J.E. The distribution of colony-forming cells among spleen colonies. *J Cell Comp Physiol*. 1963;62: 327-336. doi 10.1002/jcp.1030620313
- Sleptsov A.A., Nazarenko M.S., Puzyrev V.P. Common in atherogenesis and carcinogenesis: clonal hematopoiesis. *Russ J Cardiol*. 2023; 28(10):57-65. doi 10.15829/1560-4071-2023-5511
- Spees J.L., Lee R.H., Gregory C.A. Mechanisms of mesenchymal stem/stromal cell function. *Stem Cell Res Ther*. 2016;7(1):125. doi 10.1186/s13287-016-0363-7
- Steensma D.P. Clinical implications of clonal hematopoiesis. *Mayo Clin Proc*. 2018;93(8):1122-1130. doi 10.1016/j.mayocp.2018.04.002
- Symington L.S. End resection at double-strand breaks: mechanism and regulation. *Cold Spring Harb Perspect Biol*. 2014;6(8):a016436. doi 10.1101/cshperspect.a016436
- Vahidy F.S., Rahbar M.H., Zhu H., Rowan P.J., Bambhroliya A.B., Savitz S.I. Systematic review and meta-analysis of bone marrow-derived mononuclear cells in animal models of ischemic stroke. *Stroke*. 2016;47(6):1632-1639. doi 10.1161/STROKEAHA.116.012701
- Vatolin S.Y., Okhapkina E.V., Matveeva N.M., Shilov A.G., Baiborodin S.I., Philimonenko V.V., Zhdanova N.S., Serov O.L. Scheduled perturbation in DNA during in vitro differentiation of mouse embryo-derived cells. *Mol Reprod Dev*. 1997;47(1):1-10. doi 10.1002/(SICI)1098-2795(199705)47:1<1::AID-MRD1>3.0.CO;2-R
- Velten L., Haas S.F., Raffel S., Blaszkiewicz S., Islam N., Hennig B.P., Hirche C., ... Ho A.D., Huber W., Trumpp A., Essers M.A.G., Steinmetz L.M. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*. 2017;19(4):271-281. doi 10.1038/ncb3493
- Vriend L.E.M., Krawczyk P.M. Nick-initiated homologous recombination: protecting the genome, one strand at a time. *DNA Repair (Amst)*. 2017;50:1-13. doi 10.1016/j.dnarep.2016.12.005
- Wang X., Ma S., Yang B., Huang T., Meng N., Xu L., Xing Q., ... Zhang T., Wu J., Yang G.L., Guan F., Wang J. Resveratrol promotes hUC-MSCs engraftment and neural repair in a mouse model of Alzheimer's disease. *Behav Brain Res*. 2018;339:297-304. doi 10.1016/j.bbr.2017.10.032
- Watcham S., Kucinski I., Gottgens B. New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing. *Blood*. 2019;133(13):1415-1426. doi 10.1182/blood-2018-08-835355
- Waterman R.S., Tomchuck S.L., Henkle S.L., Betancourt A.M. A new mesenchymal stem cell (MSC) paradigm: polarization into a pro-inflammatory MSC1 or an immunosuppressive MSC2 phenotype. *PLoS One*. 2010;5(4):e10088. doi 10.1371/journal.pone.0010088
- Watson C.J., Papula A.L., Poon G.Y.P., Wong W.H., Young A.L., Druley T.E., Fisher D.S., Blundell J.R. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science*. 2020; 367(6485):1449-1454. doi 10.1126/science.aay9333
- Weiskopf D., Weinberger B., Grubeck-Loebenstein B. The aging of the immune system. *Transpl Int*. 2009;22(11):1041-1050. doi 10.1111/j.1432-2277.2009.00927.x
- Wright K.T., El Masri W., Osman A., Chowdhury J., Johnson W.E.B. Concise review: bone marrow for the treatment of spinal cord injury: mechanisms and clinical applications. *Stem Cells*. 2011;29(2): 169-178. doi 10.1002/stem.570
- Xu S.Y. Sequence-specific DNA nicking endonucleases. *Biomol Concepts*. 2015;6(4):253-267. doi 10.1515/bmc-2015-0016

- Yakubov L.A., Rogachev V.A., Likhacheva A.C., Bogachev S.S., Sebeleva T.E., Shilov A.G., Baiborodin S.I., Petrova N.A., Mechetina L.V., Shurdov M.A., Wickstrom E. Natural human gene correction by small extracellular genomic DNA fragments. *Cell Cycle*. 2007;6(18):2293-2301. doi 10.4161/cc.6.18.4729
- Yakubov L.A., Taranov O.S., Sidorov S.V., Nikonov S.D., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bogachev S.S. The concept of natural genome reconstruction. Part 1. Basic provisions of the “natural genome reconstruction” concept. Changing the genome of hematopoietic stem cells using several natural cellular mechanisms that are inherent in the hematopoietic cell and determine its biological status as “the source of the body’s reparative potential”. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2024;28(7): 696-705. doi 10.18699/vjgb-24-78
- Yin C., Heit B. Cellular responses to the efferocytosis of apoptotic cells. *Front Immunol*. 2021;12:631714. doi 10.3389/fimmu.2021.631714
- Yuan N., Wei W., Ji L., Qian J., Jin Z., Liu H., Xu L., ... He Y., Wang M., Tang L., Fang Y., Wang J. Young donor hematopoietic stem cells revitalize aged or damaged bone marrow niche by transdifferentiating into functional niche cells. *Aging Cell*. 2023;22(8):e13889. doi 10.1111/ace1.13889
- Yunir E., Kurniawan F., Rezaprasga E., Wijaya I.P., Suroyo I., Matondang S., Irawan C., Soewondo P. Autologous bone-marrow vs. peripheral blood mononuclear cells therapy for peripheral artery disease in diabetic patients. *Int J Stem Cells*. 2021;14(1):21-32. doi 10.15283/ijsc20088
- Zhang M., Huang B. The multi-differentiation potential of peripheral blood mononuclear cells. *Stem Cell Res Ther*. 2012;3(6):48. doi 10.1186/scrt139
- Zilio N., Ulrich H.D. Exploring the SSBreakome: genome-wide mapping of DNA single-strand breaks by next-generation sequencing. *FEBS J*. 2021;288(13):3948-3961. doi 10.1111/FEBS.15568

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

Received July 5, 2025. Revised July 14, 2025. Accepted December 12, 2025.

doi 10.18699/vjgb-26-58

Effects of chronic psychosocial stress on the development of mouse oocytes and preimplantation embryos

T.N. Igonina ¹ , I.D. Chekashov ^{1, 2}, A.L. Levinson², S.Ya. Amstislavsky ¹¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia² Novosibirsk State University, Novosibirsk, Russia egik_00@mail.ru

Abstract. The impact of psychosocial chronic stress on mammalian oocyte maturation, fertilization and early stages of embryonic development remains poorly understood. This study addresses the effects of chronic psychosocial stress on the reproductive outcome in female mice, i.e. the development of *in vitro*- and *in vivo*-derived embryos. The model of chronic stress used in the study comprised a 21-day protocol consisting of a period of social isolation followed by the overcrowding. Two experiments were conducted, varying in the type of fertilization. In experiment 1, female mice were stressed at the folliculogenesis stages; then oocyte maturation and *in vitro* fertilization were performed, and the early development of the *in vitro*-derived embryos was studied. Experiment 2 differed in that fertilization was performed *in vivo*, and the resulting *in vivo*-derived embryos were cultured *in vitro* since the two-cell stage. To assess preimplantation embryo development, blastocysts were fixed, stained with the TUNEL/DAPI method and analyzed using fluorescence microscopy, i.e. the number of interphase nuclei and the apoptosis index were estimated. The results of Experiment 1 showed that chronic stress did not affect oocyte maturation or their fertilization capacity. However, embryos from the stress group contained fewer interphase nuclei ($p < 0.001$), which points to a lower cleavage rate. Meanwhile, the apoptosis rate in these blastocysts was comparable to controls. Experiment 2 showed that chronic stress caused a decrease in the proportion of embryos that achieved the blastocyst stage during the culture period and an increase in the proportion of morulae ($p < 0.01$), as well as a decrease in the number of interphase nuclei in blastocysts ($p < 0.001$). Experiments demonstrated that the chronic psychosocial stress exerts a moderate but significant effect on early embryonic development, primarily via the reduced proliferative activity of embryonic cells. These results obtained in mice have a translational value for reproductive medicine and highlight the importance of maternal stress when analyzing ART outcomes.

Key words: chronic psychosocial stress; oocytes; embryos; *in vitro* fertilization; early embryonic development; TUNEL assay; apoptosis; assisted reproductive technologies (ART)

For citation: Igonina T.N., Chekashov I.D., Levinson A.L., Amstislavsky S.Ya. Effects of chronic psychosocial stress on the development of mouse oocytes and preimplantation embryos. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2026;30(4):557-568. doi 10.18699/vjgb-26-58

Funding. The work was supported by the Russian Science Foundation, project No. 23–25–00139. Animal maintenance was supported by Russian Governmental Project No. FWNR-2026-0028 for ICG SB RAS.

Acknowledgements. The work was performed using the equipment of the Center for Genetic Resources of Laboratory Animals at the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, supported by the Ministry of Science and Higher Education of the Russian Federation (Unique Project Identifier RFMEF162119X0023).

Author contribution. Idea of the work and experimental design – S.Ya.A., T.N.I., A.L.L.; experiments – T.N.I., I.D.Ch.; processing and discussion of results – T.N.I., I.D.Ch., A.L.L., S.Ya.A.; manuscript writing and editing – T.N.I., I.D.Ch., A.L.L., S.Ya.A.

Эффекты хронического психосоциального стресса на развитие ооцитов и преимплантационных эмбрионов мышей

Т.Н. Игонина ¹ , И.Д. Чекашов ^{1, 2}, А.Л. Левинсон², С.Я. Амстиславский ¹¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия egik_00@mail.ru

Аннотация. Эффекты хронического психосоциального стресса на созревание ооцитов, оплодотворение и ранние этапы развития эмбрионов млекопитающих остаются недостаточно изученными. Настоящее исследование направлено на изучение влияния хронического психосоциального стресса на репродуктивные показатели самок мышей и развитие эмбрионов, полученных после оплодотворения *in vitro* и *in vivo*. Хронический стресс

моделировали с помощью 21-дневного протокола, включавшего период социальной изоляции, за которым следовало содержание в условиях высокой скученности. Проведены два эксперимента с воздействием стресса на стадии фолликулогенеза, различавшихся способом оплодотворения. В эксперименте 1 у самок мышей после воздействия стресса были извлечены ооциты, проведено дозревание ооцитов *in vitro*, экстракорпоральное оплодотворение и исследовано раннее развитие полученных эмбрионов. Эксперимент 2 отличался тем, что оплодотворение было *in vivo*, полученные эмбрионы культивировали *in vitro* с двухклеточной стадии. Для оценки преимплантационного развития эмбрионов бластоцисты фиксировали, окрашивали методом TUNEL/DAPI и анализировали с помощью флуоресцентной микроскопии, подсчитывая число интерфазных ядер и индекс апоптоза. Эксперимент 1 показал, что хронический стресс не оказывает влияния на дозревание ооцитов и их способность к оплодотворению. Однако эмбрионы из группы «стресс» содержали меньшее число интерфазных ядер ($p < 0.001$), что указывает на замедленную скорость дробления. Между тем уровень апоптоза в этих бластоцистах был сопоставим с таковым в контрольной группе. В эксперименте 2 хронический стресс вызвал снижение процентной доли эмбрионов, достигших стадии бластоцисты в течение периода культивирования, и увеличение доли морул ($p < 0.01$), а также снижение числа интерфазных ядер в бластоцистах ($p < 0.001$). Эксперименты показали, что хронический психосоциальный стресс оказывает умеренное, но значимое влияние на раннее эмбриональное развитие, главным образом за счет снижения пролиферативной активности клеток. Эти результаты, полученные на мышах, имеют трансляционное значение для репродуктивной медицины и подчеркивают важность учета материнского стресса при анализе исходов ВРТ.

Ключевые слова: хронический психосоциальный стресс; ооциты; эмбрионы; оплодотворение *in vitro*; раннее эмбриональное развитие; TUNEL-анализ; апоптоз; вспомогательные репродуктивные технологии (ВРТ)

Introduction

Chronic psychosocial stress has become an essential feature of life in a modern urban agglomeration. Urbanization is associated with a complex of predominantly psychological stressors, including social overload (crowding, excess of superficial contacts) combined with a deficit of meaningful connections, professional loads, and uncertainty. These stressors adversely affect the psychoemotional state and reproductive function (Ochnik et al., 2024; Orquiza, 2024).

Chronic stress activates the hypothalamic-pituitary-adrenal (HPA) axis, resulting in an increase in glucocorticoid synthesis (James et al., 2023). This affects the reproductive function through direct and indirect mechanisms (Zhai et al., 2020; Bhaumik et al., 2023; Jeon et al., 2023).

The direct effects are mediated by glucocorticoid receptors present in ovarian cells (granulosa, theca cells) and in the early embryo (Bhaumik et al., 2023; Jeon et al., 2023). In the ovary, this leads to the suppression of proliferation and induction of apoptosis of granulosa cells, disruption of steroidogenesis, and deterioration of oocyte quality (Prasad et al., 2016; Bhaumik et al., 2023). At the preimplantation stage of embryo development, elevated maternal corticosterone levels can directly suppress embryonic cell division, reduce the number of cells in the blastocyst, and lead to delayed hatching (Liu et al., 2012; Zhai et al., 2020).

An indirect mechanism of stress influence on reproduction is the suppression of the hypothalamic-pituitary-gonadal (HPG) axis. High glucocorticoid levels inhibit gonadotropin-releasing hormone (GnRH) secretion in the hypothalamus, which leads to a decrease in the production of gonadotropins, i. e. luteinizing (LH) and follicle-stimulating (FSH) hormones (Zhou et al., 2019). A lack of these hormones disrupts key ovarian processes: folliculogenesis, ovulation, and steroid synthesis (Zhai et al., 2020).

Such complex dysregulation causes risks for reproductive health and can contribute to the development of infertility, the prevalence of which reaches 12.6–17.5 % among couples of reproductive age (Cox et al., 2022). Assisted reproductive technologies (ARTs) are increasingly used to treat infertility in clinical practice (Abdullah et al., 2023). However, data are accumulating that chronic stress is a significant factor capable of influencing the effectiveness of ART (Zhou et al., 2019; Levinson et al., 2022). Most research in this area is focused on studying systemic hormonal changes (Valsamakis et al., 2019), while the effect of stress factors on oocyte maturation and early embryo development in the context of ART application remains poorly studied. The significance of such studies nowadays is high as the number of women resorting to ART procedures increases (Chambers et al., 2021).

Experiments on laboratory animals confirm that stressors of various natures disrupt the HPG axis, changing the hormonal profile in female mice; such changes negatively affect the female reproductive system (Zhai et al., 2020). However, there are only few studies that address the effects of psychosocial stress on female reproductive function and on the outcome of ARTs (Levinson et al., 2022).

In the present study, we used the model of chronic psychosocial stress previously adapted by us for female mice (Lebedeva et al., 2024; Igonina et al., 2024a, b). The main focus of the work was the study of oocyte maturation *in vitro* and *in vivo*, measured as the effectiveness of IVF, and the subsequent embryo development under chronic psychosocial stress. The novelty of the approach includes modeling chronic psychosocial stress in female mice combining with subsequent assessment of its impact on the key stages of ART protocols, from oocyte maturation to embryo culture, which ensures the high translational value of this work.

The aim of this study is to elucidate the effect of chronic psychosocial stress on the reproductive function in female mice and on early development of *in vitro*-derived and *in vivo*-derived embryos.

Materials and methods

Experimental animals. Studies were conducted on sexually mature females ($n = 27$) and fertile males ($n = 25$) of the CD1 mouse strain aged 2.5 months, raised in the specific pathogen free (SPF) vivarium of the Institute of Cytology and Genetics SB RAS (ICG SB RAS). Animals were housed in individually ventilated OptiMice cages (Animal Care, United States) measuring $34.3 \times 29.2 \times 15.5$ cm (length $\times$ width $\times$ height) at a temperature of $22\text{--}24^\circ\text{C}$ and relative humidity of $40\text{--}50\%$. Sterile fractionated birch chips for laboratory animals (TU 16.10.23-001-0084157135-2019) were used as bedding. All animals had free access to standardized autoclaved Delta Feeds chew for laboratory mice and rats (BioPro, Russia) and purified water “Severyanka” (Ekoproekt, Russia) enriched with mineral supplements. A light regime of 12 h light/12 h dark (lights on at 3:00 am) was maintained throughout the entire experiment period.

Chronic psychosocial stress modeling. Animals of the experimental group were kept under conditions of social isolation (one female per cage) for 11 days; then they were transferred to crowding conditions (social overload): 11 females per cage for the next 10 days. The total duration of the stress protocol was 21 days. Animals in the control group were kept under standard vivarium conditions: groups of 5 females per cage. This stress model was verified by evaluation of corticosterone levels in the blood serum.

Experimental design. Female mice were randomly distributed between two groups: control group – no stress; stress group – chronic psychosocial stress according to the protocol described above.

To evaluate the effect of chronic psychosocial stress on reproductive function and early embryonic development, two sequential experiments were conducted, differing in the method of fertilization (Fig. 1). All the females used in the experiments were hormonally stimulated for superovulation (protocol presented below).

Experiment 1. Effect of stress on oocyte maturation, *in vitro* fertilization, and early embryo development.

Females from both groups (control, $n = 6$; stress, $n = 6$) were euthanized four hours after hCG administration by plunging to carbon dioxide (CO_2). Blood and ovaries were collected. Immature oocytes at the germinal vesicle (GV) stage were separated from ovarian tissue. Oocytes were matured *in vitro*, then fertilized using IVF, and the resulting embryos were cultured in a CO_2 incubator for 96 hours. The *in vitro*-derived embryos were then fixed in 4 % formalin dissolved in phosphate-buffered saline, stained with TUNEL/DAPI, and analyzed using fluorescence microscopy.

Experiment 2. Effect of stress on the development of *in vivo* fertilized embryos. Females from both groups (control, $n = 8$; stress, $n = 7$) were mated with fertile males. The animals were euthanized 48 hours after pairing using carbon dioxide, and blood was collected. Embryos were collected from the oviducts at the stage of 2–4-cells. The collected embryos were cultured *in vitro* in a CO_2 incubator for 68 hours. After culturing, the embryos were fixed, stained with TUNEL/DAPI, and analyzed using fluorescence microscopy.

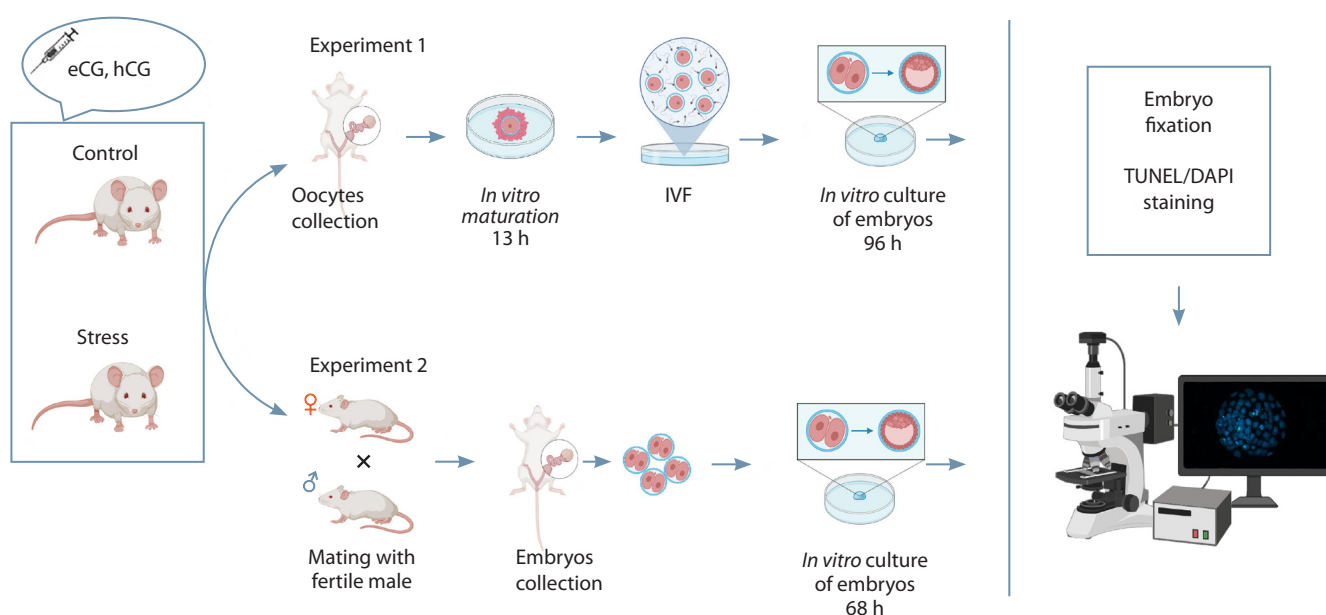


Fig. 1. Experimental design.

eCG – equine chorionic gonadotropin, hCG – human chorionic gonadotropin.

Corticosterone assay in blood serum. Blood (0.5 ml) was collected into tubes immediately after decapitation of the animals. Samples were incubated for 25 min at room temperature for clot formation and centrifuged (1,000g, 10 min; Hermle Z 326, Germany). The resulting serum was stored at -80°C . Serum corticosterone concentration was determined by enzyme-linked immunosorbent assay using the Corticosterone rat/mouse-IFA kit (Chema, Russia) according to manufacturer's recommendations. Optical density was measured at 450 nm on a plate spectrophotometer (Thermo Fisher Scientific, United States) 15 min after stopping the reaction. The concentration was calculated using a calibration curve.

Hormonal stimulation of ovaries. For superovulation, mice were injected intraperitoneally with 7.5 IU of equine chorionic gonadotropin (eCG, folligon, Intervet, Netherlands), and after 46 hours, 7.5 IU of human chorionic gonadotropin (hCG, chorulon, Intervet, Netherlands) in accordance with the standard superovulation protocol accepted for mice (Shindo et al., 2022).

Immature oocyte collection and *in vitro* maturation. Euthanasia was performed 4 hours after hCG injection. Ovaries were removed from the surrounding tissues and placed into pre-warmed FertiCult Flushing medium (FertiPro, Belgium). Immature oocytes at the germinal vesicle (GV) stage were carefully separated from ovarian tissue using a thin needle and tweezers. Oocytes were transferred onto Petri dishes (Corning Incorporated, United States), placed into 50 μL drops of HTF medium pre-equilibrated in a CO_2 incubator, and covered with sterile mineral oil. The culturing was carried out in a CO_2 incubator (New BrunswickTM Galaxy 48R, Eppendorf, Germany) at 37°C , 5 % CO_2 , and 90 % humidity for 13–14 hours. Oocyte maturity was assessed by the presence of the first polar body; matured oocytes were then used for IVF.

Collection of epididymal sperm. Epididymides were taken from euthanized males. The tissue was placed in 100 μL of pre-warmed HTF medium onto Petri dishes (Corning Incorporated, United States) and cut to allow sperm cells to flow out. The dishes with sperm cells were incubated for 30 minutes at 37°C and 5 % CO_2 aiming at spermatozoa capacitation. Then 10 μL of the suspension was transferred to 60 μL of fresh HTF medium and incubated for another 30 minutes. The resulting material was used for IVF.

***In vitro* fertilization.** Mature oocytes were incubated for 30 minutes in 60 μL drops of pre-equilibrated HTF medium under mineral oil (FertiPro, Belgium). Then 10 μL of epididymal sperm suspension was added, and the dishes with medium drops containing oocytes and sperm cells were placed in a CO_2 incubator for 4 hours. After incubation, oocytes were washed and transferred to KSOM AA medium for further culturing. Successful fertilization was confirmed by the presence of two pronuclei in the oocytes.

Collection of *in vivo*-derived embryos. Immediately after hCG injection, females were mated with fertile males. Mating success was determined after 18 hours by the presence of a copulatory plug and/or spermatozoa in vaginal smears. The fertilized mice were euthanized 48 hours after hCG injection. Then the oviducts and uterus were removed and flushed with FertiCult Flushing medium solution (FertiPro, Belgium) to obtain embryos. Embryos at the 2–4 blastomere stage were collected for *in vitro* cultivation.

***In vitro* embryo culture.** Embryos were cultured in KSOM AA medium in 20 μL drops under mineral oil in a CO_2 incubator (New BrunswickTM Galaxy 48R, Eppendorf, Germany) at 37°C , 5 % CO_2 , and 90 % humidity.

The embryo quality and the rate of embryo development were assessed daily using light microscopy (Leica S80 APO stereomicroscope, Germany). For embryos at cleavage stages, the number, shape, and size of blastomere, as well as their general morphology, were analyzed. For blastocysts, blastocoel size, inner cell mass (ICM) compactness, and trophectoderm (TE) cell density were assessed. An embryo was considered normally developing if it corresponded to the expected stage after the culture period and had no signs of damage (vacuolization, cytoplasm granulation), the degree of fragmentation did not exceed 20 %, blastocysts had a compact ICM, and the blastocoel was surrounded by TE cells. If there was no cell number increase in a particular embryo or the transition to the next developmental stage within 24 hours of observation not happened within 24 hours of *in vitro* culturing, embryos were classified as non-developing. In addition to the qualitative assessments (developing/non-developing), the percentage of embryos that reached each specific stage was used in the subsequent statistical analysis of *in vitro* development dynamics.

DAPI and TUNEL staining of blastocysts followed by fluorescence microscopy. Embryos were fixed and stained using the TUNEL FITC kit (Vazyme, China) according to the manufacturer's protocol. After three washes with phosphate-buffered saline, samples were treated with proteinase K (5 min), incubated in equilibration buffer (15 min), and then in TUNEL solution (1 hour, 37°C). Thereafter, the embryos were washed, stained with DAPI (5 min), and mounted on slides. Microscopic evaluation was performed with a Zeiss Axio Imager 2 microscope (Germany) using DAPI-FITC filters. Nuclei and apoptotic cells were counted using ImageJ software. The apoptosis index was calculated as the ratio of apoptotic cells to the total number of cells.

We analyzed 124 randomly selected embryos obtained *in vitro* by IVF (65 in the control group and 59 in the stress group) and 274 embryos obtained *in vivo* after natural mating (166 in the control group and 108 in the stress group).

Preparation of culture media. Human tubal fluid (HTF) medium (Quinn et al., 1985) was prepared for IVF. A modification of this medium (HTFm) was used for oocyte matu-

ration. Embryos were cultured in KSOM AA according to (Chatot et al., 1989). All media were prepared independently from individual components (Table S1)¹ in purified water, sterilized by filtration through a 0.22- μ m membrane, and equilibrated in a CO₂ incubator under a layer of mineral oil at 37 °C for ≥ 4 h.

Statistical analysis. Data were analyzed using STATISTICA v. 8.0 software package (StatSoft Inc). Normality of distribution was checked by the Shapiro–Wilk test. Parameters with normal distribution of data (corticosterone level, percentage of mature and fertilized oocytes) were compared using Student’s *t*-test. The dynamics of embryo development were assessed by two-way repeated measures ANOVA. Data are presented as $M \pm SEM$. Parameters with non-normal distribution (number of interphase nuclei, apoptosis index) were analyzed using the Mann–Whitney U-test and presented as median with 25 % and 75 % quartiles. Data on embryo development stages were compared using the χ^2 (chi-square) test. The significance level was $p < 0.05$.

Results

Verification of the stress model

The level of corticosterone in the blood serum in females subjected to chronic psychosocial stress was significantly higher ($p < 0.05$, Student’s *t*-test) than in the control group (Fig. 2). This result verifies the stress model used.

Effect of stress on oocyte maturation, *in vitro* fertilization, and early embryo development

From females of the control group ($n = 6$) and the group subjected to chronic stress ($n = 6$), 184 and 140 immature oocytes at the germinal vesicle stage (GV) were obtained, respectively. After *in vitro* culture for 13–14 hours, 155 (84.2 %) oocytes in the control group and 105 (75.0 %) in the stress group matured to the MII stage. Their subsequent fertilization by IVF resulted in zygote formation in 80.4 % (119/155) in the control group and 84.6 % (88/105) in the stress group. Student’s *t*-test of the obtained data revealed no significant differences between the groups either in the percentage of oocyte maturation ($p > 0.05$) or in fertilization efficiency ($p > 0.05$) (Fig. 3a).

Embryo development *in vitro* proceeded similarly in the control and stress groups at all stages of culturing (Fig. 3b; Table 1). Two-way repeated measures ANOVA revealed no significant effect of chronic psychosocial stress: there was neither a main effect of the “Stress” factor ($F(1,10) = 1.2$, $p > 0.05$) nor its interaction with the duration of culturing ($F(3,30) < 1$, $p > 0.05$). Instead, a significant effect of the “Duration of culturing” factor was revealed ($F(3,30) = 7.8$, $p < 0.01$), reflecting the general dynamics of embryonic development *in vitro*.

Embryos obtained by IVF from the stressed females, after 96 hours of *in vitro* culture, were characterized by a significantly lower number of interphase nuclei compared with the

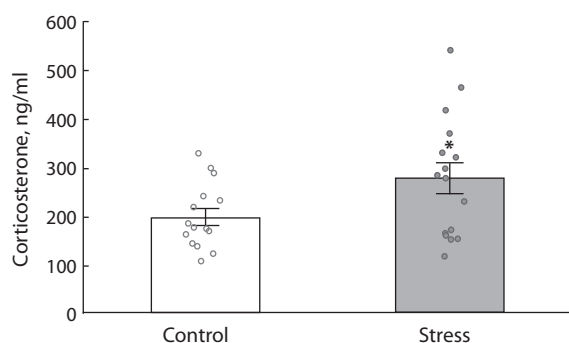


Fig. 2. Blood serum corticosterone concentration in control and chronically stressed animals.

Data are presented as $M \pm SEM$. Number of animals/samples: $n = 15$ (control group); $n = 16$ (stress group). * $p < 0.05$ compared with the control group (Student’s *t*-test).

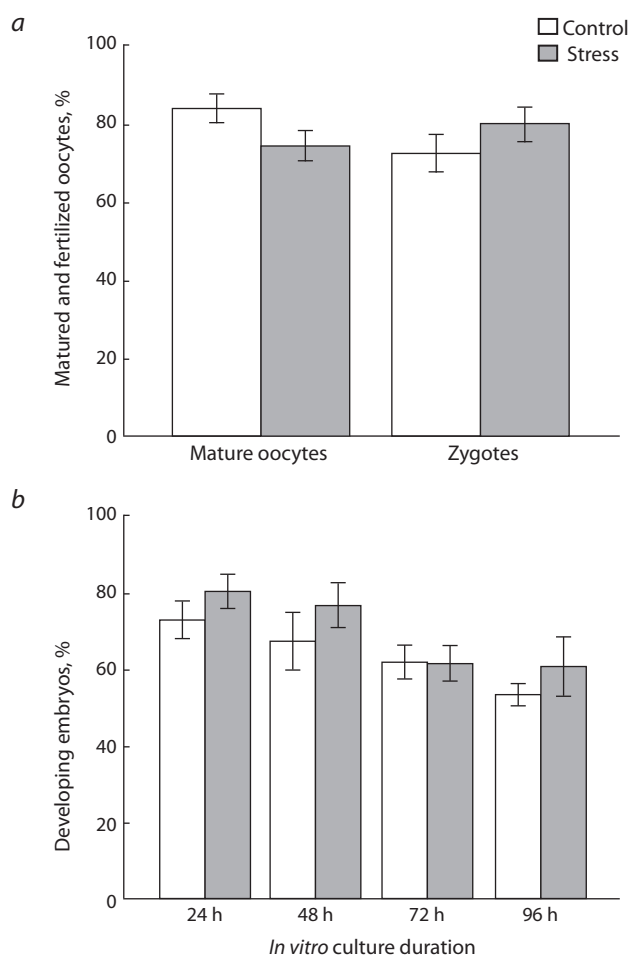


Fig. 3. Effect of chronic psychosocial stress on oocyte maturation, their fertilization potential, and *in vitro* embryo development.

a – left: oocytes that reached MII stage (number of mature oocytes in the control group: $n = 155$, in the stress group: $n = 105$); right: percentage of zygotes formed after IVF (number of zygotes in the control group: $n = 119$, in the stress group: $n = 88$); b – embryo development at different intervals of culturing (initial number of embryos placed in culture in the control group: $n = 119$, in the stress group: $n = 88$). Data are presented as $M \pm SEM$ (average per female).

¹ Supplementary Table S1 and Figure S1 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Igonina_Engl_30_4.pdf

Table 1. Stages of development of *in vitro* fertilized embryos after 96 hours of culture

Group	Morulae	Blastocysts	Developmental arrest
			% (n)
Control	43.1 (28)	50.8 (33)	6.2 (4)
Stress	39.0 (23)	44.1 (26)	16.9 (10)

control group ($p < 0.001$, Mann–Whitney U-test; Fig. 4a). The proportions of apoptotic nuclei did not differ significantly (Fig. 4b).

**Effect of stress on the development
of *in vivo* fertilized embryos**

From the control group females ($n = 8$) and the females subjected to chronic stress ($n = 7$), 139 and 115 embryos were obtained, respectively. In the control group, 127 (91.4 %) embryos were at the 2-cell stage, and 12 (8.6 %) were at the 3–4-cell stage; in the stress group, 96 (83.5 %) and 19 (16.5 %) embryos were at the 2-cell and 3–4-cell stages, respectively.

Chronic psychosocial stress did not affect the fertility of female mice, as there was no difference between the control (17.4 ± 2.3) and stressed (16.4 ± 1.8) females ($p > 0.05$, Student’s *t*-test) in the average number of embryos per female on day 2 of *in vivo* development.

The rates of the subsequent *in vitro* embryo development were comparable in the control and stress groups throughout the entire culture period (Fig. S2). Repeated measures ANOVA revealed no statistically significant effect of the “Stress” factor ($F(1,13) = 1.20$; $p > 0.05$), the “Duration of culturing” factor ($F(2,26) = 2.06$; $p > 0.05$), or interaction between these factors ($F(2,26) = 2.82$; $p > 0.05$).

At the late stage of *in vitro* culturing (68 h), shifts in the distribution of embryo development stages were revealed in the group subjected to chronic stress: the proportion of blastocysts was lower, and the proportion of morulae was higher compared with the control group ($p < 0.01$; Table 2).

Subsequent analysis of embryos cultured *in vitro* for 68 h revealed a significant decrease in the number of interphase nuclei in the stress group compared with the control group ($p < 0.001$, Mann–Whitney U-test; Fig. 5a). This result points

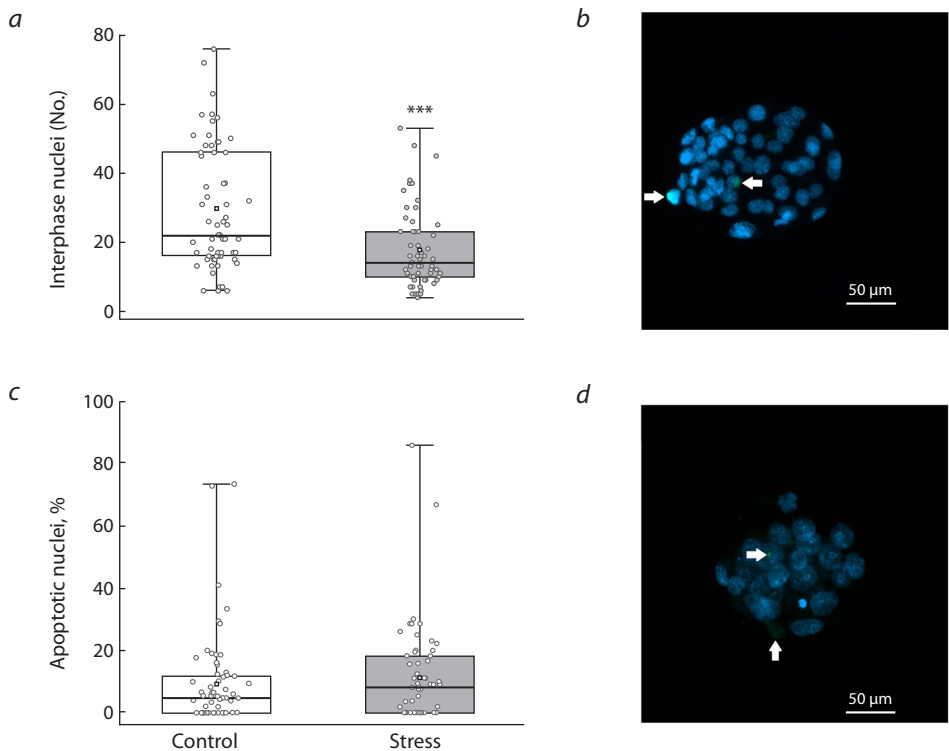


Fig. 4. Effect of chronic psychosocial stress on the number of interphase nuclei (a) and the percentage of apoptotic nuclei (c) in mouse embryos fertilized *in vitro*. Representative images of control (b) and stress (d) group embryos after DAPI (blue, nuclei) and TUNEL (green, apoptotic cells, indicated by arrows) staining. Scale bar 50 μm. Data are presented as median with 25 % and 75 % quartiles. *** $p < 0.001$ (Mann–Whitney U-test). Number of embryos studied: $n = 65$ in the control group, $n = 59$ in the stress group.

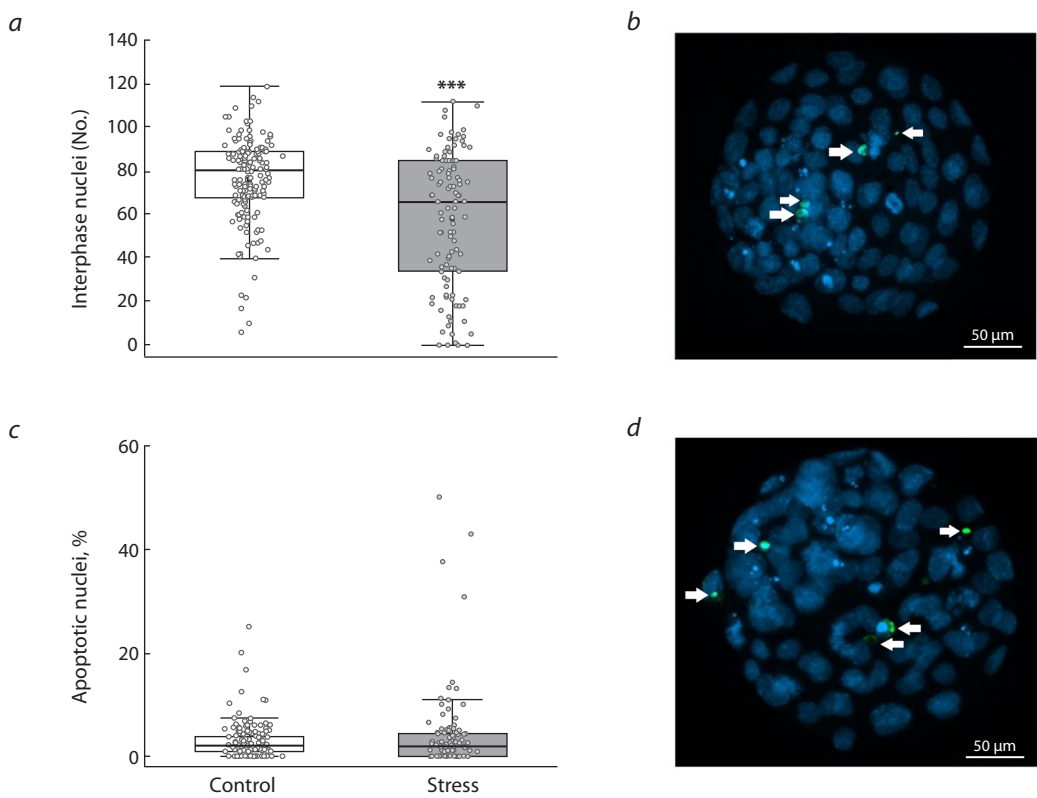


Fig. 5. Effect of chronic psychosocial stress on (a) the number of interphase nuclei and (c) the percentage of apoptotic nuclei in *in vivo* fertilized mouse embryos after 68 hours of *in vitro* culture. Representative images of (b) control and (d) stress group embryos after DAPI (blue, nuclei) and TUNEL (green, apoptotic cells, indicated by arrows) staining. Scale bar 50 μ m. Data are presented as median with 25 % and 75 % quartiles. *** $p < 0.001$ (Mann–Whitney U-test). Numbers of embryos studied: $n = 166$ in the control group, $n = 108$ in the stress group.

Table 2. Stages of development of *in vivo* fertilized embryos after 68 hours of *in vitro* culture

Group	Morulae	Blastocysts	Developmental arrest
	% (n)		
Control	1.7 (3)	94.2 (162)	4.1 (7)
Stress	8.8 (10)**	82.5 (94)**	8.8 (10)

** $p < 0.01$ compared with the control group (χ^2 test).

to a decrease in the proliferative activity of embryonic cells under the influence of stress. However, the proportion of apoptotic cells in the embryos did not differ significantly between the groups ($p > 0.05$, Mann–Whitney U-test), demonstrating the absence of a pronounced effect of stress on the rates of programmed cell death at this stage of development (Fig. 5b).

Discussion

Hans Selye was one of the first to define stress from a biological point of view as a nonspecific reaction of the body to a threat to homeostasis (Selye, 1950). Stressors, perceived as a

real or potential danger, trigger a cascade of neuroendocrine reactions, the activation of the HPA axis in particular (James et al., 2023). Depending on the duration of exposure, stressors can cause acute or chronic stress (Godoy et al., 2018). In addition, stressors are classified by their nature: physiological, which are associated with a threat to physical well-being, and psychological, which are associated with psychoemotional and psychosocial effects (Godoy et al., 2018). Chronic stress is known to lead to various pathologies (Kalisch et al., 2024). Studies of the effect of chronic psychoemotional stress on the reproductive system often employ models involving mice, such as “chronic unpredictable stress” and “restriction” (Gao Y. et al., 2016; Levinson et al., 2022). These stressors have both psychological and physical effects, and can cause undesirable consequences, including weight loss, adversely affecting the general condition of the animals (Gao Y. et al., 2016; Furman et al., 2022). In some studies, presenting a predator, e. g. a hungry cat, to a mouse, is used as a model of psychoemotional stress, which creates excessive psychogenic impact associated with a potential threat to life (Liu et al., 2012). Meanwhile, the stressful effects that a person faces in a modern megalopolis are not associated with an immediate threat to life and are mostly

psychosocial in nature, while physical stressors are nowadays less relevant (Orquiza, 2024).

In a modern urban environment, a person may simultaneously experience a deficit of meaningful social connections and an abundance of superficial or unwanted contacts associated with high population density and limited privacy (Ham-moud et al., 2021; Zhang Z. et al., 2023). This combination of factors enhances the experience of loneliness and social stress, which is confirmed by data on the link between urban crowding and increased physiological stress responses (Lederbogen et al., 2011). Alternating periods of social isolation and forced involvement in intensive social interaction can exacerbate the psychoemotional load and may have a negative impact on the psychophysiological state of a person (Hawkley, Cacioppo, 2010; Lederbogen et al., 2011). Animal studies show that increasing population density and associated crowding exert a pronounced effect on reproductive processes (Suvorov, 2021).

In studies performed on laboratory animals, the effects of social isolation and crowding are addressed so far not in combination, but separately (Gavrilov et al., 2022). In the present study, we used a model based on the alternation of periods of isolation and subsequent overcrowding. A model of this sort was initially offered for rats (Gadek-Michalska et al., 2019) and later adapted for female mice in our studies (Lebedeva et al., 2024; Igonina et al., 2024a, b). Unlike traditional stress models, this model seems to be more relevant to the social conditions faced by women in a modern megalopolis (Ochnik et al., 2024; Orquiza, 2024).

Our results showed that alternating periods of isolation and overcrowding are accompanied by an increase in corticosterone levels in the blood of female mice, which confirms the effectiveness of this model for inducing stress. The results of the present study showed that psychosocial stress did not disrupt oocyte maturation *in vitro* or affect the oocyte ability to be fertilized. We had demonstrated previously that under *in vivo* conditions, chronic psychosocial stress led to a decrease in oocyte quality and maturity, accumulation of reactive oxygen species (ROS) in such oocytes, and enhanced apoptosis in cumulus cells (Lebedeva et al., 2024; Igonina et al., 2024b). The pronounced effect of stress on oocyte maturation *in vivo* versus its absence *in vitro* is probably explained by the fact that culture media lack the physiological microenvironment necessary for the transmission of the stress signal. This microenvironment, including follicle cells, paracrine and endocrine signals, is known to be necessary for normal oocyte maturation process *in vivo* (Gilchrist et al., 2004; Yeo et al., 2009). We assume that stress primarily disrupts these communication pathways (Huang, Zeng, 2021; Igonina et al., 2024a). As the direct effect of stress on the oocyte appears to be less significant than its indirect effects through follicular cells and hormonal background (Prasad et al., 2016), oocytes can maintain competence *in vitro* conditions, where systemic influence is minimized, even after exposure of the female to stress.

Animal studies employing various stress models show that chronic stress causes similar disturbances in folliculogenesis, leading to oocyte maturation impairment. Housing animals under crowded conditions led to high levels of corticosterone and angiotensin II, disrupted hormonal balance, induction of granulosa cell apoptosis, and reduced ovarian reserve (Kim, You, 2022). Stress associated with immobilization caused disruption of meiosis and oocyte death, leading to a decrease in ovarian reserve (Jiang et al., 2023) and an increase in the number of abnormal oocytes (Tsuji et al., 2021). In addition, this type of stress led to disruption of meiotic spindle morphology in oocytes (Sun et al., 2018) and reduced their developmental potential both *in vitro* and *in vivo* (Zhang S.Y. et al., 2011; Wu X.F. et al., 2015; Zhao et al., 2020). Chronic unpredictable stress had a similar effect, causing a decrease in ovarian reserve (Gao L. et al., 2020), deterioration of quality, and decrease in the number of oocytes (Wu L.M. et al., 2012a, b). In some cases, it led to the state of anovulation (Kala, Nivsarkar, 2016), disrupted granulosa cell proliferation, and accelerated their aging (Sun et al., 2021; Ma et al., 2024).

The primary systemic pathway mediating the effects of chronic stress is activation of the HPA axis and a sustained increase in circulating glucocorticoid levels (James et al., 2023). Until recently, the literature addressing the question whether or not glucocorticoids directly affect oocytes was area of controversy (Bhaumik et al., 2023); however, it has been shown not long ago that mouse oocytes lack glucocorticoid receptors (GR) (Cincotta et al., 2024). Nevertheless, it is well documented that oocyte maturation is sensitive to elevated glucocorticoid levels via indirect effects mediated by follicular somatic cells, i. e. granulosa and cumulus cells, which express GR (Bhaumik et al., 2023; Cincotta et al., 2024). Glucocorticoid action is realized predominantly through the classical genomic pathway involving binding to the cytoplasmic GR, its translocation to the nucleus, and subsequent regulation of transcription, as well as through rapid non-genomic and isoform-specific mechanisms, resulting in a complex and tissue-specific cellular response (Bhaumik et al., 2023).

In addition, activation of the HPA axis suppresses the HPG axis, including effects at the levels of the hypothalamus and pituitary gland (Bhaumik et al., 2023). It leads to reduced synthesis of sex steroids and disruption of the hormonal milieu required for normal folliculogenesis (Kalantaridou et al., 2004; Whirlledge, Cidlowski, 2013). It is worth noting that hormonal stimulation of superovulation was used in the present study, which may compensate for this deficiency and, thereby, attenuate this pathway of stress influence. However, as demonstrated by our data and by results of some other studies, stress continues to exert a negative impact on oocytes through additional mechanisms, primarily via direct disruption of the local ovarian microenvironment, manifesting itself as the induction of granulosa cell apoptosis and alterations in the secretion of paracrine factors (Prasad et al., 2016; Lebedeva et

al., 2024; Igonina et al., 2024a). A key element of this process is the action of glucocorticoids on follicular somatic cells (Bhaumik et al., 2023; Cincotta et al., 2024).

Thus, glucocorticoid signaling alters the metabolism of follicular somatic cells, the secretion of local regulatory factors, and the redox status, which in turn induces changes of the oocyte cytoplasm, including the accumulation of maternal mRNA, protein, and epigenetic enzymes (Wu X.F. et al., 2015; Joseph, Whirlledge, 2017; Sun et al., 2018). The stress-induced deficit in metabolic and hormonal support from the follicular epithelium brings about suboptimal conditions for the final stages of oocyte maturation.

It should be noted that, in addition to indirect effects mediated by somatic cells, direct sensitivity of oocytes to certain stress-induced signals and hormones, such as corticotropin-releasing factor, has been demonstrated (Zaidan et al., 2013). However, the role of this pathway appears to be substantially smaller than the complex effects mediated through alterations of the follicular microenvironment.

The results of the present study show that embryos obtained from stressed females are characterized by a reduced number of cells in blastocysts regardless of the fertilization method, and, in the case of *in vivo* fertilization, also by a delay in reaching the blastocyst stage. The time needed to achieve the blastocyst stage and the number of cells in the blastocyst are key indicators of preimplantation development in mammals (Ajduk, Zernicka-Goetz, 2012). A reduction in cell number, in particular, may point to a slowdown in cleavage rates and, more generally, it may be indicative of a low rate of early embryonic development (Nishizono et al., 2017).

A crucial aspect of our findings is that the reduction in blastocyst cell number was observed regardless of whether embryos were obtained via natural *in vivo* fertilization or by *in vitro* fertilization of oocytes retrieved from a stress-altered microenvironment but matured *in vitro* under controlled conditions. Our data support the previously proposed hypothesis that stress-induced changes in the oocyte transcriptome are manifested in embryos derived from these oocytes (Roth, 2018). It appears likely that stress exposure during the early stages of folliculogenesis leads to persistent functional alterations in oocytes, which are attributable, in particular, to epigenetic reprogramming.

Indeed, the earliest preimplantation stages of mammalian embryonic development critically depend on the pool of maternal mRNAs and proteins accumulated in oocytes, the synthesis of which is completed prior to fertilization (Wang et al., 2019; Kojima et al., 2025). These maternal transcripts provide the molecular basis for zygotic genome activation and support rapid cleavage cycles during the early stages of embryonic development, until the embryo's own transcriptome begins to govern developmental processes (Wang et al., 2019; Kojima et al., 2025). Stress-induced alterations in the abundance, composition, or functional activity of these molecules can disrupt

this delicate balance, thereby adversely affecting reproductive success and subsequent stages of embryo development (Roth, 2018). This mechanism may explain the phenomenon observed in our study as well as in others: a reduction in blastocyst cell number and delayed development in embryos from stressed females with the absence of apparent morphological defects in the oocytes (Liu et al., 2012; Casillas et al., 2021).

The hypothesis that the pre-fertilization influence over the oocyte can program embryonic development appears to be broadly applicable, as it is supported by studies not only in mammals (Roth, 2018) but also in fish (Li et al., 2012). In fish, exposure of oocytes to cortisol prior to fertilization results in persistent changes in the embryonic transcriptome and modulates embryonic growth (Li et al., 2012). Importantly, in mammals, unlike in fish, this mechanism appears to be indirect. Since mouse oocytes do not express classical glucocorticoid receptors (Cincotta et al., 2024), the stress signal mainly disrupts the function of follicular somatic cells, which in turn creates conditions for latent reprogramming of the developing gamete.

Thus, our results demonstrate that chronic psychosocial stress, mediated by activation of the HPA axis and disruption of the ovarian microenvironment, induces functional reprogramming of oocytes. This reprogramming does not affect oocyte morphology but predetermines reduced embryo quality and delayed embryonic development. The identified mechanism provides a biological rationale for clinically observed associations between stress and reduced efficacy of assisted reproductive technologies (ART) (Koumparou et al., 2021; Abdoli et al., 2025; Lu et al., 2025). In contrast to the existing inconsistencies in epidemiological data, which may be related to difficulties in the objective assessment of stress (Zanettoullis et al., 2024; Hu et al., 2025), our study indicates that exposure to chronic stress can generate latent oocyte defects with potential consequences for the successful application of reproductive technologies.

Conclusion

Our study supports the hypothesis of the negative impact of chronic psychosocial stress on reproductive function and provides an experimental basis for the interpretation of clinical observations. The model adequately reproduces key aspects of stress exposure experienced by residents of modern megalopolises. Investigation of the effects of psychoemotional stress on the ovary (Igonina et al., 2024a) and the consequences of these alterations for oocyte development and early embryogenesis opens avenues for the development of novel strategies to mitigate these changes (Wu L.M. et al., 2012b; Igonina et al., 2024b). Integration of such approaches into clinical practice may enable the development of personalized infertility treatment strategies aimed at reducing the negative impact of chronic stress and improving the outcomes of ART programs.

Ethical standards compliance. All studies were approved by the Bioethics Committee of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (protocol No. 143 of March 15, 2023) and complied with the requirements of the European Directive 2010/63/EU on the protection of animals used for scientific purposes.

References

- Abdoli S., Gholami A.H., Masoumi S.Z., Najafi-Vosough R., Azimi M., Jenabi E., Aliabadi S., Soltanian A.R., Ghaleiha A., Pilehvari S. The association of anxiety and perceived stress with *in vitro* fertilization outcomes in infertile women: a cross-sectional study. *J Hum Reprod Sci.* 2025;18(1):16-22. doi 10.4103/jhrs.jhrs_168_24
- Abdullah K.A.L., Atazhanova T., Chavez-Badiola A., Shivhare S.B. Automation in ART: paving the way for the future of infertility treatment. *Reprod Sci.* 2023;30(4):1006-1016. doi 10.1007/s43032-022-00941-y
- Ajduk A., Zernicka-Goetz M. Advances in embryo selection methods. *F1000 Biol Rep.* 2012;4:11. doi 10.3410/B4-11
- Bhaumik S., Lockett J., Cuffe J., Clifton V.L. Glucocorticoids and their receptor isoforms: roles in female reproduction, pregnancy, and foetal development. *Biology.* 2023;12(8):1104. doi 10.3390/biology12081104
- Casillas F., Betancourt M., Juárez-Rojas L., Ducolomb Y., López A., Ávila-Quintero A., Zamora J., Ommati M.M., Retana-Márquez S. Chronic stress detrimentally affects *in vivo* maturation in rat oocytes and oocyte viability at all phases of the estrous cycle. *Animals (Basel).* 2021;11(9):2478. doi 10.3390/ani11092478
- Chambers G.M., Dyer S., Zegers-Hochschild F., de Mouzon J., Ishihara O., Banker M., Mansour R., Kupka M.S., Adamson G.D. International Committee for Monitoring Assisted Reproductive Technologies world report: assisted reproductive technology, 2014. *Hum Reprod.* 2021;36(11):2921-2934. doi 10.1093/humrep/deab198
- Chatot C.L., Ziomek C.A., Bavister B.D., Lewis J.L., Torres I. An improved culture medium supports development of random-bred 1-cell mouse embryos *in vitro*. *J Reprod Fertil.* 1989;86(2):679-688. doi 10.1530/jrf.0.0860679
- Cincotta S.A., Richardson N., Foecke M.H., Laird D.J. Differential susceptibility of male and female germ cells to glucocorticoid-mediated signaling. *eLife.* 2024;12:RP90164. doi 10.7554/eLife.90164
- Cox C.M., Thoma M.E., Tchangalova N., Mburu G., Bornstein M.J., Johnson C.L., Kiarie J. Infertility prevalence and the methods of estimation from 1990 to 2021: a systematic review and meta-analysis. *Hum Reprod Open.* 2022;2022(4):hoac051. doi 10.1093/hropen/hoac051
- Furman O., Tsoory M., Chen A. Differential chronic social stress models in male and female mice. *Eur J Neurosci.* 2022;55(9-10):2777-2793. doi 10.1111/ejn.15481
- Gadek-Michalska A., Tadeusz J., Bugajski A., Bugajski J. Chronic isolation stress affects subsequent crowding stress-induced brain Nitric Oxide Synthase (NOS) isoforms and Hypothalamic-Pituitary-Adrenal (HPA) axis responses. *Neurotox Res.* 2019;36(3):523-539. doi 10.1007/s12640-019-00067-1
- Gao L., Zhao F., Zhang Y., Wang W., Cao Q. Diminished ovarian reserve induced by chronic unpredictable stress in C57BL/6 mice. *Gynecol Endocrinol.* 2020;36(1):49-54. doi 10.1080/09513590.2019.1631274
- Gao Y., Chen F., Kong Q.Q., Ning S.F., Yuan H.J., Lian H.Y., Luo M.J., Tan J.H. Stresses on female mice impair oocyte developmental potential: effects of stress severity and duration on oocytes at the growing follicle stage. *Reprod Sci.* 2016;23(9):1148-1157. doi 10.1177/1933719116630416
- Gavrilov V.V., Onufriev M.V., Moiseeva Y.V., Alexandrov Y.I., Gulyaeva N.V. Chronic social isolation stress and crowding in rats have different effects on learning an operant behavior and the state of the hypothalamo-hypophyseal-adrenocortical system. *Neurosci Behav Physiol.* 2022;52(5):698-704. doi 10.1007/s11055-022-01295-3
- Gilchrist R.B., Ritter L.J., Armstrong D.T. Oocyte-somatic cell interactions during follicle development in mammals. *Anim Reprod Sci.* 2004;82-83:431-446. doi 10.1016/j.anireprosci.2004.05.017
- Godoy L.D., Rossignoli M.T., Delfino-Pereira P., Garcia-Cairasco N., de Lima Umeoka E.H. A comprehensive overview on stress neurobiology: basic concepts and clinical implications. *Front Behav Neurosci.* 2018;12:127. doi 10.3389/fnbeh.2018.00127
- Hammoud R., Tognin S., Bakolis I., Ivanova D., Fitzpatrick N., Burgess L., Smythe M., Gibbons J., Davidson N., Mechelli A. Lonely in a crowd: investigating the association between overcrowding and loneliness using smartphone technologies. *Sci Rep.* 2021;11(1):24134. doi 10.1038/s41598-021-03398-2
- Hawkey L.C., Cacioppo J.T. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. *Ann Behav Med.* 2010;40(2):218-227. doi 10.1007/s12160-010-9210-8
- Hu Y., Wang W., Ma W., Wang W., Ren W., Wang S., Fu F., Li Y. Impact of psychological stress on ovarian function: insights, mechanisms and intervention strategies (Review). *Int J Mol Med.* 2025;55(2):34. doi 10.3892/ijmm.2024.5475
- Huang J., Zeng H. The influence of environmental factors on ovarian function, follicular genesis, and oocyte quality. *Adv Exp Med Biol.* 2021;1300:41-62. doi 10.1007/978-981-33-4187-6_3
- Igonina T., Lebedeva D., Tsybko A., Rozhkova I., Babochkina T., Levinson A., Amstislavsky S. Chronic psychosocial stress affects insulin-like growth factor 1 and its receptors in mouse ovaries. *Reprod Fertil Dev.* 2024a;36:RD24101. doi 10.1071/RD24101
- Igonina T.N., Lebedeva D.A., Shavshaeva N.A., Brusentsev E.Yu., Levinson A., Amstislavsky S. Effects of *in vivo* administration of insulin-like growth factor-1 on oocyte and embryo development in mice under chronic psychosocial stress. *J Evol Biochem Phys.* 2024b;60:1725-1740. doi 10.1134/S0022093024050065
- James K.A., Stromin J.I., Steenkamp N., Combrinck M.I. Understanding the relationships between physiological and psychosocial stress, cortisol and cognition. *Front Endocrinol.* 2023;14:1085950. doi 10.3389/fendo.2023.1085950
- Jeon H., Choi Y., Brännström M., Akin J.W., Curry T.E., Jo M. Cortisol/glucocorticoid receptor: a critical mediator of the ovulatory process and luteinization in human periovulatory follicles. *Hum Reprod.* 2023;38(4):671-685. doi 10.1093/humrep/dead017
- Jiang Y., Xu J., Tao C., Lin X., Li K., Liu Q., ... Sun Y., Zhang F., Kang Y., Xu C., Zhang L. Chronic stress induces meiotic arrest failure and ovarian reserve decline via the cAMP signaling pathway. *Front Endocrinol (Lausanne).* 2023;14:1177061. doi 10.3389/fendo.2023.1177061
- Joseph D.N., Whirledge S. Stress and the HPA axis: balancing homeostasis and fertility. *Int J Mol Sci.* 2017;18(10):2224. doi 10.3390/ijms18102224
- Kala M., Nivsarkar M. Role of cortisol and superoxide dismutase in psychological stress induced anovulation. *Gen Comp Endocrinol.* 2016;225:117-124. doi 10.1016/j.ygcen.2015.09.010
- Kalantaridou S.N., Makriganakis A., Zoumakis E., Chrousos G.P. Reproductive functions of corticotropin-releasing hormone. Research and potential clinical utility of antalarmins (CRH receptor type 1 antagonists). *Am J Reprod Immunol.* 2004;51(4):269-274. doi 10.1111/j.1600-0897.2004.00155.x
- Kalisch R., Russo S.J., Müller M.B. Neurobiology and systems biology of stress resilience. *Physiol Rev.* 2024;104(3):1205-1263. doi 10.1152/physrev.00042.2023

- Kim J., You S. High housing density-induced chronic stress diminishes ovarian reserve via granulosa cell apoptosis by angiotensin II over-expression in mice. *Int J Mol Sci.* 2022;23(15):8614. doi 10.3390/ijms23158614
- Kojima M.L., Hoppe C., Giraldez A.J. The maternal-to-zygotic transition: reprogramming of the cytoplasm and nucleus. *Nat Rev Genet.* 2025;26(4):245-267. doi 10.1038/s41576-024-00792-0
- Koumparou M., Bakas P., Pantos K., Economou M., Chrousos G. Stress management and In Vitro Fertilization (IVF): a pilot randomized controlled trial. *Psychiatriki.* 2021;32(4):290-299. doi 10.22365/jpsych.2021.029
- Lebedeva D.A., Igonina T.N., Brusentsev E.Yu., Shavshaeva N.A., Amstislavskij S.Ya. Effects of ovarian stimulation with gonadotropins in the conditions of chronic psychosocial stress on the quality of murine oocyte. *Russian Journal of Physiology.* 2024;110(6):930-944. doi 10.31857/S0869813924060044 (in Russian)
- Lederbogen F., Kirsch P., Haddad L., Streif F., Tost H., Schuch P., Wüst S., Pruessner J.C., Rietschel M., Deuschle M., Meyer-Lindenberg A. City living and urban upbringing affect neural social stress processing in humans. *Nature.* 2011;474(7352):498-501. doi 10.1038/nature10190
- Levinson A.L., Igonina T.N., Rozhkova I.N., Brusentsev E.Yu., Amstislavskij S.Ya. Psycho-emotional stress, folliculogenesis, and reproductive technologies: clinical and experimental data. *Vavilovskii Zhurnal Genetiki i Selektii = Vavilov J Genet Breed.* 2022;26(5):431-441. doi 10.18699/VJGB-22-53
- Li M., Leatherland J.F., Vijayan M.M., King W.A., Madan P. Glucocorticoid receptor activation following elevated oocyte cortisol content is associated with zygote activation, early embryo cell division, and IGF system gene responses in rainbow trout. *J Endocrinol.* 2012;215(1):137-149. doi 10.1530/JOE-12-0030
- Liu Y.X., Cheng Y.N., Miao Y.L., Wei D.L., Zhao L.H., Luo M.J., Tan J.H. Psychological stress on female mice diminishes the developmental potential of oocytes: a study using the predatory stress model. *PLoS One.* 2012;7(10):e48083. doi 10.1371/journal.pone.0048083
- Lu Q., Cheng Y., Zhou Z., Fan J., Chen J., Yan C., Zeng X., Yang J., Wang X. Effects of emotions on IVF/ICSI outcomes in infertile women: a systematic review and meta-analysis. *J Assist Reprod Genet.* 2025;42(4):1083-1099. doi 10.1007/s10815-025-03388-7
- Ma J., Wang L., Yang D., Luo J., Gao J., Wang J., Guo H., Li J., Wang F., Wu J., Hu R. Chronic stress causes ovarian fibrosis to impair female fertility in mice. *Cell Signal.* 2024;122:111334. doi 10.1016/j.cellsig.2024.111334
- Nishizono H., Uno K., Abe H. Cleavage speed and blastomere number in DBA/2J compared with C57BL/6J mouse embryos. *J Am Assoc Lab Anim Sci.* 2017;56(1):11-17
- Ochnik D., Buława B., Nagel P., Gachowski M., Budziński M. Urbanization, loneliness and mental health model – a cross-sectional network analysis with a representative sample. *Sci Rep.* 2024;14(1):24974. doi 10.1038/s41598-024-76813-z
- Orquiza J.C. Self-stress: a new perspective on stress and moral disorders of civilization. *J Org Psychol.* 2024;24(1). doi 10.33423/jop.v24i1.6885
- Prasad S., Tiwari M., Pandey A.N., Shrivastav T.G., Chaube S.K. Impact of stress on oocyte quality and reproductive outcome. *J Biomed Sci.* 2016;23:36. doi 10.1186/s12929-016-0253-4
- Quinn P., Kerin J.F., Warnes G.M. Improved pregnancy rate in human in vitro fertilization with the use of a medium based on the composition of human tubal fluid. *Fertil Steril.* 1985;44(4):493-498. doi 10.1016/s0015-0282(16)48918-1
- Roth Z. Stress-induced alterations in oocyte transcripts are further expressed in the developing blastocyst. *Mol Reprod Dev.* 2018;85(11):821-835. doi 10.1002/mrd.23045
- Selye H. Stress and the general adaptation syndrome. *Br Med J.* 1950;1(4667):1383-1392. doi 10.1136/bmj.1.4667.1383
- Shindo M., Miyado K., Kang W., Fukami M., Miyado M. Efficient superovulation and egg collection from mice. *Bio Protoc.* 2022;12(11):e4439. doi 10.21769/BioProtoc.4439
- Sun J., Guo Y., Zhang Q., Bu S., Li B., Wang Q., Lai D. Chronic restraint stress disturbs meiotic resumption through APC/C-mediated cyclin B1 excessive degradation in mouse oocytes. *Cell Cycle.* 2018;17(13):1591-1601. doi 10.1080/15384101.2018.1471316
- Sun J., Guo Y., Fan Y., Wang Q., Zhang Q., Lai D. Decreased expression of IDH1 by chronic unpredictable stress suppresses proliferation and accelerates senescence of granulosa cells through ROS activated MAPK signaling pathways. *Free Radic Biol Med.* 2021;169:122-136. doi 10.1016/j.freeradbiomed.2021.04.016
- Suvorov A. Population numbers and reproductive health. *Endocrinology.* 2021;162(11):bqab154. doi 10.1210/endo/bqab154
- Tsuji A., Ikeda Y., Murakami M., Kitagishi Y., Matsuda S. d-Leucine protects oocytes from chronic psychological stress in mice. *Reprod Med Biol.* 2021;20(4):477-484. doi 10.1002/rmb.12396
- Valsamakis G., Chrousos G., Mastorakos G. Stress, female reproduction and pregnancy. *Psychoneuroendocrinology.* 2019;100:48-57. doi 10.1016/j.psyneuen.2018.09.031
- Wang Y., Liu Q., Tang F., Yan L., Qiao J. Epigenetic regulation and risk factors during the development of human gametes and early embryos. *Annu Rev Genomics Hum Genet.* 2019;20:21-40. doi 10.1146/annurev-genom-083118-015143
- Whirledge S., Cidlowski J.A. A role for glucocorticoids in stress-impaired reproduction: beyond the hypothalamus and pituitary. *Endocrinology.* 2013;154(12):4450-4468. doi 10.1210/en.2013-1652
- Wu L.M., Hu M.H., Tong X.H., Han H., Shen N., Jin R.T., Wang W., Zhou G.X., He G.P., Liu Y.S. Chronic unpredictable stress decreases expression of brain-derived neurotrophic factor (BDNF) in mouse ovaries: relationship to oocytes developmental potential. *PLoS One.* 2012a;7(12):e52331. doi 10.1371/journal.pone.0052331
- Wu L.M., Liu Y.S., Tong X.H., Shen N., Jin R.T., Han H., Hu M.H., Wang W., Zhou G.X. Inhibition of follicular development induced by chronic unpredictable stress is associated with growth and differentiation factor 9 and gonadotropin in mice. *Biol Reprod.* 2012b;86(4):121. doi 10.1095/biolreprod.111.093468
- Wu X.F., Yuan H.J., Li H., Gong S., Lin J., Miao Y.L., Wang T.Y., Tan J.H. Restraint stress on female mice diminishes the developmental potential of oocytes: roles of chromatin configuration and histone modification in germinal vesicle stage oocytes. *Biol Reprod.* 2015;92(1):13. doi 10.1095/biolreprod.114.124396
- Yeo C.X., Gilchrist R.B., Lane M. Disruption of bidirectional oocyte-cumulus paracrine signaling during in vitro maturation reduces subsequent mouse oocyte developmental competence. *Biol Reprod.* 2009;80(5):1072-1080. doi 10.1095/biolreprod.108.073908
- Zaidan H., Leshem M., Gaisler-Salomon I. Prereproductive stress to female rats alters corticotropin releasing factor type 1 expression in ova and behavior and brain corticotropin releasing factor type 1 expression in offspring. *Biol Psychiatry.* 2013;74(9):680-687. doi 10.1016/j.biopsych.2013.04.014
- Zanetoullis A.T., Mastorakos G., Vakas P., Vlahos N., Valsamakis G. Effect of stress on each of the stages of the IVF procedure: a systematic review. *Int J Mol Sci.* 2024;25(2):726. doi 10.3390/ijms25020726
- Zhai Q.Y., Wang J.J., Tian Y., Liu X., Song Z. Review of psychological stress on oocyte and early embryonic development in female mice. *Reprod Biol Endocrinol.* 2020;18(1):101. doi 10.1186/s12958-020-00657-1

- Zhang S.Y., Wang J.Z., Li J.J., Wei D.L., Sui H.S., Zhang Z.H., Zhou P., Tan J.H. Maternal restraint stress diminishes the developmental potential of oocytes. *Biol Reprod.* 2011;84(4):672-681. doi [10.1095/biolreprod.110.087890](https://doi.org/10.1095/biolreprod.110.087890)
- Zhang Z., Měchurová K., Resch B., Amegbor P., Sabel C.E. Assessing the association between overcrowding and human physiological stress response in different urban contexts: a case study in Salzburg, Austria. *Int J Health Geogr.* 2023;22(1):15. doi [10.1186/s12942-023-00334-7](https://doi.org/10.1186/s12942-023-00334-7)
- Zhao X.Y., Li Z.B., Yuan H.J., Han X., Wu J.S., Feng X.Y., Zhang M., Tan J.H. Restraint stress and elevation of corticotrophin-releasing hormone in female mice impair oocyte competence through activation of the tumour necrosis factor α (TNF- α) system. *Reprod Fertil Dev.* 2020;32(9):862-872. doi [10.1071/RD20002](https://doi.org/10.1071/RD20002)
- Zhou F.J., Cai Y.N., Dong Y.Z. Stress increases the risk of pregnancy failure in couples undergoing IVF. *Stress.* 2019;22(4):414-420. doi [10.1080/10253890.2019.1584181](https://doi.org/10.1080/10253890.2019.1584181)

Conflict of interest. The authors declare no explicit or potential conflicts of interest related to the publication of this article.

Received September 26, 2025. Revised February 4, 2026. Accepted February 5, 2026.

doi 10.18699/vjgb-26-59

Issues of interspecific hybridization between *Solanum lycopersicum* L. and *Solanum sisymbriifolium* Lam.

A.V. Vishnyakova , A.P. Fedotov , E.V. Osminina , A.Z. Martirosyan , A.D. Kobyashova ,
A.A. Alexandrova , A.A. Mironov , S.G. Monakhos 

¹ Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, Russia

² Lomonosov Moscow State University, Moscow, Russia

³ Skolkovo Institute of Science and Technology, Moscow, Russia

 a.vishnyakova@rgau-msha.ru

Abstract. Interspecific hybridization plays a crucial role in tomato (*Solanum lycopersicum* L.) breeding for introducing beneficial traits from wild relatives, such as resistance to biotic and abiotic stress. Sticky nightshade (*Solanum sisymbriifolium* Lam.) is a promising source for the introgression of desirable traits. Reports on hybridization between *S. lycopersicum* and *S. sisymbriifolium* remain contradictory, differently describing the progeny as doubled haploids or interspecific hybrids. In the present study, we clarify the nature of the progeny derived from hybridization between *S. lycopersicum* and *S. sisymbriifolium* by analyzing the early stages of ovule development and characterizing the morphological and cytogenetic features of the resulting plants. Interspecific hybridization between *S. lycopersicum* and *S. sisymbriifolium* encounters postzygotic reproductive barriers. Only a small proportion of developing ovules in male-sterile tomato lines, whether self-pollinated or pollinated with *S. sisymbriifolium* followed the normal developmental pathway. In most cases, either ovule development was arrested, or parthenocarpic seed-like bodies formed instead of normal seeds. Embryo rescue enabled the recovery of 12 plants resulting from interspecific hybridization of *S. lycopersicum* and *S. sisymbriifolium*. Morphologically, the plants closely resembled the *S. lycopersicum* parent, although they displayed several traits distinctive from those of the parental tomato lines. Notably, yellow-orange mature fruits developed in progeny from the cross of green-fruited *S. lycopersicum* and red-fruited *S. sisymbriifolium*. Analysis of chromosome numbers in root meristems revealed mixoploidy ($2n = 16-26$), and meiotic analysis during microsporogenesis showed multiple aberrations in meiosis. Thus, comprehensive embryological, morphological, and cytogenetic analyses provide evidence for the hybrid origin of the obtained plants. This confirms the possibility of overcoming postzygotic barriers between these species and opens avenues for the utilization of *S. sisymbriifolium* in tomato breeding programs.

Key words: tomato; sticky nightshade; *Solanum lycopersicum*; *Solanum sisymbriifolium*; interspecific hybridization; embryo rescue; microsporogenesis; embryology; mixoploidy

For citation: Vishnyakova A.V., Fedotov A.P., Osminina E.V., Martirosyan A.Z., Kobyashova A.D., Alexandrova A.A., Mironov A.A., Monakhos S.G. Issues of interspecific hybridization between *Solanum lycopersicum* L. and *Solanum sisymbriifolium* Lam. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2026;30(4):569-580. doi 10.18699/vjgb-26-59

Funding. The work was carried out as part of the research project commissioned by the Ministry of Agriculture of the Russian Federation and funded by the federal budget in 2025.

Некоторые аспекты межвидовой гибридизации *Solanum lycopersicum* L. и *Solanum sisymbriifolium* Lam.

А.В. Вишнякова , А.П. Федотов , Е.В. Осминина , А.З. Мартиросян , А.Д. Кобяшова ,
А.А. Александрова , А.А. Миронов , С.Г. Монахос 

¹ Российский государственный аграрный университет – МСХА имени К.А. Тимирязева, Москва, Россия

² Московский государственный университет имени М.В. Ломоносова, Москва, Россия

³ Сколковский институт науки и технологий, Москва, Россия

 a.vishnyakova@rgau-msha.ru

Аннотация. Межвидовая гибридизация играет большую роль в селекции томата (*Solanum lycopersicum* L.), позволяет обогащать генофонд новыми признаками и создавать сорта и гибриды с устойчивостью к биотическим и абиотическим факторам. Перспективным источником для интрогрессии ценных признаков может быть паслен гулявниколистый (*Solanum sisymbriifolium* Lam.). Сведения о результатах гибридизации между *S. lycopersicum* и *S. sisymbriifolium* противоречивы: одни исследователи описывают полученные растения как удвоенные

гаплоиды, другие – как типичные межвидовые гибриды. В настоящей работе мы уточнили природу потомства от гибридизации *S. lycopersicum* и *S. sisymbriifolium* путем анализа начальных стадий развития семязачатков и описания морфологических и цитогенетических особенностей полученных растений. Получение межвидовых гибридов между *S. lycopersicum* и *S. sisymbriifolium* сопряжено с постзиготическими барьерами нескрещиваемости. Сравнительный анализ развивающихся семязачатков от самоопыления линий томата с функциональной мужской стерильностью и от их опыления *S. sisymbriifolium* выявил, что лишь часть семязачатков развивается по типичному пути, в то время как большинство либо останавливается в развитии, либо формирует партенокарпические структуры с псевдоэмбриональной тканью. С использованием эмбриокультуры нами было получено 12 растений от гибридизации *S. lycopersicum* и *S. sisymbriifolium*. По морфологическим признакам все растения были близки *S. lycopersicum*, однако отмечено появление признаков, не характерных для родительских линий томата, например желто-оранжевой окраски зрелых плодов у растений, полученных при скрещивании линии зеленоплодных томатов с красноплодным пасленом гулявниколистным. При анализе числа хромосом в меристемах корня наблюдали миксоплоидию ($2n = 16-26$), а при микроспорогенезе регистрировали множественные нарушения мейоза. Таким образом, комплексный эмбриологический, морфологический и цитогенетический анализ предоставляет доказательства гибридного происхождения полученных растений, что подтверждает возможность преодоления постзиготических барьеров между данными видами и открывает путь к использованию *S. sisymbriifolium* в селекции томата.

Ключевые слова: томат; паслен гулявниколистный; *Solanum lycopersicum*; *Solanum sisymbriifolium*; межвидовая гибридизация; эмбриокультура; микроспорогенез; эмбриология; миксоплоидия

Introduction

Tomato (*Solanum lycopersicum* L.) is one of the world's most economically important vegetable crops. However, its productivity is significantly limited by its susceptibility to biotic and abiotic stresses (Sadashiva et al., 2017). Thus, breeding strategies that develop genotypes with genetic resistance are crucial (Abbas et al., 2024). One of the key approaches to solving this problem is interspecific hybridization, which enables the introduction of valuable alleles from related wild species into the cultivated tomato genome (Ajaharuddin et al., 2024).

The introgression of disease resistance genes from other species into tomatoes has been employed for long and is quite successful (Rubio et al., 2016; Yerasu et al., 2023). Currently, most of the sources of resistance used in tomato breeding are wild species of the *Lycopersicon* section. Species from other sections of the *Solanum* genus are in lesser use (Foolad, 2007), as their gene pool availability is restricted by incompatibility barriers (Abbas et al., 2024).

Sticky nightshade (*Solanum sisymbriifolium* Lam.), a distant relative of the tomato, is regarded by some researchers as a promising source of disease resistance. It has been found to be resistant to bacterial wilt (Collonnier et al., 2003) and verticillium wilt (Piosik et al., 2019). According to W.G. Flier et al. (2003), *S. sisymbriifolium* exhibits hypersensitivity when inoculated with several European isolates of *Phytophthora infestans*, a potato pathogen. In field tests in central Russia, *S. sisymbriifolium* demonstrated high resistance to tomato late blight, according to our observations. Additionally, researchers have reported resistance to carmine spider mites (Piosik et al., 2019) and nematodes (Hajihassani et al., 2020). *S. sisymbriifolium* has a high adaptive potential and is currently widespread in many countries (Biswas et al., 2023).

Hybridization between distantly related species, such as tomato and sticky nightshade, encounters serious post-zygotic barriers. However, it has been demonstrated that sticky night-

shade pollen tubes successfully grow within tomato pistils and germinate in ovules (Bal, Abak, 2003; Piosik et al., 2019). Interspecific hybridization between *S. lycopersicum* and *S. sisymbriifolium* has been attempted by several research teams. However, the results were inconsistent: D. Chambonet (1996) and U. Bal and K. Abak (2003) obtained plants phenotypically similar to tomatoes and considered them to be doubled haploids. In contrast, L. Piosik et al. (2019) produced interspecific hybrids that resembled nightshades. Thus, the nature of plants obtained through interspecific hybridization, as well as their morphological and cytological characteristics, remains obscure.

The goal of this study was to produce plants resulting from the cross between *S. lycopersicum* and *S. sisymbriifolium* and to perform a comprehensive analysis of their characteristics using comparative embryology, cytogenetics, and morphological trait analysis.

Materials and methods

Plant material and hybridization. The tomato samples used for hybridization were obtained from the Timofeev Breeding Station, Moscow. These samples include beef tomato lines Roz Son and Roz st9 and cherry tomato lines st8, st6, st6(lik), and st4. Lines Roz Son, st8, st6, st6(lik), and st4 are generations 6–8 of inbred lines. Roz st9 is the third generation, homozygous for six resistance genes against fusarium wilt, late blight, bronze leaf of tomato, verticillium wilt, nematodes, and tomato mosaic virus. All lines exhibit functional male sterility, self-compatibility, and indeterminate growth. Fruits of Roz Son and Roz st9 are pink; st8, green; and st6(lik), st6, and st4, red. The sticky nightshade sample Sn was also obtained from the Timofeev Breeding Station. These plants are characterized by self-incompatibility, determinate growth, and intensely red fruits measuring 2–3 cm in diameter.

Hybridization of tomato and sticky nightshade was conducted from 9 to 11 a.m. The crosses were reciprocal. For

the control, the tomato parental lines were self-pollinated. Pollination was performed on emasculated buds: for the tomato, at the lemon-yellow bud stage; for the nightshade, at the white bud stage, one or two days before the flowers opened. Immediately after emasculation, pollination was performed with freshly collected pollen, and the stigma was isolated with cotton wool. A total of 1,165 tomato flowers were pollinated with sticky nightshade, not counting the flowers used for histological analysis. The numbers of pollinations performed by cross combination were: Roz Son – 156, Roz st9 – 298, st8 – 185, st6(lik) – 190, st6 – 132, and st4 – 204. At least five plants of each tomato sample and six plants of nightshade were involved in the hybridization.

Histology. Tomato flowers and developing fruits were fixed in a formaldehyde–alcohol–acetic acid (FAA) solution (Ruzin, 1999). At subsequent steps, a minimum of two samples (flower or developing fruit) were gathered from each plant line up to 15 days after pollination (DAP). They were analyzed either whole or after longitudinal dissection. Individual ovules (at least 10 from each plant line) or ovules attached to the placenta were isolated from samples taken after 15 DAP and used for further study. Samples were dehydrated through a graded ethanol series, cleared in xylene, and embedded in paraffin. Serial sections 10–15 μm thick were obtained with a Microm HM355S rotary microtome (Thermo Fisher Scientific). Sections were stained with 0.1 % Alcian Blue in 3 % acetic acid and 1 % aqueous Safranin (Barykina et al., 2004) or stained with PAS reaction followed by Ehrlich's hematoxylin staining (Ruzin, 1999). Staining was carried out manually or using a Varistain Gemini ES automated stainer (Thermo Fisher Scientific). Slides were examined under an AxioPlan 2 light microscope (Carl Zeiss) equipped with a differential interference contrast (DIC) module, or with a BX61VS slide scanner (Olympus). Image processing was performed using Adobe Photoshop and ImageJ software.

Embryo rescue. Fruits were selected for cultivation on nutrient media on 16 to 36 days after pollination. Fruit set was counted as the fruits were harvested. The fruits were sterilized in a 2 % sodium hypochlorite solution and then washed three times in sterile distilled water. Enlarged ovules were isolated and transferred to a nutrient medium. We used for ovule culture the nutrient media described by D.J. Harberd (1969) and Ł. Piosik et al. (2019). If the ovule integument had no ruptures and the embryos did not germinate on the solidified nutrient medium, the embryoids were removed from the enlarged ovules after three weeks and cultivated on a medium with the same composition. To regenerate shoots from callus tissue, a solidified Murashige–Skoog (MS) medium containing 30 g/L sucrose, 5 mg/L BAP, and 0.01 mg/L IAA was used. If the ovules germinated or formed organogenic callus with shoots, they were transferred to an MS medium with agar, containing 20 g/L sucrose and 0.01 mg/L each of BAP and IAA. In cases of difficulty in shoot rooting, a MS nutrient medium containing 20 g/L sucrose, 2 mg/L NAA, and 8 g/L agar was used.

Phenotyping of the obtained plants. The st8×Sn (No. 1, 2, and 3) and st6×Sn plants (No. 1, 2, 3, 4, 5, and 6), as well as the parental lines st6, st8, and sticky nightshade Sn, were examined with regard to the morphological characteristics of their shoots, flowers and fruits, in accordance with the DUS testing guidelines of the State Commission of the Russian Federation for Selection Achievements Test and Protection (<https://gossortrf.ru/publication/#publication-methodics>).

Cytological studies of the chromosome numbers of the obtained plants in mitotic and meiotic cells. Cytological preparations were made from anther and root meristematic cells using the SteamDrop method (Kirov et al., 2014). The prepared slides were examined under an AXIOSKOP 40 microscope (Carl Zeiss) at 1000× magnification using immersion oil. At least 10 cells per sample were examined.

Statistical analysis. Chi-square analysis was used to compare fruit set following the pollination of tomato lines by *S. sisymbriifolium*. Pearson's chi-square test was used to assess the significance of differences between genotypes at a significance level of $p = 0.05$. For pairwise comparisons between genotypes, the Bonferroni correction was applied to adjust the critical significance level. Data analysis and graphical visualization of the results were performed using the Microsoft Excel and R Studio (version 2025.09.2+418) software packages.

Results

Typical pattern of embryogenesis following self-pollination in *S. lycopersicum* line Roz st9

In order to compare the early stages of embryogenesis, *S. lycopersicum* line Roz st9 was chosen as a control line. Only mature unitegmatic tenuinucellate ovules with completely developed embryo sacs typical of tomatoes were observed in the ovary of control plants one day before anthesis (Fig. 1A). The ovules exhibited a multicellular endosperm and a linear proembryo by day 5 after self-pollination (Fig. 1B). Simultaneously, the integumentary tapetum (IT), also referred to as endothelium, originated from the inner epidermis of the integument, beginning its development on the chalazal pole. The parenchyma cells of the integument adjacent to the IT developed thickened cell walls, and their contents began to degenerate, forming a “death zone” (DZ).

Further seed development was accompanied by an increase in size due to both the proliferation of cells at the periphery of the developing seed coat and the continuous expansion of the embryo and endosperm (Fig. 1C–E). The embryo reached the early globular stage by 10–15 DAP (Fig. 1C). The accumulation of storage sugars began in the endosperm cells, which were tightly attached to the embryo. The outer part of the endosperm was fully covered by a single-layered IT. Spherical embryos with a well-developed suspensor at the late globular stage (Fig. 1D) and embryos at the heart stage (Fig. 1E) were observed in developing seeds at 24 DAP. The endosperm cells

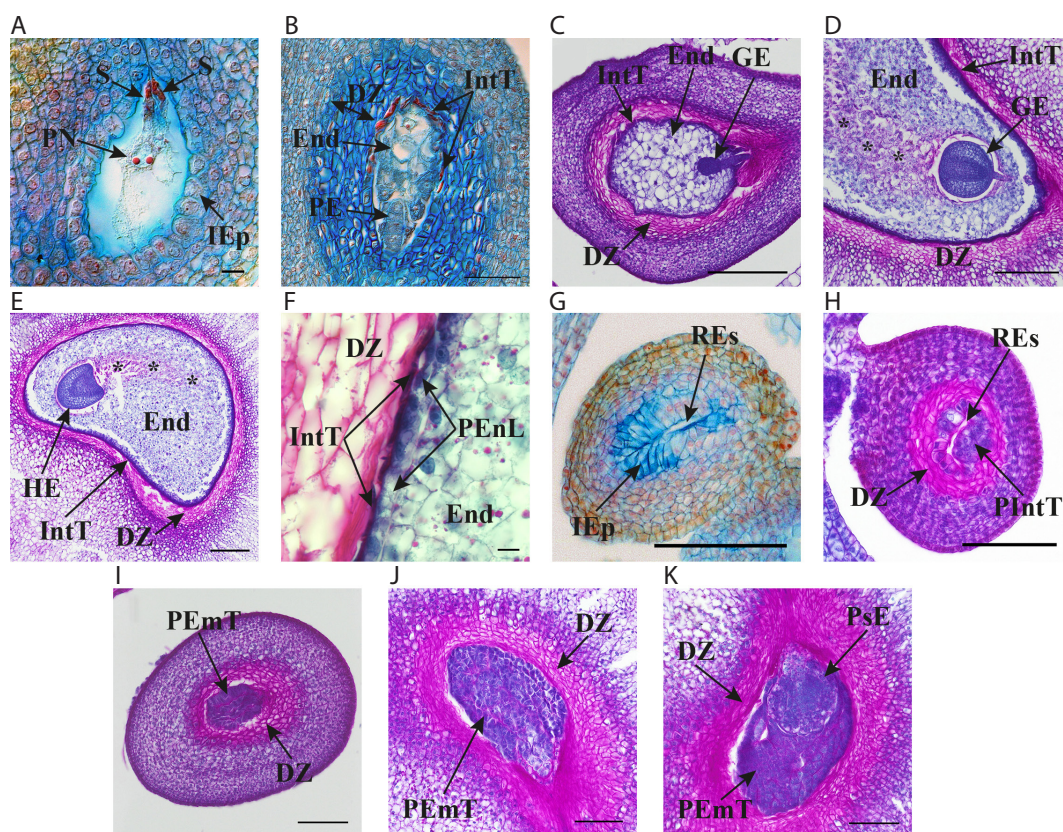


Fig. 1. Typical seed development following self-pollination in *S. lycopersicum* line Roz st9 (A–F) and observed aberrations leading to the formation of seed-like bodies (G–K). A – structure of the embryo sac and surrounding tissues one day before anthesis (0 DAP); B – endosperm and proembryo in the developing seed, 5 DAP; C – early globular embryo, 15 DAP; D – late globular embryo, 24 DAP; E – developing embryo at the heart stage, 24 DAP; F – endosperm contact zone with surrounding tissues of the developing seed coat, 24 DAP; G – cells of the inner epidermis of the integument grow into the collapsed embryo sac, 5 DAP; H – development of the death zone around the endothelium, 10 DAP; I – pseudoembryonic tissue development, 15 DAP; J – seed-like body filled with pseudoembryonic tissue, 24 DAP; K – pseudoembryo surrounded by pseudoembryonic tissue, 24 DAP.

DZ – death zone; End – endosperm; GE – globular embryo; HE – heart-shaped embryo; IEp – inner epidermis of the integument; IntT – integumentary tapetum; PE – proembryo; PEnL – peripheral endosperm cell layer; PN – polar nuclei; PsE – pseudoembryo; S – synergid; PEmT – pseudoembryonic tissue; PIntT – proliferating integumentary tapetum; REs – remnants of the degenerated embryo sac. Asterisks indicate the region of endosperm cells being degraded by the embryo. Scale bars: A, F – 10 μ m, B – 50 μ m, G, K – 100 μ m, C–E – 200 μ m.

adjacent to the IT differed from neighboring cells in having smaller sizes, flattening along the IT, and noticeable vacuolization; thus, they can be considered a separate peripheral layer of endosperm cells (Fig. 1F). The embryos were surrounded by a well-developed endosperm cavity, whereas a number of endosperm cells underwent degradation due to nutrient uptake by the embryo (Fig. 1D, E). At 28 DAP, embryos at the torpedo stage were detected in developing seeds.

Deviations from the typical embryogenic pattern after self-pollination in *S. lycopersicum* line Roz st9

Embryo sac degeneration was the most common deviation in ovule development after self-pollination of tomato lines. Such ovules were characterized by the destruction and shrinkage of the embryo sac by 5 DAP (Fig. 1G). Ovule development ceased when the inner epidermis cells of the integument no-

ticeably elongated and filled the space left by the collapsed embryo sac.

In some cases, the cells of the inner epidermis of the integument did not elongate following embryo sac collapse at 5–10 DAP but instead differentiated into IT and started to proliferate, and this process was accompanied by the formation of DZ around them (Fig. 1H). Such ovules continued to grow and produce seed-like bodies, which were smaller than ordinary seeds. The proliferating IT formed a pseudoembryonic tissue, distinguished by the formation of cells clusters, thicker cell walls at the cluster borders, lack of storage sugars during the subsequent ovule development, intense staining characteristic of IT (Fig. 1I, J), and direct contact with the DZ. One sample collected at 24 DAP contained a body resembling an embryo at the late globular stage at the micropylar side surrounded by pseudoembryonic tissue (Fig. 1K). Since its nature is still un-

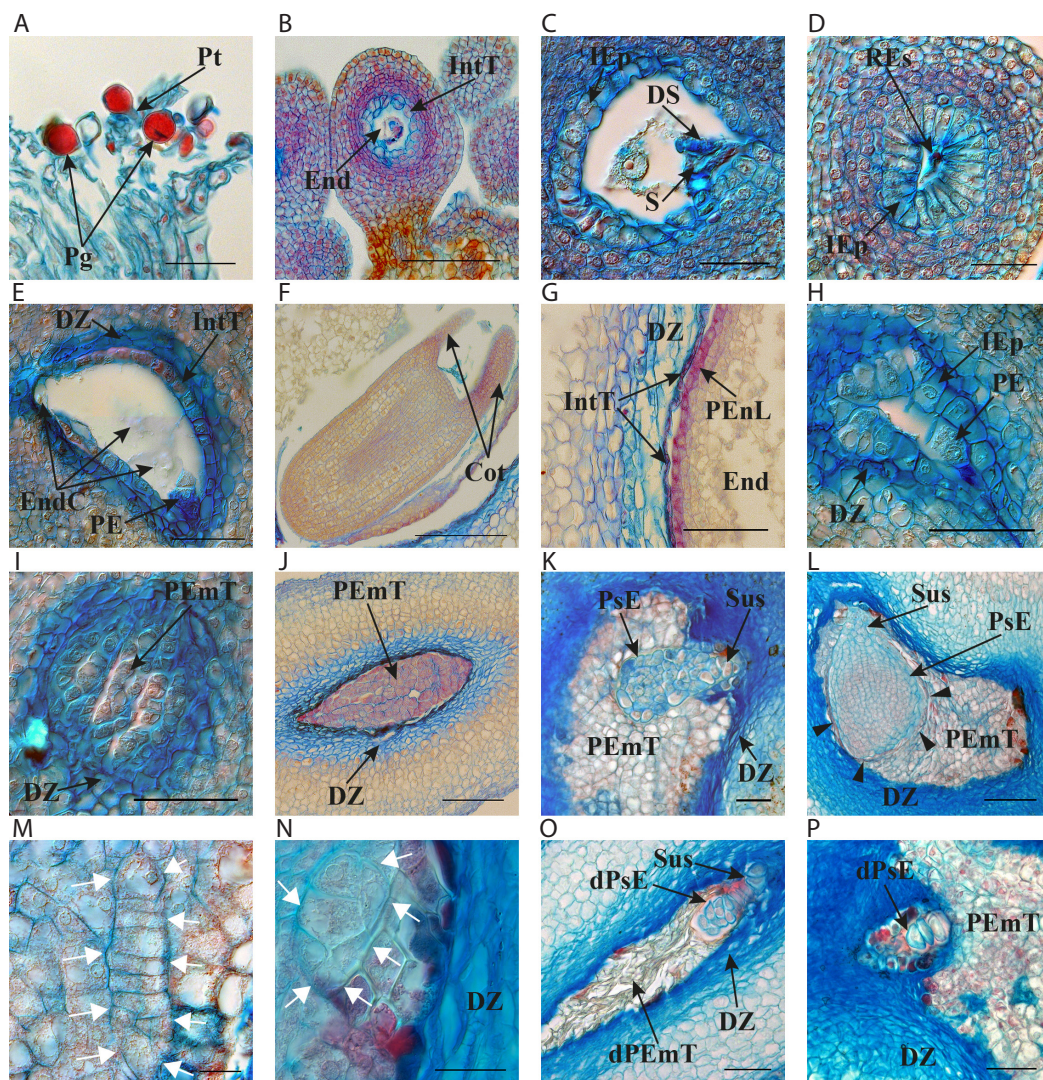


Fig. 2. Development of seeds and seed-like bodies after pollination of *S. lycopersicum* with *S. sisymbriifolium* pollen. A – germinated pollen grains on the receptive surface of the stigma, 3 DAP; B – early endosperm cell divisions, 3 DAP; C – ovule with degenerated synergid, 3 DAP; D – cells of the inner epidermis of the integument grow into the collapsed embryo sac, 3 DAP; E – proembryo surrounded by cellular endosperm (nuclei out of focus), 15 DAP; F – embryo at the torpedo-shaped stage, 15 DAP; G – endosperm contact zone with surrounding tissues of the developing seed coat, 15 DAP; H – cells of the inner epidermis of the integument grow into the collapsed embryo sac with the presence of the proembryo, 15 DAP; I – seed-like body with pseudoembryonic tissue surrounded by DZ, 15 DAP; J – seed-like body with proliferated pseudoembryonic tissue and developed DZ, 15 DAP; K – pseudoembryo at the globular stage of development, 28 DAP; L – large pseudoembryo at the globular stage of development, 30 DAP; M – cell cluster in the pseudoembryonic tissue, 28 DAP; N – contact zone between pseudoembryonic tissue and DZ, 30 DAP; O – degenerating pseudoembryonic tissue and pseudoembryo, 29 DAP; P – degenerating pseudoembryo with intact pseudoembryonic tissue, 30 DAP.

Cot – cotyledons; dPsE – degenerating pseudoembryo; dPEmT – degenerating pseudoembryonic tissue; DS – degenerated synergid; DZ – death zone; End – endosperm; EndC – endosperm cells; IEp – inner epidermis of the integument; IntT – integumentary tapetum; PEmT – pseudoembryonic tissue; PsE – pseudoembryo; REs – remnants of the degenerated embryo sac; S – synergid; Sus – suspensor. White arrows indicate the boundaries of the cell cluster within the pseudoembryonic tissue. Black triangles indicate elongated cells of the pseudoembryonic tissue at the embryo boundary. Scale bars: M, N – 25 µm, C, D – 30 µm, A, E, H, I, K, O, P – 50 µm, B, G, J, L – 100 µm, F – 200 µm.

known, we hereafter refer to such bodies as “pseudoembryos”. It was characterized by a large, broad suspensor that contacted the DZ and by the absence of visible cellular differentiation, including protoderm cells. The cells of the pseudoembryonic tissue were closely attached to the pseudoembryo.

Pattern of embryogenesis in crosses between *S. lycopersicum* and *S. sisymbriifolium*

Germinating pollen grains were observed in all analyzed cross combinations after hybridization between *S. lycopersicum* and *S. sisymbriifolium* (Fig. 2A). Only mature ovules with fully

developed embryo sacs were observed in all examined ovaries at 1 DAP. Although no growth of pollen tubes into the ovules was observed, at 3 DAP ovules with one or two degenerated synergids were found, and initial divisions of endosperm cells were detected (Fig. 2B, C). The onset of IT differentiation was also noted. In some ovules, the embryo sac degenerated and cells of the inner epidermis of the integument elongated (Fig. 2D). Such ovules were arrested in their development and were present in samples at all subsequent developmental stages.

Samples collected at 10–15 DAP revealed significant variation in the observed stages of embryogenesis depending on the line examined. Seeds following the typical course of embryogenesis for tomatoes were observed in the Roz st9 and Roz Son lines. In developing seeds of the Roz st9 line, a linear proembryo was present together with an endosperm consisting of a small number of cells, which was surrounded by IT and a forming DZ (Fig. 2E). In contrast, in the Roz Son line at 15 DAP, embryos at the torpedo stage surrounded by an endosperm cavity (Fig. 2F), a well-developed endosperm with a differentiated layer of peripheral cells surrounded by IT, and a well-developed DZ were detected (Fig. 2G). In control plants, the corresponding stages of seed development in the Roz st9 and Roz Son lines were observed at 5 and 28 DAP, respectively. In ovary samples of the st6 line at 15 DAP, ovules without discernible changes or with a degenerated synergid were observed.

In all lines examined, most ovules at 15 DAP were arrested in development: their embryo sac degenerated to be replaced by elongated cells of the inner epidermis of the integument. It was noted in some ovules of the Roz st9 line that although the cells of the inner epidermis of the integument began to elongate, cells not belonging to the integument were present on the micropylar side, and the DZ began to differentiate (Fig. 2H). Some seed-like bodies in the Roz st9 line (at an earlier developmental stage) and in the Roz Son line (at a later developmental stage) were occupied by pseudoembryonic tissue, as in the control plants (Fig. 2I, J).

In all lines, only two types of seed-like bodies were found in the examined samples at 20–30 DAP: those filled with pseudoembryonic tissue and those containing pseudoembryonic tissue together with an embryo on the micropylar side (Fig. 2K, L). No seeds following the developmental pathway typical of tomato were detected. The pseudoembryonic tissues in both types of seed-like bodies shared some features (Fig. 2M, N): formation of clusters of cells with thicker cell walls at the boundaries of the clusters, absence of storage substances, and direct contact with the DZ without any specialized cells. Some pseudoembryonic tissue cells, most often at the periphery, had granular cytoplasm. They were highly stained with Safranin (Fig. 2L, N).

The drop-shaped pseudoembryo was always located on the micropylar side, and its narrow part, resembling a suspensor, contacted the DZ (Fig. 2K, L). In some pseudoembryos, the suspensor was linear and looked as a single file of cells, whereas in larger pseudoembryos it consisted of several files of

cells that widened acropetally, making it impossible to draw a precise boundary between the pseudoembryo and the suspensor (Fig. 2K, L). The outermost cells were relatively flattened only in the largest pseudoembryos (Fig. 2L), whereas in most cases no visible cellular differentiation, including protoderm cells, was observed (Fig. 2K). Some pseudoembryos were covered by a cuticle (Fig. 2K). The adjacent pseudoembryonic tissue cells were always closely appressed to the pseudoembryo. They showed no signs of degradation. In some seed-like bodies, degeneration of both the pseudoembryo and the pseudoembryonic tissue (Fig. 2O) or only of the pseudoembryo itself (Fig. 2P) was observed.

Fruit set in the hybridization of *S. lycopersicum* and *S. sisymbriifolium*, and the cultivation of isolated ovules and embryos

The average fruit set rate for the beef tomato lines Roz Son and Roz st9 in 2024 was 82.6 ± 2.1 %. For the cherry tomato lines st8, st6(lik), st6 and st4, it was 86.9 ± 6.7 %. No fruit set was observed when nightshade was pollinated with tomato pollen in any of the cross combinations.

The number of developing ovules and fruit set in crosses between *S. lycopersicum* and *S. sisymbriifolium* depended more on the genotype of the maternal plant than on its group membership (Fig. 3). The beef tomato line Roz st9 and the cherry tomato line st6(lik) exhibited high fruit sets. The cherry tomato line st8 exhibited a low fruit set. Fruit set was also influenced by weather conditions of that year; on the average, fruit set was lower in 2023 for all genotypes except st8, though the overall trend was consistent across genotypes.

In 2023, we isolated 123 ovules of the Roz st9×Sn combination, 87 ovules of the Roz Son×Sn combination, 42 ovules of the st6×Sn combination, 72 ovules of the st6(lik)×Sn combination, 45 ovules of the st8×Sn combination, and 88 ovules of the st4×Sn combination. No regenerated plants were obtained from isolated embryos cultured in 2023.

In 2024, we isolated 275 enlarged ovules of the Roz st9×Sn combination, 97 ovules of Roz Son×Sn, 127 ovules of st6×Sn, 304 ovules of st6(lik)×Sn, 117 ovules of st8×Sn, and 178 ovules of st4×Sn. During the cultivation of enlarged ovules on nutrient media, browning or whitening of the ovules with their subsequent death was predominantly observed. An increase in the size of the ovules, rupture of the integument, and germination of the embryo were observed only in the st8×Sn combination in 2 % of the isolated ovules. In the st6×Sn and st6(lik)×Sn combinations, the ovules increased in size without integument rupturing. They were opened manually and the embryos were isolated. Nonorganogenic callus formed from 58 % of the extracted st6×Sn embryos, while organogenic callus formed from 30 %, shoots were obtained from the latter, and the remaining embryos died. In the st6(lik)×Sn combination, organogenic callus formed from only 1 % of the embryos, and plants were obtained from it. As a result, we obtained three plants from the st8×Sn combination, six plants from the st6×Sn combination, two plants from the st6(lik)×Sn combination, and one plant from the st4×Sn combination.

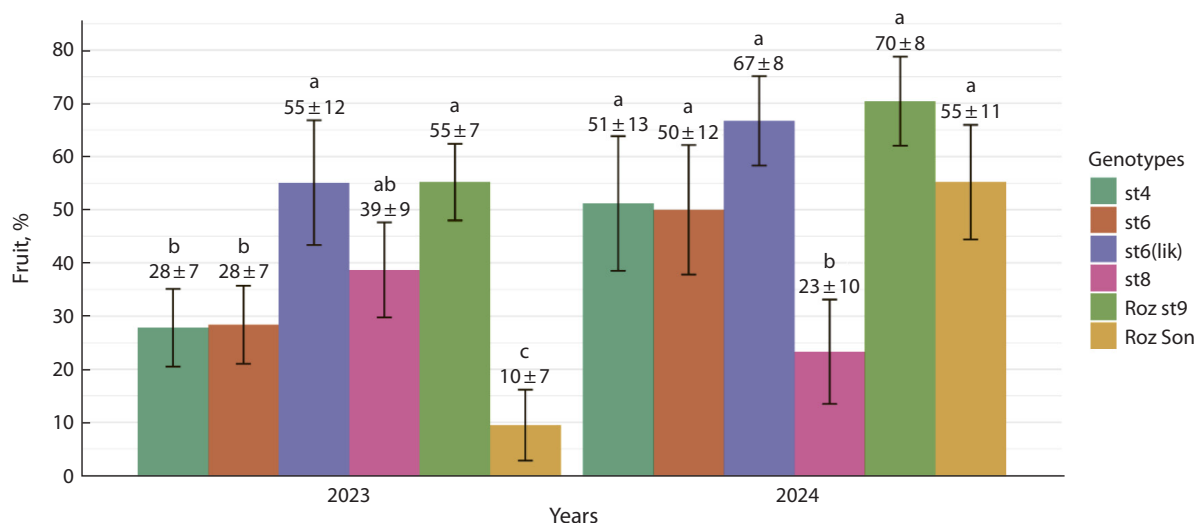


Fig. 3. Fruit set in the hybridization of *S. lycopersicum* and *S. sisymbriifolium*.

Chart bars labeled with the same lowercase letters (a, b, c) do not differ significantly at the 5 % significance level ($P \leq 0.05$).

In the *in vitro* combinations involving beef tomatoes, either embryo death or callus formation was observed. Plant regeneration from the callus was not successful. Callus was obtained from 25 % of the embryos introduced into culture in the Roz Son×Sn combination and from 58 % in the Roz st9×Sn combination.

Morphological traits of the plants obtained

A detailed analysis of the progeny resulting from the *S. lycopersicum* × *S. sisymbriifolium* hybridization and of the parental forms was conducted to identify phenotypic differences from the parental forms. Forty traits were taken into consideration. In what follows, we dwell on the plants derived from the st8×Sn and st6×Sn combinations (Supplementary Material)¹.

The plants resulting from the *S. lycopersicum* × *S. sisymbriifolium* hybridization had an indeterminate growth habit, though the intensity of their growth processes varied. At the time of measurement, the plants obtained from the st8×Sn cross ranged in height from 152 to 186 cm, whereas the st8 line measured 230 cm and the nightshade measured 160 cm. A similar trend was observed in the analysis of the st6×Sn plants: the st6 parent reached a height of 230 cm, whereas the progeny ranged from 150 to 210 cm in height.

The leaf position relative to the stem in the plants obtained from the st6×Sn and st8×Sn hybrid combinations was semi-erect, as in the tomato parental lines; however, in plants No. 3, 4, and 6 of the st6×Sn combination, a semi-vertical leaf position was observed, which was not characteristic of either parental form. Based on leaf blade characters, both intermediate expressions of traits and the emergence of new traits were identified (Table S1).

Analysis of the morphological traits of fruits in plants obtained from the st8×Sn hybrid combination revealed intermedi-

ate fruit coloration between the parental forms (Table S1). The st8 tomato had yellow-green mature fruit, while the *S. sisymbriifolium* was red. All plants obtained from this combination exhibited yellow-orange coloration; the pulp also acquired an uneven yellow-orange tint (Fig. 4), with plants No. 1 and 3 showing more intense coloration than plant No. 2.

Analysis of fruit and pulp coloration in the plants obtained from the st6×Sn combination revealed no segregation: all fruits were orange-red, matching the st6 parental tomato line. However, plant st6×Sn No. 1 exhibited green striping on the fruit prior to ripening, a trait not observed in either parental form. The fruit shape in the longitudinal section demonstrated segregation: plants st6×Sn No. 3, 4, 5, and 6 displayed a heart-shaped longitudinal section similar to the st6 line, whereas plant st6×Sn No. 2 was round, like *S. sisymbriifolium*. The longitudinal section of st6×Sn No. 1 differed from both parents and was oblate.

Chromosome number analysis in meristems and in microsporogenesis

Analysis of the chromosome number in root meristems revealed mixoploidy in all plants obtained from the cross between tomato and *S. sisymbriifolium*. The chromosome number varied from 16 to 26. The proportion of cells with the expected chromosome number of $2n = 24$ varied significantly among the samples (Fig. 5A).

Of the st6×Sn combination, plant No. 4 was the most stable in terms of chromosome number in root meristems. Of the examined cells, 60 % had 24 chromosomes and 40 % had 22. In the other plants studied from the st6×Sn and st8×Sn combinations, the percentage of cells with the expected number of chromosomes was substantially lower, not exceeding 40 % (Fig. 5A). In plant st6×Sn No. 1, 45 % of the examined cells had 26 chromosomes; however, cells with 19, 22, and 25 chromosomes were also noted. In plant st6×Sn No. 2, 36 %

¹ Supplementary Tables S1 and S2 are available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Vish_Engl_30_4.pdf

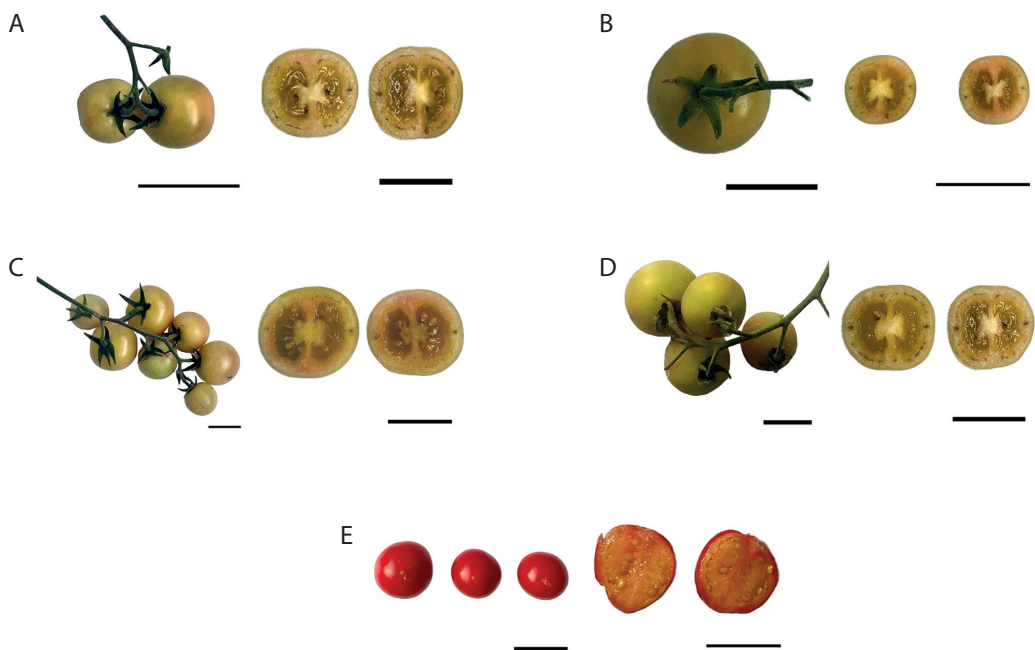


Fig. 4. Fruit and pulp coloration in plants obtained from the st8xSn combination and the parental genotypes: A – st8xSn No. 1; B – st8xSn No. 2; C – st8xSn No. 3; D – tomato line st8; E – *S. sisymbriifolium* Sn.
Scale bars: A – 5 cm, B–E – 2.5 cm.

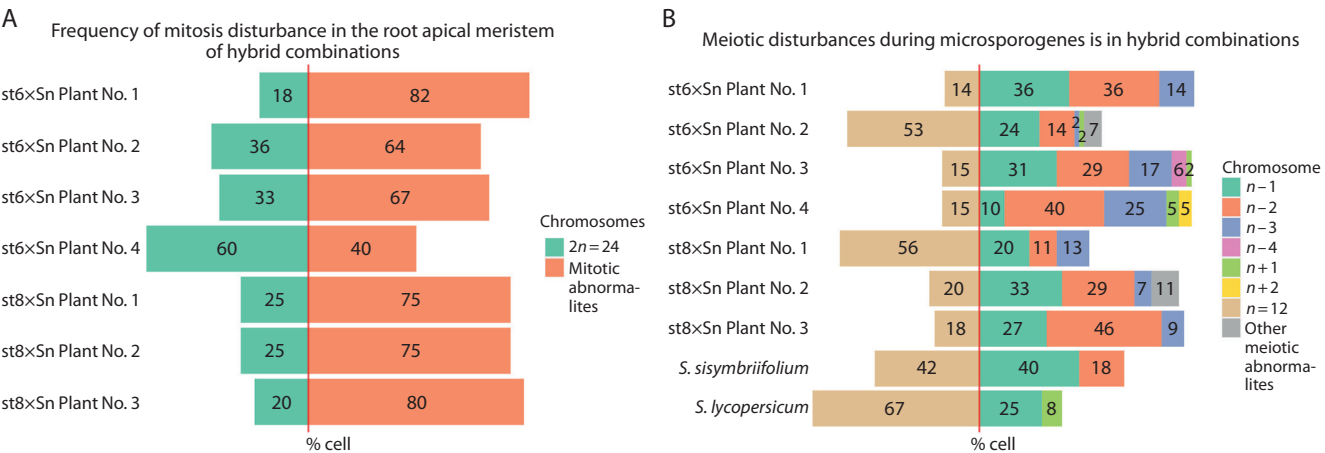


Fig. 5. Distribution of cells with the expected chromosome number and cells with altered chromosome numbers (A) in mitosis in meristems and (B) in meiosis during microsporogenesis.

of the examined cells had 26 chromosomes, while other cells had 16, 20, or 22 chromosomes. In plant st6xSn No. 3, 33 % of cells had 26 chromosomes, 25 % had 22, and a few had 20. Chromosome numbers in cells from the st8xSn plants varied from 20 to 23, and no cells with excess numbers were found.

Analysis of microsporogenesis in the parental tomato forms showed that 33 % of cells had altered chromosome numbers, whereas *S. sisymbriifolium* showed 58 %. In plants obtained from their hybridization, the proportion of cells with altered chromosome numbers and other meiotic aberrations increased to 86 % (Fig. 5B). The number of chromosomes in the meiotic

cells of plants derived from the cross between tomato and *S. sisymbriifolium* varied broadly, from 8 to 14 chromosomes per cell.

The most frequent division irregularities were those leading to the formation of aneuploid microspores in tetrads with chromosome numbers of $n - 1$, $n - 2$ and $n - 3$ (Fig. 5B). Additionally, spontaneous chromosome doubling was observed in plant No. 2 of the st6xSn combination (Fig. 6D). Division irregularities associated with uneven chromosome distribution during telophase II of meiosis were found in plants No. 1 and 2 of the st8xSn combination (Fig. 6E, F) and in st6xSn

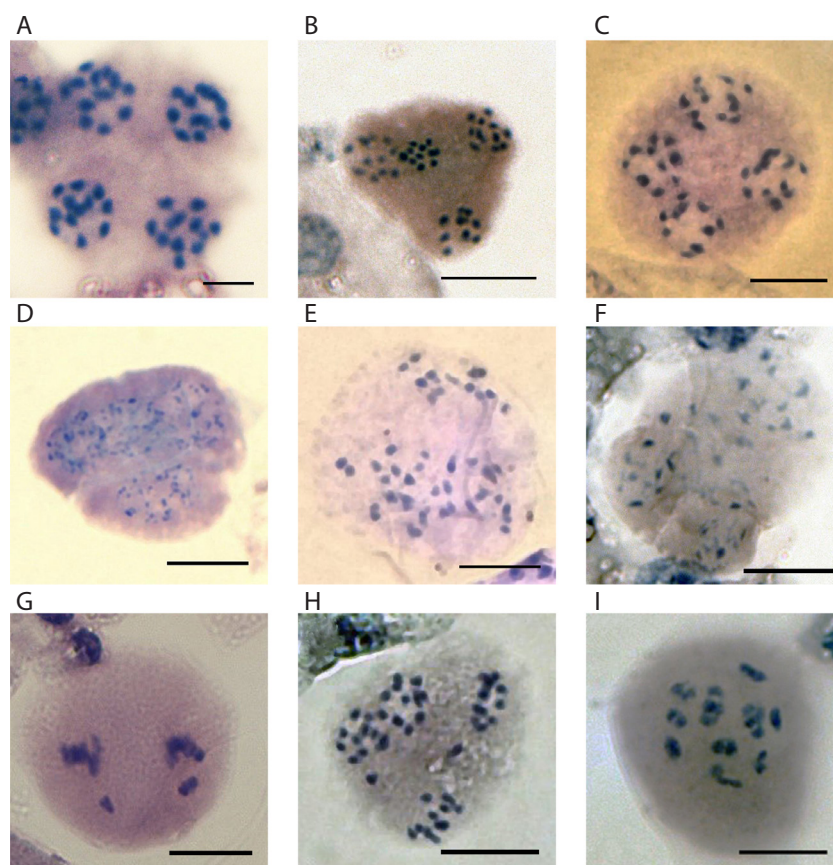


Fig. 6. Fragments of cytological preparations with chromosomes of meiotic cells: A – sticky nightshade (telophase II); B – st6xSn No. 2 (telophase II); C – tomato (telophase II); D – st6xSn No. 2 (telophase II); E, F – st8xSn No. 2 (telophase II); G – st8xSn No. 2 (anaphase I); H – st8xSn No. 3 (telophase II); I – st6xSn No. 3 (prophase I).

Scale bars 10 μm.

No. 4. Meiotic irregularities such as nondisjunction or lagging chromosomes during metaphase and telophase II were noted in plants No. 2 and 3 of the st6xSn cross and plants No. 1–3 of st8xSn (Fig. 6G, H). We also noted the formation of univalents, trivalents, and bivalents with a single chiasma in st6xSn No. 2 and 3, and st8xSn No. 2 (Fig. 6I).

Discussion

Determining the nature of plants obtained through crossing remains one of the primary challenges in any interspecific hybridization study. Previous reports on *S. lycopersicum* × *S. sisymbriifolium* hybrids produced conflicting results. Some researchers classified the progeny as haploids or doubled haploids (Chambonnet, 1996; Bal, Abak, 2003), while others confirmed their status as true interspecific hybrids (Piosik et al., 2019). The plants obtained in this study were phenotypically similar to the tomato parent. This is consistent with the results of U. Bal and K. Abak (2003); however, we hold the view that these plants are of hybrid origin.

Only mature embryo sacs (Fig. 1A) typical of tomatoes (Cooper, 1931) were recognized in both the control line and all examined plants from lines pollinated with *S. sisym-*

briifolium pollen one day before anthesis. Accordingly, all aspects of ovule development observed in this study are linked to fertilization or post-zygotic stages of development. In the analyzed *S. lycopersicum* lines, four variants of ovule development were observed following either self-pollination or hybridization with *S. sisymbriifolium*. During the typical developmental course (Fig. 1B–F, 2A–C, 2E–G), normal tomato seeds formed, characterized by a well-developed embryo surrounded by a cellular endosperm, IT, and DZ. In addition, three alternative developmental types were detected: (i) developmental arrest (Fig. 1G; 2D); (ii) formation of pseudoembryonic tissue (Fig. 1H–J; 2G, I); and (iii) formation of pseudoembryonic tissue with a pseudoembryo (Fig. 1K; 2K–P). The last two patterns resulted in the production of seed-like bodies.

The most prevalent type of ovule development in all examined plants was developmental arrest caused by the collapse of the embryo sac and the filling of the resulting gap by elongated inner epidermal cells of the integument (Fig. 1G; 2D) without IT differentiation. This variant of ovule development has been described by several authors, and the origin of the observed elongated cells from the inner epidermis cells of

the integument is unquestionable (Palamarchuk et al., 1975; Chaban et al., 2020). However, I. Chaban et al. (2020) reported that ovules with this type of development undergo further development, which was not observed in our study.

A third alternative outcome of the interspecific hybridization (Fig. 2K–P) rather than of self-pollination (Fig. 1K) proved to be more typical. A pseudoembryo surrounded by pseudoembryonic tissue developed inside seed-like bodies, a pattern not reported in earlier studies (Chambonnet, 1996; Bal, Abak, 2003; Piosik et al., 2019). In these cases, all observed pseudoembryos reached the globular or late-globular stage (Fig. 1K; 2K, L). Numerous studies on Solanaceae hybrids have shown that seed development often stops at the globular stage, which marks a critical step in overcoming the post-zygotic interspecific barrier to crossability (Baek et al., 2016; Roth et al., 2018; Piosik et al., 2019). Scientists, however, seldom discuss the nature of the tissues surrounding the embryo and often overlook the possibility of pseudoembryonic tissue formation in tomato. Hybrid seeds are usually described as either developing a normal endosperm or having degenerated endosperm accompanied by the formation of a multilayered IT (Baek et al., 2016; Roth et al., 2018), which does not resemble the pseudoembryonic tissue observed in this study.

Particular consideration should be given to the characteristics of the pseudoembryos surrounded by pseudoembryonic tissue. Similar abnormalities have been reported in several angiosperms with unitegmic ovules, where adventive embryos may arise from IT (Kapil, Tiwari, 1978). In our material, however, each pseudoembryo developed strictly at the micropylar pole, which contrasts with adventive embryos originating from IT in random sites. Evidence supporting hybrid origin comes from ovules at early stages after pollination, where cells presumably corresponding to an embryo were found at the same position, the micropylar pole, despite the beginning of IT proliferation (Fig. 2H). These observations suggest that pseudoembryonic tissue forms from IT, while the embryo develops at the micropylar pole. As we did not observe endosperm degeneration in ovules containing a pseudoembryo, it seems unlikely that the seed-like bodies resulted from replacement of degenerated endosperm with pseudoembryonic tissue. Taken together, our findings indicate that the pseudoembryos formed in this study may be of hybrid origin.

In this research, we isolated ovules on nutrient medium. Consequently, the *in vitro* material consisted of normally developing seeds along with seed-like bodies, either parthenocarpic or possessing embryos. During the cultivation of isolated ovules, relatively rapid embryo and plantlet growth on the nutrient medium was observed only in the st8×Sn combination. The plantlets rapidly grew upon transplanting into soil, which is consistent with the observations of U. Bal and K. Abak (2003). In the combinations st6×Sn, st4×Sn, and st6(lik)×Sn, plant regeneration occurred through the formation of organogenic callus from the isolated embryoids. These plantlets developed slowly, and the majority of them failed to root. A similar growth trend was noted by Ł. Piosik et al. (2019).

With regard to morphological traits, all obtained plants resembled *S. lycopersicum*, consistent with the results of U. Bal and K. Abak (2003) and D. Chambonnet (1996). This is notable because interspecific tomato hybrids often exhibit phenotypes with prominent features of the paternal component (Gavrilenko et al., 2001; Piosik et al., 2019). However, detailed analysis of the phenotypic characteristics of the studied plants revealed certain traits of *S. sisymbriifolium*, along with the emergence of traits absent from either parental form. We rule out the possibility of phenotypic segregation in the progeny of the parental tomato line, as we employed genetically stable multigenerational inbred lines.

Cytogenetic analysis of root meristems revealed mixoploidy, with chromosome numbers ranging from 16 to 26. U. Bal and K. Abak (2003) also reported the occurrence of mixoploidy ($2n = 24–26$) in the meristems of plants derived from *S. lycopersicum* × *S. sisymbriifolium* hybridization; however, they interpreted this phenomenon as a sign of chromosome doubling in haploid plants. In contrast to the results of U. Bal and K. Abak (2003), a significant proportion of cells in our study contained fewer than 24 chromosomes. Notably, plants from the st8×Sn combination exhibited no cells with an increased chromosome count. This reduction is likely attributable to the elimination of *S. sisymbriifolium* chromosomes. Mixoploidy in meristems can serve as both a marker of hybrid origin and a response to plant regeneration through callus culture. However, the st8×Sn plants were obtained through direct morphogenesis without a callus stage, which further supports their hybrid origin. Furthermore, E.W. Lindstrom and K. Koos (1931) obtained a doubled haploid from the callus of a haploid tomato; notably, its microsporogenesis was characterized as regular, producing microspores with a consistent count of $n = 12$ chromosomes.

Researchers (Chambonnet, 1996; Bal, Abak, 2003; Piosik et al., 2019) who experimented with the hybridization of *S. lycopersicum* and *S. sisymbriifolium* did not study microsporogenesis in the obtained plants. T. Gavrilenko et al. (2001) obtained a somatic hybrid of *S. lycopersicum* and *S. tuberosum* and noted the presence of univalents in meiotic cells, as well as lagging chromosomes during anaphase, meiosis telophase I, and prophase II. Our study also revealed numerous meiotic irregularities in the obtained plants, manifesting themselves in chromosome lagging, uneven distribution of chromosomes between daughter cells, and the formation of univalents, trivalents, and bivalents with a single chiasma. Meanwhile, the *S. lycopersicum* and *S. sisymbriifolium* lines used in our study had aneuploid cells ($n - 1$), which may be attributed to high temperatures (33–37 °C) during plant cultivation (Schindfessel et al., 2023).

On the grounds of the multitude of processes studied that follow the pollination of *S. lycopersicum* with *S. sisymbriifolium* pollen, the production of plants in embryoculture and the differences in phenotypes between the parental tomato lines used and the obtained plants, and their cytogenetic features, we consider the obtained plants to be interspecific hybrids. Further research is required to confirm the hybrid origin of

the obtained plants, including the development of specific molecular markers or the optimization of the GISH (genomic *in situ* hybridization) protocol, which is beyond the scope of this study. However, this does not diminish the importance of our results, which help us understand what happens during the hybridization of these species. They also show that there is a chance that valuable traits from *S. sisymbriifolium* can be added to the resulting plants, as we have shown that they can produce fruit.

Conclusion

In crosses between *S. lycopersicum* and *S. sisymbriifolium*, two alternative developmental types leading to the formation of seed-like bodies were found in developing fruits, in addition to normally developing seeds and ovules arrested during early development. Some of these seed-like bodies represented parthenocarpic seeds filled with pseudoembryonic tissue formed as a result of proliferation of the integumentary tapetum. Certain seed-like bodies containing pseudoembryonic tissue also enclosed an embryo arrested at the globular stage. Its position at the micropylar pole and the absence of evidence for an adventive origin suggest that the embryo may be of hybrid nature. The developmental arrest of embryos and the formation of pseudoembryonic tissue in place of endosperm provide further evidence for postzygotic barriers that hinder the production of interspecific hybrids between *S. lycopersicum* and *S. sisymbriifolium*.

We used the method of isolating ovules and embryos from them to hybridize cherry tomato lines and *S. sisymbriifolium* and obtained 12 viable plants. While all of the obtained plants had tomato-like phenotypes, they demonstrated high chromosomal instability in their meristematic cells. Additionally, we studied the microsporogenesis process for the first time in *S. lycopersicum* and *S. sisymbriifolium* crosses and discovered a significant number of meiosis I and II abnormalities, such as uneven chromosome distribution between daughter cells, lagging chromosomes, and the formation of univalents and trivalents. This finding supports the hypothesis that the obtained samples are hybrids.

The obtained results add to our understanding of the processes occurring during distant hybridization and open up the possibility of using *S. sisymbriifolium* as a source of valuable traits in tomato breeding.

References

- Abbas M.M., Ghani M.A., Ziaf K., Khan T.N., Jahangir M.M., Awan F.S., Hussain M.M. Development and characterization of F₁ hybrid tomato by inter and intraspecific hybridization. *Pak J Agric Sci.* 2024;61(1):199-207. doi 10.21162/PAKJAS/24.218
- Ajaharuddin S.M., Lal M., Yadav A., Kumar N., Dhakad A., Sinha G., Singh B.P., Upadhyay A. Breeding for resistance against pest and diseases in tomatoes: a review. *J Sci Res Rep.* 2024;30(6):469-479. doi 10.9734/jsrr/2024/v30i62063
- Baek Y.S., Royer S.M., Broz A.K., Covey P.A., López-Casado G., Nuñez R., Kear P.J., Bonierbale M., Orillo M., van der Knaap E., Stack S.M., McClure B., Chetelat R.T., Bedinger P.A. Interspecific reproductive barriers between sympatric populations of wild tomato species (*Solanum* section *Lycopersicon*). *Am J Bot.* 2016;103(11):1964-1978. doi 10.3732/ajb.1600356
- Bal U., Abak K. Attempts of haploidy induction in tomato (*Lycopersicon esculentum* Mill.) via gynogenesis I: pollination with *Solanum sisymbriifolium* Lam. pollen. *Pak J Biol Sci.* 2003;6(8):745-749. doi 10.3923/pjbs.2003.745.749
- Barykina R.P., Veselova T.D., Devyatov A.G., Dzhalilova Kh.Kh., Iljina G.M., Chubatova N.V. Handbook of Botanical Microtechnique. Principles and methods. Moscow: MSU Publ., 2004 (in Russian)
- Biswas D., Haque S.M., Ghosh B. *Solanum sisymbriifolium* Lam.: an underutilised plant with future prospects in nutrition and medicine. *Proc Indian Natl Sci Acad.* 2023;89(3):445-469. doi 10.1007/s43538-023-00191-9
- Chaban I., Baranova E., Kononenko N., Khaliluev M., Smirnova E. Distinct differentiation characteristics of endothelium determine its ability to form pseudo-embryos in tomato ovules. *Int J Mol Sci.* 2020;21(1):12. doi 10.3390/ijms21010012
- Chambonnet D. Essais d'haploïdisation de la tomate. In: Report D'Activite 1995-1996 Station D'Amelioration Des Plantes Maraicheres D'Avignon-Montfavet, 1996;84-85
- Collonnier C., Fock I., Daunay M.C., Servaes A., Vedel F., Siljak-Yakovlev S., Souvannavong V., Sihachakr D. Somatic hybrids between *Solanum melongena* and *S. sisymbriifolium*, as a useful source of resistance against bacterial and fungal wilts. *Plant Sci.* 2003;164(5):849-861. doi 10.1016/S0168-9452(03)00075-X
- Cooper D.C. Macrosporogenesis and the development of the macrogametophyte of *Lycopersicon esculentum*. *Am J Bot.* 1931;18(9):739-748. doi 10.1002/j.1537-2197.1931.tb04945.x
- Flier W.G., Van Den Bosch G.B.M., Turkensteen L.J. Epidemiological importance of *Solanum sisymbriifolium*, *S. nigrum* and *S. dulcamara* as alternative hosts for *Phytophthora infestans*. *Plant Pathol.* 2003;52(5):595-603. doi 10.1046/j.1365-3059.2003.00922.x
- Foolad M.R. Genome mapping and molecular breeding of tomato. *Int J Plant Genomics.* 2007;2007(1):64358. doi 10.1155/2007/64358
- Gavrilenko T., Thieme R., Rokka V.M. Cytogenetic analysis of *Lycopersicon esculentum* (+) *Solanum etuberosum* somatic hybrids and their androgenetic regenerants. *Theor Appl Genet.* 2001;103(2):231-239. doi 10.1007/s001220100626
- Hajihassani A., Rutter W.B., Schwarz T., Woldemeskel M., Ali M.E., Hamidi N. Characterization of resistance to major tropical root-knot nematodes (*Meloidogyne* spp.) in *Solanum sisymbriifolium*. *Phytopathology.* 2020;110(3):666-673. doi 10.1094/PHYTO-10-19-0393-R
- Harberd D.J. A simple effective embryo culture technique for brassica. *Euphytica.* 1969;18:425-429. doi 10.1007/BF00397792
- Kapil R.N., Tiwari S.C. The integumentary tapetum. *Bot Rev.* 1978;44(4):457-490. doi 10.1007/BF02860847
- Kirov I., Divashuk M., Van Laere K., Soloviev A., Khrustaleva L. An easy "SteamDrop" method for high quality plant chromosome preparation. *Mol Cytogenet.* 2014;7(1):21. doi 10.1186/1755-8166-7-21
- Lindstrom E.W., Koos K. Cyto-genetic investigations of a haploid tomato and its diploid and tetraploid progeny. *Am J Bot.* 1931;18(6):398-410. doi 10.1002/j.1537-2197.1931.tb09599.x
- Palamarchuk I.A., Veselova T.D., Antonova T.S. Peculiarities of the reproductive organ development in glasshouse tomatoes. *Vestnik Moskovskogo Universiteta.* 1975;8(6):113-116 (in Russian)
- Piosik Ł., Ruta-Piosik M., Zenkteler M., Zenkteler E. Development of interspecific hybrids between *Solanum lycopersicum* L. and *S. sisymbriifolium* Lam. via embryo calli. *Euphytica.* 2019;215(2):31. doi 10.1007/s10681-019-2358-9
- Roth M., Florez-Rueda A.M., Griesser S., Paris M., Städler T. Incidence and developmental timing of endosperm failure in post-zygotic isola-

- tion between wild tomato lineages. *Ann Bot.* 2018;121(1):107-118. doi [10.1093/aob/mcx133](https://doi.org/10.1093/aob/mcx133)
- Rubio F., Alonso A., García-Martínez S., Ruiz J.J. Introgression of virus-resistance genes into traditional Spanish tomato cultivars (*Solanum lycopersicum* L.): effects on yield and quality. *Sci Hortic.* 2016; 198:183-190. doi [10.1016/j.scienta.2015.11.025](https://doi.org/10.1016/j.scienta.2015.11.025)
- Ruzin S.E. Plant Microtechnique and Microscopy. Oxford University Press, 1999
- Sadashiva A.T., Hanson P., Reddy M.K., Ravishankar K.V., Prasad M., Prasanna H.C., Madhavi Reddy K., ... Kumar M., Kaushal A., Rai A.K., Jatav V., Bhat L. Breeding tomato (*Solanum lycopersicum* L.) for resistance to biotic and abiotic stresses. *J Hortl Sci.* 2017;12(2):91-105. doi [10.24154/jhs.v12i2.2](https://doi.org/10.24154/jhs.v12i2.2)
- Schindfessel C., De Storme N., Trinh H.K., Geelen D. Asynapsis and meiotic restitution in tomato male meiosis induced by heat stress. *Front Plant Sci.* 2023;14:1210092. doi [10.3389/fpls.2023.1210092](https://doi.org/10.3389/fpls.2023.1210092)
- Yerasu S.R., Gowda M.T., Krishna R., Kaur C., Singh S., Saha S., Kumar R. Development of interspecific hybrids through embryo rescue for heat-stable nematode resistance (*Mi-9* gene) from *Solanum arcanum* in tomato. *Plant Cell Tiss Organ Cult.* 2023;154(3):703-711. doi [10.1007/s11240-023-02544-0](https://doi.org/10.1007/s11240-023-02544-0)

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

Received November 29, 2025. Revised February 18, 2026. Accepted March 23, 2026.

doi 10.18699/vjgb-26-60

Differentiation of subgenomes in StY-genomic species of the genus *Elymus* (Triticeae, Poaceae) from the territory of Russia according to sequencing data of the nuclear gene *GBSS1* (*waxy*)

A.V. Agafonov ¹ , E.V. Shabanova (Kobozeva) ¹, A.A. Bondar ², O.V. Dorogina ¹¹ Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia² Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia agalex@mail.ru

Abstract. According to the latest data, 55 species of the genus *Elymus* are distributed in Russia, carrying three subgenomic combinations (StH, StY, and StHY). Among them, the StY-genomic group comprises only 10 species, with varying degrees of understanding of their evolutionary relationships. Ancestral taxa of the genus *Pseudoroegneria* donated the St subgenome to all modern species of the genus *Elymus*. The StY-genomic group of species possesses an additional Y subgenome of unknown origin, which, according to some data, is close to the St subgenome. We studied the phylogenetic relationships between StY-genomic species from Russia, five species of the genus *Pseudoroegneria*, and five species of the genus *Hordeum* from the NCBI GenBank by comparing the nucleotide sequences of the nuclear gene *GBSS1* from exons 9 to 14. One of the main objectives was a comparative analysis of phylogenetic patterns based on (i) more conservative exons and (ii) introns that do not encode the amino acid sequences of the enzyme. All gene variants from the St subgenomes of the studied species are divided into three clusters according to three marker groups of nucleotide sequence of clones in the genus *Pseudoroegneria*: Central Asian (St₁) with the reference species *P. strigosa*, North American (St₂) with the marker species *P. spicata*, and Middle Eastern (St₃) with two species, *P. tauri* and *P. libanotica*, to which the Eastern European species *P. stipifolia* gravitates. The H subgenome, originating from ancestral taxa of the genus *Hordeum* (*Criteseion*), was not detected in any of the studied species. The cluster of Y subgenome in the “introns” variant is generally visually less differentiated than in the “exons” variant. This fact contradicts the established notion that gene regions responsible for the synthesis of enzymatic molecules are more conserved. Among the most notable characteristics in the comparison of nucleotide sequences is the presence of a special Far Eastern race of *E. gmelinii* with St₃ sequences instead of St₁ and the location of the sequences of all clones of the North Kazakhstan accession *E. fedtschenkoi* KSA-0935 in the St₂ cluster instead of St₃.

Key words: molecular markers; phylogeny; subgenome; taxonomy; *Elymus*; *Pseudoroegneria*

For citation: Agafonov A.V., Shabanova (Kobozeva) E.V., Bondar A.A., Dorogina O.V. Differentiation of subgenomes in StY-genomic species of the genus *Elymus* (Triticeae, Poaceae) from the territory of Russia according to sequencing data of the nuclear gene *GBSS1* (*waxy*). *Vavilovskii Zhurnal Genetiki i Selektzii* = *Vavilov J Genet Breed*. 2026;30(4):581-592. doi 10.18699/vjgb-26-60

Funding. The study was carried out within the framework of State Assignment of CSBG SB RAS No. 126021617423-1 “Development of scientific foundations and technologies for the assessment, conservation, and restoration of natural and cultural biodiversity, including rare plant species, and their rational use”. Materials of the bio-resource scientific collections of the CSBG SB RAS USU No. 440534 and USU No. 440537 (Herbaria NS, NSK) were used. DNA sequencing was performed in Joint Core Facility “Genomics” (FWGN-2025-0017, Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk).

Дифференциация субгеномов среди StY-геномных видов рода *Elymus* (Triticeae, Poaceae) с территории России по данным секвенирования ядерного гена *GBSS1* (*waxy*)

А.В. Агафонов ¹ , Е.В. Шабанова (Кобозева) ¹, А.А. Бондарь ², О.В. Дорогина ¹¹ Центральный сибирский ботанический сад Сибирского отделения Российской академии наук, Новосибирск, Россия² Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия agalex@mail.ru

Аннотация. По последним данным, на территории России распространены 55 видов рода *Elymus*, которые являются носителями трех субгеномных комбинаций: **StH**, **StY** и **StHY**. Среди них **StY**-геномная группа насчитывает всего 10 видов с разной степенью изученности эволюционных связей. Предковые таксоны рода *Pseudoroegneria* были донорами субгенома **St** для всех современных видов рода *Elymus*. **StY**-геномная группа видов обладает дополнительным субгеномом **Y** с неизвестным происхождением, который по некоторым сведениям близок к субгеному **St**. Нами изучены филогенетические взаимоотношения между **StY**-геномными видами с территории России, пятью видами рода *Pseudoroegneria* и пятью видами рода *Hordeum* из Генбанка NCBI путем сравнения нуклеотидных последовательностей ядерного гена *GBSS1* с 9-го по 14-й экзон. Проведен анализ участков более консервативных экзонов в сравнении с интронами, которые не кодируют аминокислотные последовательности фермента. Все варианты гена из субгеномов **St** у изученных видов подразделяются на три кластера сообразно трем маркерным группам нуклеотидных последовательностей клонов рода *Pseudoroegneria*: центральноазиатской (**St**₁) с реперным видом *P. strigosa*, североамериканской (**St**₂) с маркирующим видом *P. spicata* и ближневосточной (**St**₃) с двумя видами – *P. tauri* и *P. libanotica*, к которым тяготеет восточноевропейский вид *P. stipifolia*. Ни у одного из изученных видов субгеном **H**, имеющий происхождение от предковых таксонов рода *Hordeum* (*Critiesion*), не обнаружен. Что касается **Y**-субгенома в варианте «интроны», то в целом этот кластер менее дифференцирован, чем в варианте «экзоны». Данный факт противоречит сложившимся представлениям о большей консервативности генных участков, ответственных за синтез ферментных молекул. Среди наиболее примечательных характеристик в сравнении нуклеотидных последовательностей следует отметить наличие особой дальневосточной расы *E. gmelinii* с последовательностями **St**₃ вместо **St**₁ и расположение последовательностей всех клонов североказахстанского образца *E. fedtschenkoi* KSA-0935 в кластере **St**₂ вместо **St**₃.

Ключевые слова: молекулярные маркеры; филогения; субгеном; таксономия; *Elymus*; *Pseudoroegneria*

Introduction

The genus of perennial allopolyploid herbs *Elymus* L. is the largest in the tribe Triticeae Dumort. (Poaceae Barnh.). There are between 150 and 200 species worldwide with different genomic constitutions, all of which are considered to be unified by the **St** subgenome. According to generalized data, this subgenome is present in different taxa in at least eleven combinations with six other subgenomes of the ancestral taxa of the tribe: **H**, **Y**, **P**, **W**, **Ns**, and **Xm** (Tan et al., 2024). In recent decades, there has been a trend toward splitting the genus *Elymus* s. l. into eight independent allopolyploid genera, including *Elymus* s. str. (**StH**), *Roegneria* C.Koch (**StY**), *Douglasdeweya* C. Yen, J.L. Yang & B.R. Baum (**StP**), *Campeioestachys* Drobow (**StYH**), *Kengyilia* C. Yen & J.L. Yang (**StYP**), *Anthosachne* Steudel (**StYW**), *Pascopyrum* Å. Löve (**StHNSxm**), and *Connorochloa* Barkworth, S.W.L. Jacobs & H.Q. Zhang (**StYWH**).

Since the genus belongs to the tertiary gene pool (**GP-3**) of major cereal crops, its representatives have often been included in detailed cytogenetic studies (Dewey, 1984). In recent decades, the cytogenetic method of determining the genomic constitution of species has been supplemented and even superseded by more technologically advanced DNA sequencing methods (Mason-Gamer, 2001; Baum et al, 2003). Researchers are shifting their focus to analyzing the origins and phylogenetic relationships between basic subgenomes using molecular genetic methods.

The main diversity of species in the **StY** genomic group is concentrated in China, Korea, and Japan. A number of contradictory data on the phylogenetic relationships between the **St** and **Y** subgenomes have been published. Specifically, in the search for a **Y** subgenome donor, hybridization was conducted between diploid *Hordeum* L. species (carriers of the **H** subgenome) and **StY**-genomic Asian *Elymus* species.

Triploid hybrids were characterized by a relatively low level of chromosome pairing in metaphase I, demonstrating extremely low homology between the **H**, **St**, and **Y** haplomes and, consequently, the absence of the **H** haplome in these *Elymus* species (Lu, von Bothmer, 1990). In addition, based on the analysis of chromosomal pairing in hybrids, a low relationship between the **St** and **Y** subgenomes was shown (Sakamoto, 1964; Lu, von Bothmer, 1989).

The nuclear ribosomal internal transcribed spacer (ITS) and chloroplast intergenic spacer *trnL-F* sequences were analyzed in 457 *Elymus* accessions containing diverse genomes. The results confirmed that the **St**, **H**, **P**, and **W** subgenomes of polyploid *Elymus* were introduced from the genera *Pseudoroegneria* (Nevski) A. Löve, *Hordeum*, *Agropyron* Gaertn., and *Australopyrum* (Tzvelev) A. Löve, respectively, but that the **St** and **Y** subgenomes phylogenetically trace back to a common ancestor (Liu et al., 2006).

In another study, one of the RAPD markers specific to the **Y** subgenome was converted into a sequence-tagged site (STS) marker. This modified STS marker confirmed the presence of a **Y** subgenome in 42 accessions of **StY**-genomic *Elymus* species (Okito et al., 2009). To identify possible donors of the **Y** subgenome, 43 accessions of the diploid *Pseudoroegneria* species with the **St** subgenome were analyzed using the STS marker. The authors concluded that some **StY**-genomic *Elymus* species harbored this particular variant of the **Y** subgenome, shared with the **St** genome of *Pseudoroegneria*, that is, a precursor subgenome designated **StY**.

Later, the single-copy nuclear gene encoding elongation factor G (*EF-G*) was analyzed among 28 accessions of polyploid *Elymus* species and 45 accessions of diploid species of the tribe Triticeae (Sun, Komatsuda, 2009). The data supported the hypothesis that the **Y** subgenome arose in a diploid species but had a different origin compared to the **St** subgenome. This

conclusion was supported by the single-copy nuclear RNA polymerase II gene (*RPB2*) from 58 accessions of *Pseudoroegneria* and *Elymus* species (Yan et al., 2011).

Further, based on the topology of the phylogenetic tree using chloroplast noncoding DNA sequences from 56 accessions of nine polyploid *Elymus* species, the authors showed that the **St** and **Y** subgenomes did not originate from the same donor, and that the **Y** subgenome likely originated from the **H** genome of *Hordeum* species (Song et al., 2015).

In a recent report, the following important conclusions were made based on the analysis of sequence data from three nuclear DNA regions (*Acc1*, *DMC1*, and *Pgk1*) and three chloroplast DNA regions (*nrITS*) (*matK*, *rbcL*, and *trnL-trnF*) (Pan et al., 2025):

1. The Triticeae polyploid species, which consist of different genome types, should be treated as a distinct genera.
2. Some polyploid species with the **St** genome have undergone independent allopolyploidization events in different regions of distribution.
3. The **Y** genome originated from an unknown or extinct diploid species closely related to *Peridictyon sanctum* (Janka Seberg, Fred. & Baden, which is distributed on the Balkan Peninsula (**Xp** genome)).

In this regard, we are inclined to support the view expressed three years earlier, which proposed an autotetraploid origin of the **Y** subgenome through recurrent hybridization (Liu et al., 2022). According to these data, the complex **St** genomes of the ancestors of the genus *Pseudoroegneria* in the polyploid state may have gained greater opportunities for intraspecific differentiation through repeated and back-hybridization. As a result, modified variants of the **St** genomes in some lineages evolved into intermediate **St^Y** genomes.

Thus, one likely scenario is the origin of the modern **Y** subgenome from taxa of the genus *Pseudoroegneria* carrying ancestral **St** subgenome variants, while the origin of the **Y** subgenome from the relatively distant **H** and **Xp** genomes appears unlikely.

As repeatedly demonstrated in the work of Dr. R. J. Mason-Gamer's research group from 1998 to 2024, the nuclear gene encoding granule-associated starch synthase 1 (*GBSSI*, *waxy*) can serve as an effective genetic marker of macro- and microevolutionary relationships between taxa of different levels. The gene's schematic was first presented using *Zea mays* L. as an example (Mason-Gamer et al., 1998), but information on its specificity among North American species of the genus *Elymus* was published three years later (Mason-Gamer, 2001).

We confirmed the feasibility of using this gene as an indicator of microevolutionary processes among species of the genus *Elymus* (Agafonov et al., 2019, 2024a, b). It is generally accepted that intron sequences are not subject to strict evolutionary selection; therefore, they can accommodate numerous intermediate microevolutionary events (substitutions, insertions, etc.) that occurred during the evolution of subgenomes. Moreover, the biological function of introns, as one of the constituent parts of the nucleotide sequences of a number of functional genes, appears to be quite important in relation to the regulation of gene expression (Koonin, 2006).

According to the latest data (Tzvelev, Probatova, 2019), 55 species of the genus *Elymus* are distributed in Russia. These are allotetra- and allohexaploids with the **StStHH**, **StStYY**, and **StStHHYY** genomes (Agafonov et al., 2020). Among them, the **StY** genome group comprises only 10 species, with varying degrees of understanding of their evolutionary relationships.

Previously, we proposed new combinations of taxa within this group in a broader sense, specifically for *E. pendulinus* (Nevski) Tzvelev s. l. (Kobozeva, Agafonov, 2015) and *E. ciliaris* (Trin.) Tzvelev s. l. (Agafonov et al., 2021; Shabanova (Kobozeva), Agafonov, 2023). Species boundaries are delineated based on the “principle of microevolutionary complexes”, that is, a set of closely related taxa that defines the phylogenetic unity of a species and meets other criteria. The base species includes intraspecific taxa based on their relationships established by reproductive and molecular genetic criteria. This principle can ensure methodological synergy between the fundamental principles of traditional taxonomy, on the one hand, and experimental biology incorporating modern approaches and methods, on the other. As a result, the total number of **StY**-genomic species in Russia has been reduced to seven.

This study aimed to trace the evolutionary relationships between the **St** subgenomes and to assess the position of the **Y** subgenome on the evolutionary tree among **StY**-genomic *Elymus* species from Russia by comparing the nucleotide sequences of *GBSSI* gene fragments of exons from 9 to 14. Specifically, the goal was to identify differences in the phylogenetic picture based on the sequences of more conserved exons compared to introns not involved in the final synthesis of enzyme molecules.

Materials and methods

Plant material. The study included 32 natural accessions of all **StY**-genomic species listed by N.N. Tzvelev and N.S. Probatova (2019). All of these accessions from our collection underwent morphological confirmation of taxonomic rank in experimental open-field plots and in a climate chamber to ensure compliance with modern taxon descriptions (Table S1)¹.

Two European species, *E. caucasicus* (C. Koch) Tzvelev and *E. panormitanus* (Parl.) Tzvelev, are not present in our living collection. The Central Asian species *E. abolinii* (Drobow) Tzvelev was additionally included in the comparative analysis, as we suspect it grows within Russia, but evidence is currently lacking.

The *GBSSI* gene sequences of five species of the genus *Pseudoroegneria* and five species of the genus *Hordeum* from the NCBI GenBank (URL: <http://www.ncbi.nlm.nih.gov/nuccore>) were used as reference sequences (see the Table). Unlike the collection samples, the accession numbers of clones from NCBI are given in the dendrograms in front of the species names.

Ancient representatives of these taxa were the donors of the **St** and **H** subgenomes, respectively, in modern species of

¹ Supplementary Table S1 is available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Agaf_Engl_30_4.pdf

Reference species accessions and nucleotide sequence numbers taken from NCBI GenBank

Species	Accession, origin	Sequence number in NCBI
<i>Pseudoroegneria strigosa</i> (Schult.) Á. Löve	PI 499637 China	EU282323 St
	PI 531755 China	AY360823 St
<i>P. spicata</i> (Pursh) Á. Löve	PI 232117 USA	AF079281 St
	PI 610986 USA	AY010999 St
<i>P. libanotica</i> (Hack.) D.R. Dewey	PI 228391 Iran	EU282324 St
<i>P. tauri</i> (Boiss. & Balansa) Á. Löve	PI 380652 Iran	EU282326 St
<i>P. stipifolia</i> (Czern. ex Nevski) Á. Löve	PI 313960 Russia	JX259496 St
<i>Hordeum jubatum</i> L.	RJMG 106 USA	AY010963 H
<i>H. bogdanii</i> Wilensky	PI 531760 China	EU282317 H
<i>H. brevisubulatum</i> (Trin.) Link	PI 401387 Iran	AY010961 H
<i>H. pusillum</i> Natt.	CIho 15654 USA	EU282321 H
<i>H. californicum</i> Covas & Stebbins	MA-138-1-4 USA	AF079273 H
<i>Bromus tectorum</i> L.	–	AY362757

the genus *Elymus*. Moreover, species of the genus *Pseudoroegneria*, confined to different ranges in Northern Eurasia, could have been the donors of the **St** subgenome in different species of the relatively young **StY** genome group of the genus *Elymus*.

For example, *Pseudoroegneria strigosa* was originally distributed in Central Asia and has also been recorded as native in Crimea and Greece (<https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:418876-1>. First published in Taxon, 1980. 29: 168).

Pseudoroegneria spicata is native to the western part of the North American continent (<https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:1159330-2>. First published in Taxon, 1980. 29: 168).

Pseudoroegneria libanotica and the following species are confined to territories in the Middle East (<https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:942121-1>. First published in P. Gustafson (ed.), Gene Manipulat. in Pl. Improv.: 1984: 272).

Pseudoroegneria tauri (<https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:914501-1>. First published in Feddes Rept. 1984. 95: 445).

The fifth species, *Pseudoroegneria stipifolia*, has a more northerly range than the previous two (<https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:914500-1>. First published in Feddes Rept. 1984. 95: 445).

With regard to the use of molecular markers in the study of the genus *Elymus*, significant results were obtained by Dr. R. Mason-Gamer and co-workers (Helfgott, Mason-Gamer, 2004; Mason-Gamer, 2013; Mason-Gamer et al., 1998, 2010a, b). In particular, their studies showed that information on the nucleotide sequences of the low-copy gene “waxy”

(granule-bound starch synthase 1 (*GBSSI*)) is consistent with cytogenetic data regarding the genomic constitution and evolutionary origin of North American (Mason-Gamer, 2001) and Asian (Mason-Gamer, 2010a) species of the genus *Elymus*.

The relatively small number of **StY**-genomic species in Russia makes it possible to trace the phylogenetic origin of the **St** subgenome in this group from the ancestral taxa of the genus *Pseudoroegneria* using data on the *GBSSI* gene sequences in 6–8 clones per species accession.

Total DNA isotation, PCR amplification, cloning in a plasmid vector, DNA sequencing and construction of phylogenetic trees. Total DNA was isolated from 20 mg of dry green mass using the NucleoSpin Plant II Kit (Macherey-Nagel, Germany) according to the manufacturer’s standard protocol. *GBSSI* gene fragment overlapping the region from exons 9 to 14 was obtained by PCR using a direct F-for primer (TGCGAGCTCGACAACATCATGCG, Mason-Gamer et al., 1998) and our modified M-bac-Alter1 (GGCGAGCG-GYGCRATCTCSTSGCC) reverse primer.

The *GBSSI* fragment was amplified under the conditions: 1X Q5 Reaction Buffer, 0.2 mM of each dNTP, 2.0 mM of free Mg²⁺, 0.3 μM of forward and reverse primers, about 0.4 ng/μL of genomic DNA, 8 u/mL of Q5 Hot Start II DNA Polymerase, with the addition of Betaine up to 1M and Q5-High GC Enhancer up to 1X. The temperature profile consisted from melting at 98 °C for 30 s, followed by 38 cycles of three stages (denaturation at 98 °C – 5 s; annealing at 69 °C – 10 s; elongation at 72 °C – 1 min), then completion at 72 °C – 2 min and storage at +4 °C. The fragments obtained were then cloned by ligation at the blunt ends into pJET2.1 vector using the CloneJET PCR Cloning Kit (Thermo Scien-

tific, USA). By use of primers from the vector region (Jet_F and Jet_R) in PCR amplification of *E. coli* colonies grown on the plates (NEB stable) we have obtained fixed variants of the cloned *GBSSI* gene. DNA fragments were then purified from PCR components by sorption on AMPure XP magnetic (Beckman Coulter, USA) and sequenced by Sanger method from both sides to assemble contigs and perform a subsequent analysis. From six to eight colony amplified DNA fragments of appropriate size containing separate *GBSSI* clones for each genomic DNA sample were used for sequencing. The sequence AY362757 of *Bromus tectorum* L. was taken as an outgroup sample when constructing dendrograms.

Multiple sequence alignment of the studied *GBSSI* fragments was performed using the MUSCLE algorithm in the Unipro UGENE ver. 31.0 program (Okonechnikov et al., 2012). The aligned sequences were used to construct phylogenetic trees using the maximum likelihood (ML) method (Felsenstein, 1981). A series of dendrograms were constructed in the MEGA ver. 11.0.13 program (Tamura et al., 2021) using the two-parameter (2-parameter distance, K2P) evolutionary model of M. Kimura (1980) for *GBSSI* fragments based on: a) exons separately; b) introns separately, as presumably more mobile sequences in microevolutionary terms. Bootstrap support values are indicated at the nodes. The nucleotide sequences of the clones of the *GBSSI* gene regions we sequenced have been deposited into the GenBank (National Center for Biotechnology Information, NCBI) database. We use the abbreviation NSC to refer to them.

Results

The dendrogram constructed using exon data without introns is shown in Figure 1. The following conclusions were drawn from its analysis.

1. All gene sequences in the **St** subgenome in the studied species are divided into three clusters according to three marker groups of the genus *Pseudoroegneria*: Central Asian (**St₁**) with the reference species *P. strigosa*, North American (**St₂**) with the marker species *P. spicata*, and Middle Eastern (**St₃**) with two species: *P. tauri* and *P. libanotica*. The Eastern European species *P. stipifolia* gravitates toward the third cluster.
2. Conventionally, the first cluster **St₁**, in addition to the marking sequences of *P. strigosa*, contains six NSCs of three varieties of *E. pendulinus* and four identical NSCs of *E. gmelinii*, of which three are Siberian and one is Far Eastern, as well as an NSC of the Central Asian species *E. abolinii*, which is widespread outside of Russia.
3. The North American cluster **St₂** contains a compact group of *E. ciliaris* varieties and a separate branch with two *P. spicata* sequences, along with the NSC of the North Kazakhstan accession *E. fedtschenkoi* KSA-0935_3 (highlighted in dark purple). It should be noted that only six **St₂** sequences were detected in this typical accession of *E. fedtschenkoi*. The genomic specificity of this accession will be described below. The sequence of the Crimean species *E. panormita-*

nus occupied a position between the two branches of this cluster with a bootstrap value of 22.

4. Somewhat unexpectedly, the third cluster **St₃** was differentiated, forming two main branches. One branch is formed by three marker species: *P. tauri*, *P. libanotica*, and *P. stipifolia*. The other branch comprises seven NSCs of *E. nevskii*–*E. fedtschenkoi* and three Far Eastern NSCs of *E. gmelinii*, two of which belong to Kamchatka accessions. It is unlikely that this division of *E. gmelinii* sequences between clusters **St₁** and **St₃** is random. Within this cluster, the sequence AUR-1744_2_Rec has separated the most. Nevertheless, we assigned the genomic formula **St₃** to all NSCs in this cluster.
5. The sequences of the **Y** subgenome are compactly located in a separate cluster, while predominantly preserving the species specificity of the large basic taxa *E. ciliaris*, *E. pendulinus*, *E. gmelinii* and the *E. nevskii*–*E. fedtschenkoi* complex. A single large clade was formed by the NSCs of the *E. nevskii*–*E. fedtschenkoi* complex, along with a group of six NSCs of *E. gmelinii*, which included the *E. nevskii* CHA-8764_3 sequence. The most isolated, as in the previous “exons” variant, were the **Y**-sequences of the European species *E. caucasicus* and *E. panormitanus*, as well as the NSCs of the Kamchatka accession *E. gmelinii* KES-9654_5.
6. An additional point worth noting is the location of the NSC of the Caucasian accession *E. caucasicus* H 2086, which, based on preliminary experiments, we classified as **St₂**. This sequence occupies a position between the **Y** clusters and the **St** complex. Unfortunately, this accession has undergone several stages of exchange between gene banks and genotype collections, so there is no reason to consider its origin reliable.

The relatively close location of the **Y** subgenome cluster to the North American **St₂** subgenome cluster requires a separate historical analysis, since, according to some authors, the **St** and **Y** subgenomes originate from different ancestors (Sun, Komatsuda, 2009; Yan et al., 2011; Song et al., 2015). However, in general, the **StY** genomic group species from Russia have three evolutionary vectors, which is consistent with our understanding of microevolutionary complexes within the genus *Elymus*.

When comparing the results obtained separately for exon (Fig. 1) and intron (Fig. 2) sequences, the following conclusions were reached.

1. In the “introns” dendrogram, all NSCs of *E. pendulinus* and four NSCs of *E. gmelinii* from the Central Asian group **St₁** are located in a single branch close to the branch of two marker sequences of *P. strigosa* and the Uzbek accession of *E. abolinii*. Differences from the “exons” variant were not only evident in the bootstrap values. All NSCs of *E. pendulinus* and *E. gmelinii* were indistinguishable from each other in the **St₁** dendrogram.
2. A similar pattern of differences was evident in cluster **St₂**, where the NSCs of the European *E. caucasicus* and *E. panormitanus* formed a separate branch. Within the

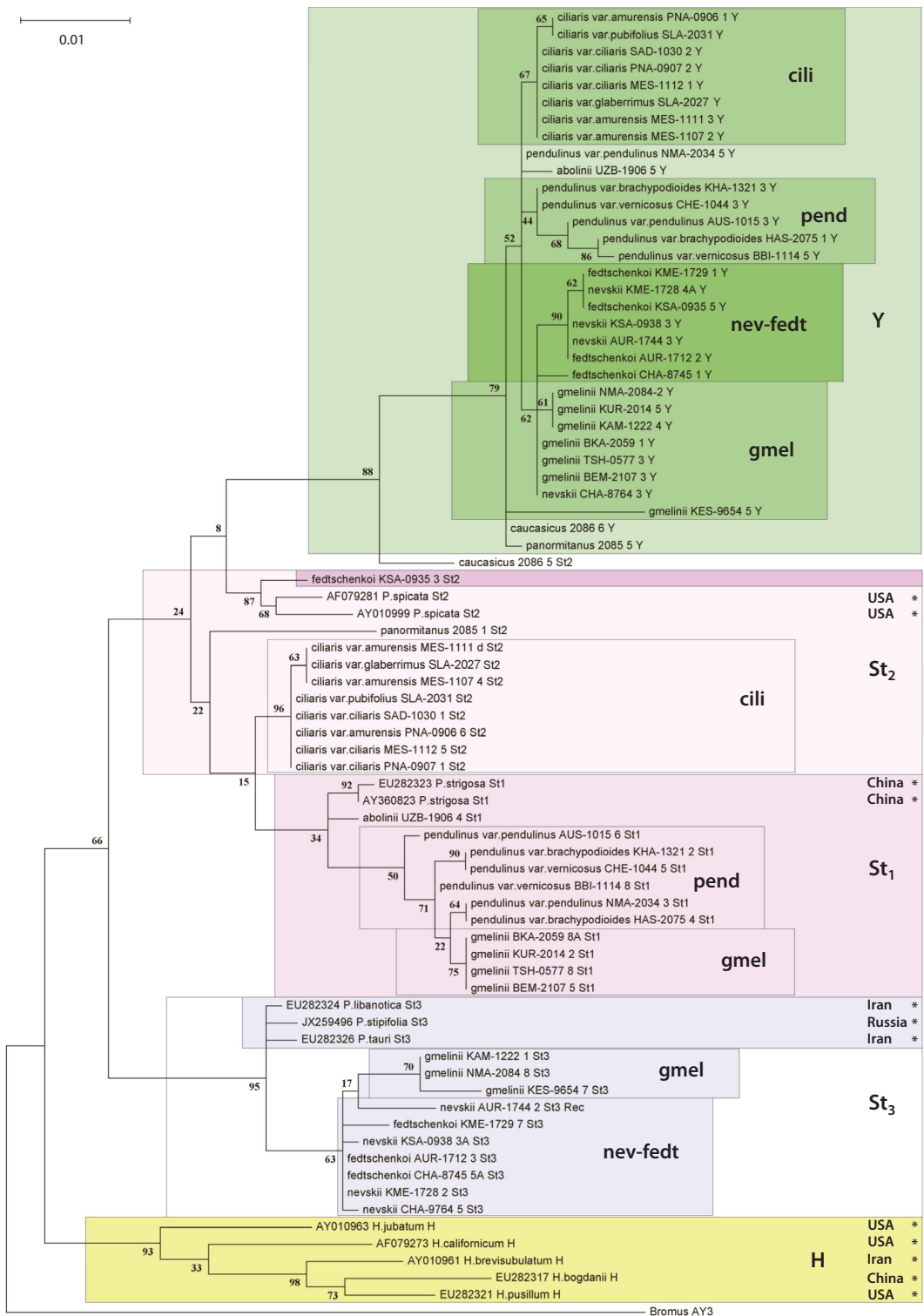


Fig. 1. ML dendrogram constructed based on the analysis of *GBSS1* gene sequences (exons 9–14 only) in the *StY*-genomic group of species from Russia compared to reference species.

Asterisks indicate monogenomic carriers of the *St* and *H* subgenomes, as well as their origin. Bootstrap support values are shown at nodes.

E. ciliaris branch, two NSCs of *E. ciliaris* var. *amurensis* and a single NSC of *E. ciliaris* var. *glaberrimus* diverged slightly from five identical varieties. The sequence of the North Kazakhstan accession of *E. fedtschenkoi* KSA-

0935_3 maintained a close relationship with the marker sequences of *P. spicata*.

3. The group of NSCs *St*₃, as in the “exons” variant, was divided into three branches, one of which included the Middle

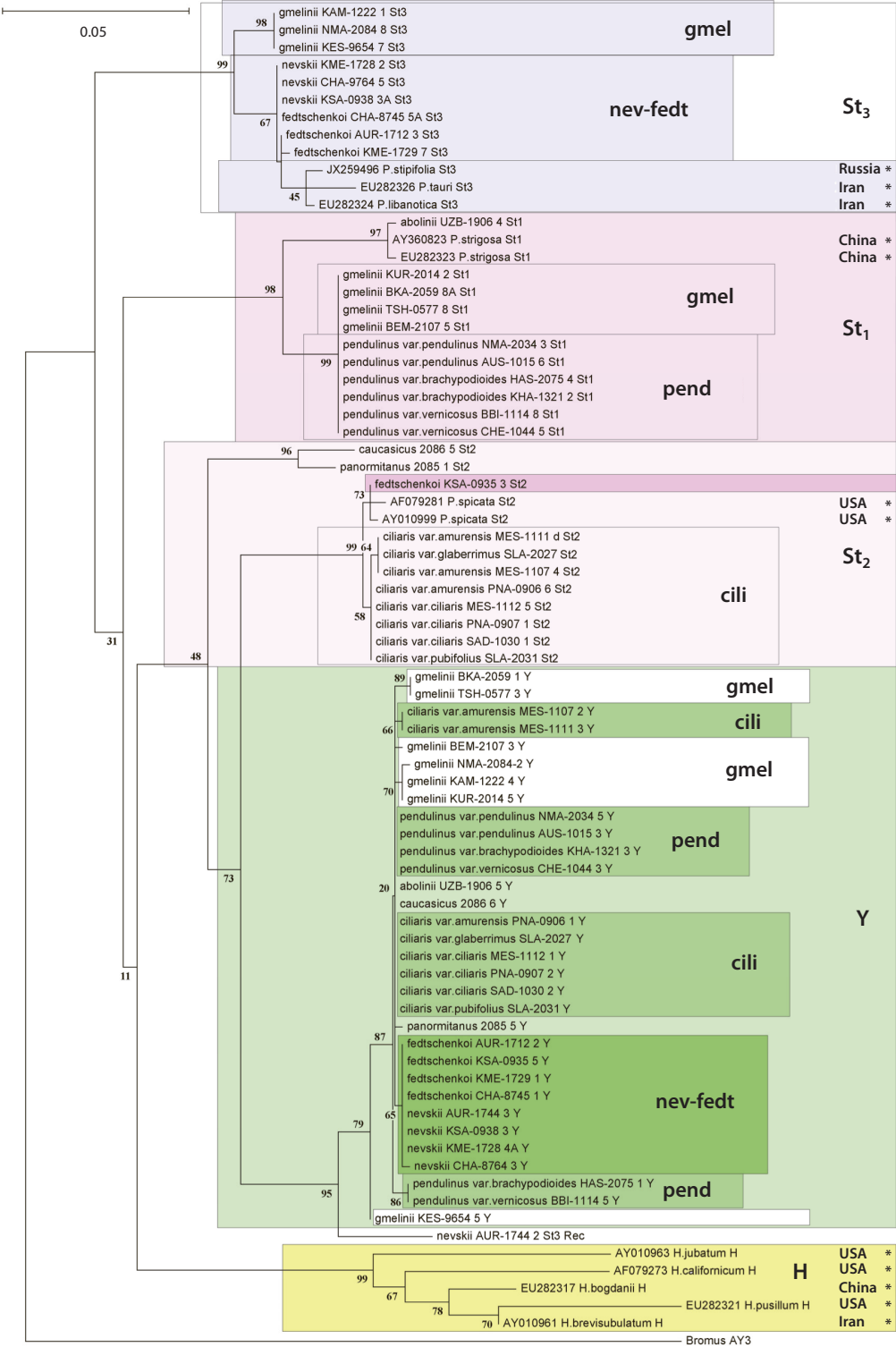


Fig. 2. ML dendrogram constructed based on the analysis of *GBSS1* gene sequences (**introns 9–13 only**) in the **StY**-genomic group of species from Russia compared to reference species. Asterisks indicate monogenic carriers of the **St** and **H** subgenomes, as well as their origin. Bootstrap support values are shown at nodes.

Eastern marker species of *Pseudoroegneria*, another united samples of the *E. nevskii*–*E. fedtschenkoi* complex (except for KSA-0935), and the third consisted of two Kamchatka and one Primorsky NSCs *E. gemlinii*.

Overall, the most unexpected finding in the **St** sequence clusters was the identification of three *E. gemlinii* accessions as belonging to a distinct Far Eastern race of this species, with the **St₃** subgenome variant instead of the **St₁** variants characteristic

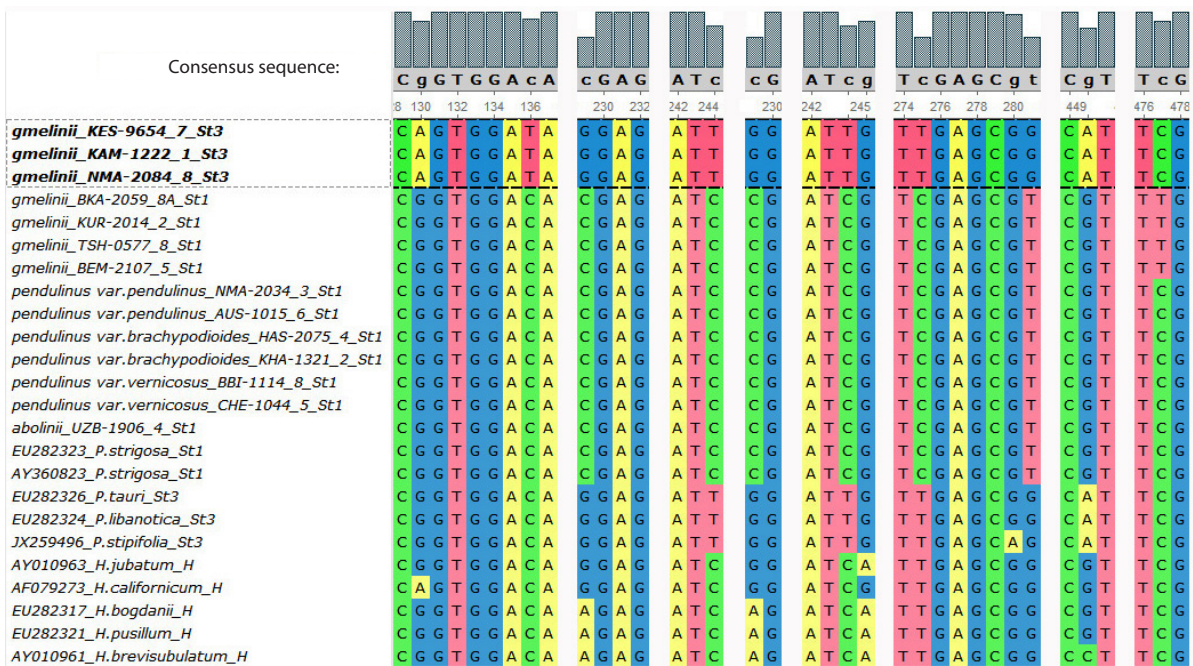


Fig. 3. Differences in the sequences of **exons** 9–14 in three clones of Far Eastern *E. gmelinii* accessions carrying the **St₃** subgenomes and clones of *E. gmelinii*, *E. pendulinus*, and *E. abolinii* carrying the **St₁** subgenomes. Sequence fragments of marker clones of *Pseudoroegneria* are shown for comparison. The numbers above indicate the nucleotide positions in the aligned sequences.

of Siberia. However, one of the Primorsky NSC *E. gmelinii* BKA-2059 was located in the **St₁** cluster along with three Siberian accessions. Furthermore, all six NSCs studied from *E. fedtschenkoi* KSA-0935 (not shown in the dendrograms) were identified as belonging to the **St₂** group instead of **St₃** in both variants.

A study of the nucleotide composition of the introns of the studied species revealed that, despite the low variability of the **Y** subgenome sequences, one nucleotide substitution and one microdeletion in the NSC of *E. gmelinii* KES-9654_5 was sufficient for it to be clearly distinguished from all the others in the dendrogram.

H-subgenomic sequences were not identified in both variants of dendrogram construction among the studied accessions, which confirms the **StY**-genomic constitution of all species.

As noted above, among the most notable characteristics in the tree topology is the presence of a distinct Far Eastern race of *E. gmelinii* with **St₃** NSCs instead of **St₁**. To elucidate possible causes, we analyzed the sets of aligned sequences in critical regions of exons and introns. In Figures 3 and 4, the horizontal frames highlight fragments of three *E. gmelinii* NSCs located in **St₃** clusters in the exon and intron dendrograms, respectively, compared to the **St₁** sequences of other species and the **St₃** sequences of reference *Pseudoroegneria* accessions. The analysis revealed at least seven positions with nucleotides identical to the **St₃** marker clones in exon sequences, and also at least twelve positions in the intron sequences with complete differences from NSCs of **St₁** subgenome. In microevolutionary

terms, this means that, having formed in ancestral genotypes, this allele became fixed in some Far Eastern populations of *E. gmelinii* and was transmitted from generation to generation virtually unchanged. However, to confirm the existence of a single Far Eastern race, it is necessary to define the presence of other distinctive characteristics, including signs of reproductive isolation.

Discussion

Thus, the arrangement of subspecific taxa (varieties) in the two clustering variants was mixed, without any discernible specificity. This fact confirms the appropriateness of grouping species units within the **StY** genome group according to their phylogenetic relationships, based on previously formulated principles (Kobozeva et al., 2011, 2017; Kobozeva, Agafonov, 2015; Agafonov et al., 2021). Moreover, the **Y** subgenome showed a low level of variability, being historically young and having not accumulated evolutionarily significant transformations compared to the more ancient **St₁–St₃** subgenomes. However, within the **Y** cluster, one can see evidence of separation of several microevolutionary complexes, such as the group of taxa from the **St₂** cluster of *E. ciliaris* and the group from the **St₃** cluster of *E. nevskii*–*E. fedtschenkoi* (highlighted in dark green in Figure 1).

Notably, the **Y** subgenome cluster in the “introns” variant is less variable than in the “exons” variant. This fact contradicts the established notion of greater conservation of gene regions responsible for the synthesis of enzymatic molecules. Outside the main branch from which most of the accessions originate

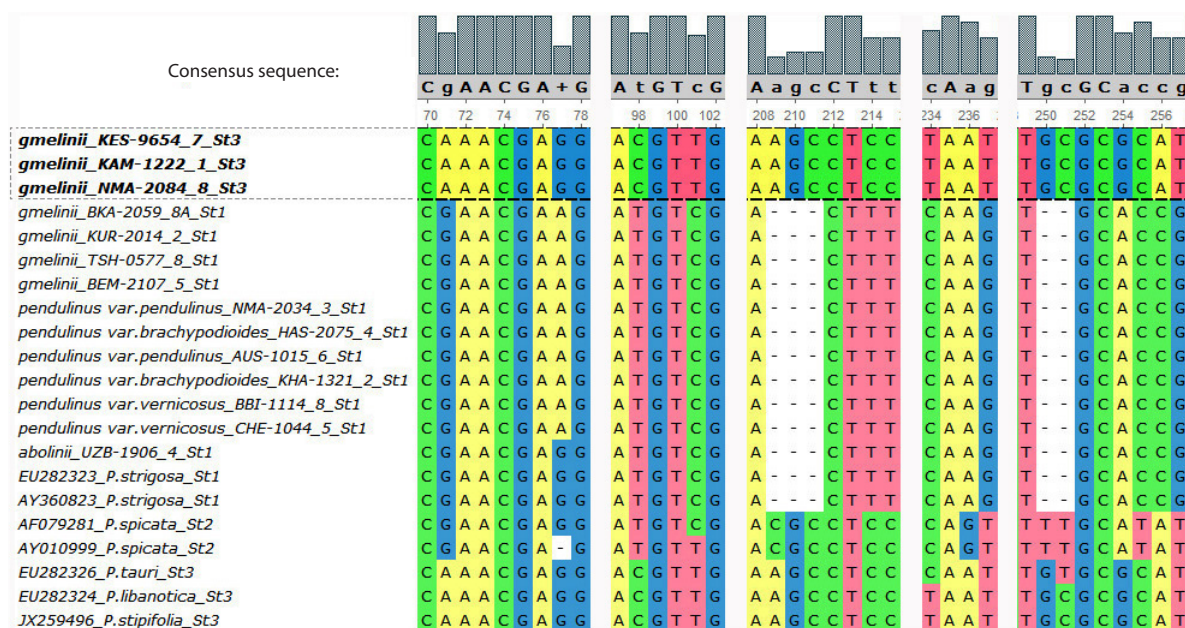


Fig. 4. Differences in the sequences of introns 9–13 in three clones of Far Eastern *E. gmelinii* accessions carrying the **St₃** subgenomes and clones of *E. gmelinii*, *E. pendulinus*, and *E. abolinii* carrying the **St₁** subgenomes.

Sequence fragments of marker clones of *Pseudoroegneria* are shown for comparison. The numbers above indicate the nucleotide positions in the aligned sequences.

lies the NSC of *E. gmelinii* – KES-9654_5_Y. The next most distant from this branch are two NSCs of *E. pendulinus*: HAS-2075 and BBI-1114. All other species assemblages are located compactly and close to each other, including *E. abolinii*, *E. caucasicus*, and *E. panormitanus*. The NSC of *E. nevskii*, AUR-1744_2_St3_Rec, was located furthest on the dendrogram, intermediate between the **Y** and **St₂** clusters, possibly recombinant. The sequence of this clone had at least 11 single substitutions identical to nucleotides in the NSCs of the **St₁**, **St₂**, and **H** subgenomes, and received its *Rec* designation due to its location on the dendrogram within the “introns” part (Fig. 2).

One notable result of the study was the location of all NSCs from the North Kazakhstan accession *E. fedtschenkoi* KSA-0935 in the **St₂** cluster instead of **St₃**. To clarify the subgenomic affiliation of the NSC *E. fedtschenkoi* KSA-0935_3_St2, aligned intron sequences in this NSC were compared with other **St₃** sequences of the *E. nevskii*–*E. fedtschenkoi* complex and reference *Pseudoroegneria* species (Fig. 5). Sequence analysis revealed that the NSC *E. fedtschenkoi* KSA-0935_3_St2 periodically carries substitutions identical to those in the NSC *P. spicata* **St₂** at positions distinct from the NSC **St₃** throughout its entire length. The possible origin of the **St₂** subgenome in *E. fedtschenkoi* can only be speculated about, given the North American range of the **St₂**-genomic species *P. spicata* and the East Asian range of the **St₂**-genomic species *E. ciliaris* s. l.

In our opinion, the group of **St_Y**-genomic species remains one of the least studied among the numerous species of the genus *Elymus* s. l. Species with the **St_H** formula have been studied better than others for two main reasons: 1) histori-

cally, species from North America and Europe have primarily attracted the attention of researchers; 2) the accumulation of knowledge about the species and structural diversity of Central and East Asia has been slower due to the vast territories and numerous mountain ranges, which create a large number of isolated ecological niches and, consequently, a multitude of local races and distinct species.

Moreover, it is believed that **St_Y**-genomic species are found in North America only as adventitious species. Reliable finds are quite rare; herbarium specimens of three species containing the **Y** subgenome have been documented: *E. ciliaris* and *E. semicostatus* (Nees ex Steud.) Melderis, as well as *E. tsukushiensis* Honda, with the genomic formula **St_{YH}** (Barkworth et al., 2007).

Currently, the number of experimental and systematic studies on the phylogeny of Asian species of the genus has increased dramatically due to the growing technological capabilities of the People’s Republic of China (Hu et al., 2013; Dong et al., 2015; Song et al., 2015; Lei et al., 2018; Liu et al., 2022; Pan et al., 2025), as well as the achievements of classical European and American scientific schools (Mason-Gamer, 2001; Leo et al., 2022, 2025; Mason-Gamer, White, 2024). In particular, the analysis of nuclear and chloroplast DNA showed that many *Elymus* species have multiple origins. The data suggest that hybridization and polyploidization were the main driving forces of evolution in increasing the biodiversity of the genus *Elymus* (Liu et al., 2006). This creates difficulties in resolving evolutionary relationships even when using high-throughput methods with a large number of genetic markers.

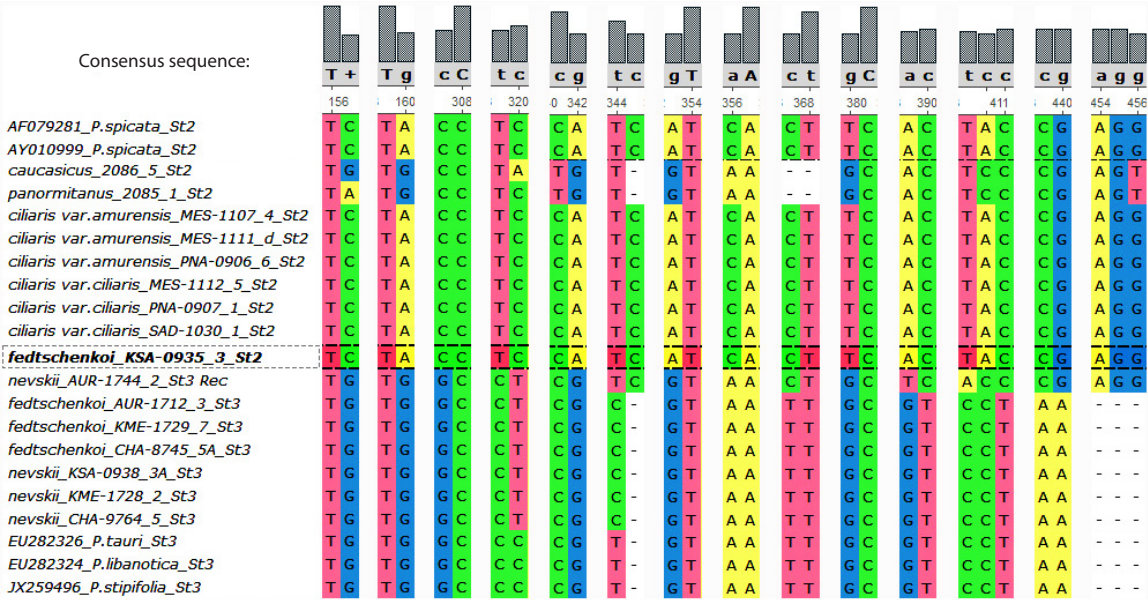


Fig. 5. The horizontal frame shows the differences in the sequences of **introns** 9–14 in the NSC of *E. fedtschenkoi* KSA-0935_3_St2 compared to those in the **St**₃ genomic accessions of the *E. nevskii*–*E. fedtschenkoi* complex and the marker sequences of *Pseudoroegneria*.

The numbers at the top indicate the nucleotide positions in the aligned sequences.

Furthermore, phenotypic plasticity, a small number of reliable morphological diagnostic characters, and a large number of taxa with the **St** subgenome in their genotypes complicate taxonomic analysis in practice (Leo et al., 2025). Nevertheless, an increasing number of new species are being described using molecular genetic data (Sha et al., 2024; Zhang et al., 2024; Alieva et al., 2025). Our results revealed signs of internal genetic divergence in *E. gmelinii*, the most widespread of all **StY**-genomic species in Russia.

The data obtained revealed a relationship between the conventional nomenclature of the **St** subgenome (**St**₁, **St**₂, and **St**₃) and a specific group of species carrying a specific subgenome. Furthermore, similarities were found in fragments of the **St**₂ and **Y** subgenome sequences, supporting the opinion of some researchers about the common origin of these subgenomes, in particular through the mechanism of recurrent hybridization (Liu et al., 2022).

Quite unexpected is the lower variability and, at the same time, relatively high species specificity among the intron sequences of the studied gene compared to the exons. This means that a detailed phylogeny of genera sharing a common and relatively ancient, yet geographically and genetically differentiated, **St** subgenome will still yield many surprises.

Conclusion

Our assembled living collection of all representatives of the **StY** genomic group from Russia allowed us to identify, at a first approximation, the features of the phylogenetic relationships between taxa. Information on reproductive relationships, heritability of diagnostic traits, and microevolutionary relations became the basis for a step-by-step revision of the currently

accepted taxonomic model of the genus *Elymus* (Tzvelev, Probatova, 2019). A broader understanding of species was used to construct the taxonomic model. The proposed model differs from the currently accepted one in that it takes into account the fundamental properties of taxa at all ranks: 1) the genomic constitution of species, as the basis for distinguishing and identifying independent genera; 2) the reproductive properties of hybrids, as an indicator of genetically determined interbreeding relationships; 3) phylogenetic relationships revealed using modern molecular and cytogenetic technologies. We consider the integrity of a species within the scope of a single microevolutionary complex, taking into account relationships with closely related taxa capable of exchanging genetic material with subsequent stabilization of sexual reproduction.

References

Agafonov A.V., Asbaganov S.V., Shabanova (Kobozeva) E.V., Morozov I.V., Bondar A.A. Genome constitution and differentiation of subgenomes in Siberian and Far Eastern endemic species of the genus *Elymus* (Poaceae) according to the sequencing of the nuclear gene *waxy*. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2019;23(7):817–826. doi 10.18699/VJ19.555

Agafonov A.V., Shabanova (Kobozeva) E.V., Asbaganov S.V., Mglinets A.V., Bogdanova V.S. Identification of genome compositions in allopolyploid species of the genus *Elymus* (Poaceae: Triticeae) in the Asian part of Russia by CAPS analysis. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2020;24(2):115–122. doi 10.18699/VJ20.606

Agafonov A.V., Shabanova (Kobozeva) E.V., Emtseva M.V., Asbaganov S.V., Dorogina O.V. Phylogenetic relationships among different morphotypes of **StY**-genomic species *Elymus ciliaris* and *E. amurensis* (Poaceae) as a unified microevolutional complex. *Botan Pacifica.* 2021;10(1):19–28. doi 10.17581/bp.2021.10101

- Agafonov A.V., Shabanova (Kobozeva) E.V., Asbaganov S.V. Micro-evolutionary relationships between the Siberian species *Elymus margaritae*, *E. komarovii*, and *E. transbaicalensis* (Poaceae) based on sequencing of the nuclear gene *GBSSI* (waxy) and hybridization. *Journal of Siberian Federal University. Biology*. 2024a;17(3):248-263 (in Russian)
- Agafonov A.V., Shabanova E.V., Emtseva M.V., Asbaganov S.V., Morozov I.V., Bondar A.A., Dorogina O.V. Phylogenetic and taxonomic relationships between morphotypes related to *Elymus caninus* (Poaceae) based on sequence of a nuclear gene *GBSSI* (waxy) and sexual hybridization. *J Syst Evol*. 2024b;62(3):520-533. doi 10.1111/jse.13006
- Alieva K.B., Peng Y., Usupbaev A., Tojibaev K.Sh., Yusupov Z., Ergashov I., Azimova D., Jiang Z. Synopsis of the genus *Elymus* (Poaceae) in Uzbekistan (Middle Asia) with a description of *Elymus uzbekistanicus* a new species from Turkestan Mts. *PhytoKeys*. 2025; 257:9-50. doi 10.3897/phytokeys.257.142950
- Barkworth M.E., Cambell J.J.N., Salomon B. *Elymus* L. In: Flora of North America. Vol. 24. New York & Oxford: Oxford University Press, 2007;288-343
- Baum B.R., Bailey L.G., Johnson D.A., Agafonov A.V. Molecular diversity of the 5S rDNA units in the *Elymus dahuricus* complex (Poaceae, Triticeae) supports the genomic constitution of **St**, **Y**, and **H** haplomes. *Can J Bot*. 2003;81(11):1091-1103. doi 10.1139/b03-102
- Dewey D.R. The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticeae. In: Gustafson J.P. (Ed.) Gene Manipulation in Plant Improvement. Stadler Genetics Symposia Series. Springer, 1984;209-279. doi 10.1007/978-1-4613-2429-4_9
- Dong Z.Z., Fan X., Sha L.N., Wang Y., Zeng J., Kang H.Y., Zhang H.Q., Wang X.L., Zhang L., Ding C.B., Yang R.W., Zhou Y.H. Phylogeny and differentiation of the **St** genome in *Elymus* L. sensu lato (Triticeae; Poaceae) based on one nuclear DNA and two chloroplast genes. *BMC Plant Biol*. 2015;15:179. doi 10.1186/s12870-015-0517-2
- Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol*. 1981;17(6):368-376. doi 10.1007/bf01734359
- Helfgott D.M., Mason-Gamer R.J. The evolution of North American *Elymus* (Triticeae, Poaceae) allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. *Syst Botany*. 2004;29(4): 850-861
- Hu Q., Yan C., Sun G. Phylogenetic analysis revealed reticulate evolution of allotetraploid *Elymus ciliaris*. *Mol Phylogenet Evol*. 2013; 69(3):805-813. doi 10.1016/j.ympev.2013.06.023
- Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980;16(2):111-120. doi 10.1007/bf01731581
- Kobozeva E.V., Agafonov A.V. Revision of the subsection *Pendulini* (Nevski) Tzelev of the genus *Elymus* L. (Poaceae). *Systematic Notes on the Materials of P.N. Krylov Herbarium of Tomsk State University*. 2015;112:22-31. doi 10.17223/20764103.112.3 (in Russian)
- Kobozeva E.V., Gerus D.E., Ovchinnikova S.V., Agafonov A.V. Taxonomic relationships between **StY** genome species *Elymus ciliaris* and *E. amurensis* (Poaceae). *Turczaninowia*. 2011;14(3):35-44 (in Russian)
- Kobozeva E.V., Asbaganov S.V., Dorogina O.V., Agafonov A.V. Variability of ISSR markers and estimation of inheritance of diagnostic characters among **StY**-genome species of *Elymus pendulinus*, *E. brachypodioides* and *E. vernicosus* (Poaceae: Triticeae). *Vavilovskii Zhurnal Genetiki i Selektii = Vavilov J Genet Breed*. 2017;21(1):135-145. doi 10.18699/VJ17.232 (in Russian)
- Koonin E.V. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol Direct*. 2006;1:1-22. doi 10.1186/1745-6150-1-22
- Lei Y.-X., Liu J., Fan X., Sha L.-N., Wang Y., Kang H.-Y., Zhou Y.-H., Zhang H.-Q. Phylogeny and molecular evolution of the *DMC1* gene in the polyploid genus *Roegneria* and its affinitive genera (Poaceae: Triticeae). *Bot J Linnean Society*. 2018;186:129-142. doi 10.1093/botlinnean/box081
- Leo J., Bengtsson T., Carlsson A.S., Brassac J., von Bothmer R. Population structure and phylogeography of *Elymus mutabilis* and its genetic relationships with *E. transbaicalensis* (Poaceae). *Nord J Bot*. 2022;4:e03520. doi 10.1111/njb.03520
- Leo J., Bengtsson T., Morales A., Carlsson A.S., von Bothmer R. Genetic structure analyses reveal multiple origins of *Elymus sensu stricto* (Poaceae). *Genet Resour Crop Evol*. 2025;72:167-185. doi 10.1007/s10722-024-01944-1
- Liu Q., Ge S., Tang H., Zhang X., Zhu G., Lu B.-R. Phylogenetic relationships in *Elymus* (Poaceae: Triticeae) based on the nuclear ribosomal internal transcribed spacer and chloroplast *trnL-F* sequences. *New Phytol*. 2006;170(2):411-420. doi 10.1111/j.1469-8137.2006.01665.x
- Liu Q.L., Liu L., Ge S., Fu L.P., Bai S.Q., Lu X., Wang Q.K., Chen W., Wang F.Y., Wang L.H., Yan X.B., Lu B.R. Endo-allopolyploidy of autopolyploids and recurrent hybridization – a possible mechanism to explain the unresolved **Y**-genome donor in polyploid *Elymus* species (Triticeae: Poaceae). *J Syst Evol*. 2022;60(2):344-360. doi 10.1111/jse.12659
- Lu B.-R., von Bothmer R. Cytological studied of dihaploid and hybrid from intergeneric cross *Elymus shandongensis* × *Triticum aestivum*. *Hereditas*. 1989;111(3):231-238. doi 10.1111/j.1601-5223.1990.tb00401.x
- Lu B.-R., von Bothmer R. Intergeneric hybridization between *Hordeum* and Asiatic *Elymus*. *Hereditas*. 1990;112(2):109-116. doi 10.1111/j.1601-5223.1990.tb00048.x
- Mason-Gamer R.J. Origin of North American *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. *Syst Bot*. 2001;26(4):757-768. doi 10.1043/0363-6445-26.4.757
- Mason-Gamer R.J. Phylogeny of a genomically diverse group of *Elymus* (Poaceae) allopolyploids reveals multiple levels of reticulation. *PLoS One*. 2013;8(11):e78449. doi 10.1371/journal.pone.0078449
- Mason-Gamer R.J., White D.M. The phylogeny of the Triticeae: resolution and phylogenetic conflict based on genome-wide nuclear loci. *Am J Bot*. 2024;111(10):e16404. doi 10.1002/ajb2.16404
- Mason-Gamer R.J., Weil C., Kellogg E.A. Granule-bound starch synthase: structure, function, and phylogenetic utility. *Mol Biol Evol*. 1998;15(12):1658-1673. doi 10.1093/oxfordjournals.molbev.a025893
- Mason-Gamer R.J., Burns M.M., Naum M. Phylogenetic relationships and reticulation among Asian *Elymus* (Poaceae) allotetraploids: analyses of three nuclear gene trees. *Mol Phylogenet Evol*. 2010a; 54(1):10-22. doi 10.1016/j.ympev.2009.10.002
- Mason-Gamer R.J., Burns M.M., Naum M. Reticulate evolutionary history of a complex group of grasses: phylogeny of *Elymus StStHH* allotetraploids based on three nuclear genes. *PLoS One*. 2010b;5(6): e10989. doi 10.1371/journal.pone.0010989
- Okito P., Mott I.W., Wu Y., Wang R.R. A **Y** genome specific STS marker in *Pseudoroegneria* and *Elymus* species (Triticeae: Gramineae). *Genome*. 2009;52(4):391-400. doi 10.1139/g09-015
- Okonechnikov K., Golosova O., Fursov M.; the UGENE team. UniPro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012; 28(8):1166-1167. doi 10.1093/bioinformatics/bts091
- Pan X., Zheng T., Zhao Y., Bao J., Fan X., Sha L., Li Y., ... Cheng Y., Zhang H., Kang H., Zhou Y., Wu D. Phylogeny and taxonomy of the polyploid species that contain **St** genome (Triticeae; Poaceae) based on four nuclear DNA and three chloroplast genes. *BMC Plant Biol*. 2025;25:183. doi 10.1186/s12870-025-06179-5
- Sakamoto S. Cytogenetic studies in the tribe Triticeae. A polyhaploid plant of *Agropyron tsukushiense* var. *transiens* Ohwi found in a state of nature. *Jpn J Genet*. 1964;39(6):393-400. doi 10.1266/jjg.39.393
- Sha L.-N., Liang X., Zhang X.-Y., Gao S., Zhang Y., Zhou Y.-H., Fan X. *Roegneria yenchiana*: a new species in the Triticeae (Poaceae) from

- the Hengduan Mountain region. *Ecol Evol.* 2024;14:e11171. doi [10.1002/ece3.11171](https://doi.org/10.1002/ece3.11171)
- Shabanova (Kobozeva) E.V., Agafonov A.V. New varieties of *Elymus ciliaris* (Poaceae: Triticeae) from the Russian Far East. *Botan Pacifica.* 2023;12(1):147-149. doi [10.17581/bp.2023.12102](https://doi.org/10.17581/bp.2023.12102)
- Song H., Nan Z.B., Tian P. Phylogenetic analysis of *Elymus* (Poaceae) in western China. *Genet Mol Res.* 2015;14(4):12228-12239. doi [10.4238/2015.October.9.11](https://doi.org/10.4238/2015.October.9.11)
- Sun G., Komatsuda T. Origin of the **Y** genome in *Elymus* and its relationship to other genomes in Triticeae based on evidence from elongation factor G (*EF-G*) gene sequences. *Mol Phylogenet Evol.* 2009;56(2):727-733. doi [10.1016/j.ympev.2010.03.037](https://doi.org/10.1016/j.ympev.2010.03.037)
- Tamura K., Stecher G., Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol.* 2021;38(7):3022-3027. doi [10.1093/molbev/msab120](https://doi.org/10.1093/molbev/msab120)
- Tan L., Wu D.D., Zhang C.B., Cheng Y.R., Sha L.N., Fan X., Kang H.Y., Wang Y., Zhang H.Q., Escudero M., Zhou Y.H. Genome constitution and evolution of *Elymus atratus* (Poaceae: Triticeae) inferred from cytogenetic and phylogenetic analysis. *Genes Genomics.* 2024;46:589-599. doi [10.1007/s13258-024-01496-9](https://doi.org/10.1007/s13258-024-01496-9)
- Tzvelev N.N., Probatova N.S. Grasses of Russia. Moscow, 2019 (in Russian)
- Yan C., Sun G., Sun D. Distinct origin of the **Y** and **St** genome in *Elymus* species: evidence from the analysis of a large sample of **St** genome species using two nuclear genes. *PLoS One.* 2011;6(10):e26853. doi [10.1371/journal.pone.0026853](https://doi.org/10.1371/journal.pone.0026853)
- Zhang Y.-C., Wei X.-X., Qin Y., Liu Y., Zhang S.-Z., Jia Z.-F., Liu W.-H. *Elymus multiramosus* (Poaceae), a new species from the north-western Qinghai-Tibetan Plateau, China. *PhytoKeys.* 2024;249:51-73. doi [10.3897/phytokeys.249.127632](https://doi.org/10.3897/phytokeys.249.127632)

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

Received September 28, 2025. Revised January 27, 2026. Accepted January 30, 2026.

doi 10.18699/vjgb-26-20

Research on genetic diversity and genomes of *Miscanthus* for optimizing their biotechnological potential

I.V. Chadaeva , D.I. Karetnikov , A.Yu. Pronozin , E.S. Starodubtseva , N.A. Omeljanchuk ,
A.V. Kochetov , D.A. Afonnikov 

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

 ichadaeva@bionet.nsc.ru

Abstract. Plants of the genus *Miscanthus* are a promising perennial energy crop, combining high biomass productivity, resistance to abiotic stress, and low agricultural technology requirements. This review summarizes recent advances in genomic and transcriptomic studies of the molecular genetic mechanisms underlying the economically valuable traits of *Miscanthus*. Data on whole-genome assemblies of key species – *M. sinensis* (*Msi*), *M. sacchariflorus* (*Msa*), *M. floridulus* (*Mfl*), and *M. lutarioriparius* (*Mlu*) – are presented using various technologies (Illumina, PacBio, Oxford Nanopore, Hi-C). Genomic studies have revealed the complex evolution of the genus, including paleoallopolyploidy, chromosomal fusions, and duplications, which accounts for the high genetic diversity of these species. Their genome assemblies at the complete chromosome level have become the basis for comparative genomics, establishing taxonomic relationships (including the recognition of *Mlu* as a subspecies of *Msa* and *Mfl* as a subtype of *Msi*), and studying synteny with related crops such as sorghum. The information on the *Miscanthus* genome allows for the complete and accurate identification of a set of genes targeting breeding for the most important biotechnological traits. At the same time, the commercial hybrid *M. × giganteus* (*M × g*) is characterized by extremely low levels of genetic polymorphism, making it vulnerable to pathogens and climate fluctuations. The integration of genetic polymorphism data, phylogeography, and functional annotation of genomes opens up opportunities for the development of new, productive, and environmentally friendly *Miscanthus* varieties through interspecific crossings, ploidy modification, and genetic engineering. These advances contribute to the optimization of the biotechnological potential of *Miscanthus* for the production of biofuels and other biomaterials and the restoration of degraded lands.

Key words: *Miscanthus*; new agricultural crop breeding; RNAseq sequencing; transcriptome; genetic diversity; stress tolerance; markers; databases; biofuels; biotechnology and industry

For citation: Chadaeva I.V., Karetnikov D.I., Pronozin A.Yu., Starodubtseva E.S., Omeljanchuk N.A., Kochetov A.V., Afonnikov D.A. Research on genetic diversity and genomes of *Miscanthus* for optimizing their biotechnological potential. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2026;30(4):593-600. doi 10.18699/vjgb-26-20

Funding. This work was supported by the state budget project FWNR-2025-0032.

Исследования генетического разнообразия и геномов мискантуса для оптимизации их биотехнологического потенциала

И.В. Чадаева , Д.И. Каретников , А.Ю. Пронозин , Э.С. Стародубцева , Н.А. Омелянчук ,
А.В. Кочетов , Д.А. Афонников 

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

 ichadaeva@bionet.nsc.ru

Аннотация. Представленный обзор посвящен интеграции геномных подходов в изучение растений рода *Miscanthus*. Мискантусы представляют собой многолетние травы, перспективные для создания энергетических культур, характеризующихся высокой продуктивностью биомассы, устойчивостью к абиотическим стрессам и низкими требованиями к агротехнологиям. Рассматриваются методы анализа генетического разнообразия, полногеномные сборки, а также современные достижения в расшифровке молекулярно-генетических механизмов, определяющих продуктивность и стрессоустойчивость, с целью применения их в селекции. Приведены данные по полногеномным сборкам ключевых видов – *M. sinensis* (*Msi*), *M. sacchariflorus* (*Msa*), *M. floridulus* (*Mfl*) и *M. lutarioriparius* (*Mlu*) – с использованием различных технологий (Illumina, PacBio, Oxford Nanopore, Hi-C). Геномные исследования выявили сложную эволюцию рода, включающую палеоаллополиплоидию, хромосомные слияния и дупликации, что обуславливает высокое генетическое разнообразие у этих видов. Полные хромосомные сборки геномов их

стали основой для сравнительной геномики, установления таксономических отношений (включая признание *Mlu* подвидом *Msa*, а *Mfl* – подтипом *Msi*) и изучения синтении с родственными культурами, такими как сорго. Полученная информация о геноме мискантусов позволяет с высокой полнотой и точностью идентифицировать набор генов, целевых для селекции по наиболее важным биотехнологическим признакам. В то же время коммерческий гибрид *M. × giganteus* (*M × g*) характеризуется крайне низким уровнем генетического полиморфизма, что делает его уязвимым к патогенам и климатическим колебаниям. Совокупность данных о генетическом полиморфизме, филогеографии и функциональной аннотации геномов открывает возможности для создания новых, продуктивных и экологически безопасных сортов мискантуса через межвидовые скрещивания, модификацию плоидности и геномную инженерию. Эти достижения способствуют оптимизации биотехнологического потенциала мискантусов в производстве биотоплива и других биоматериалов, а также восстановлению деградированных земель.

Ключевые слова: мискантус (*Miscanthus*); селекция новой сельскохозяйственной культуры; секвенирование RNAseq; транскриптомы; генетическое разнообразие; устойчивость к факторам стресса; маркеры; базы данных; биотопливо; биотехнологии и промышленность

Introduction

The genus *Miscanthus* (silvergrass) comprises tall perennial grasses with short or long rhizomes and a rigid rachis and belong to the family Poaceae, specifically to the tribe Andropogoneae. *Miscanthus* species, are native to the temperate monsoon climates of East Asia (the Russian Far East) (Dorogina et al., 2025), as well as subtropical and tropical regions of East Asia, Southeast Asia, and the Pacific Islands, with several species extending to tropical Africa (Brosse et al., 2012; Wang et al., 2021). The highest species diversification is found in East Asia, particularly in China and Japan (Hodkinson et al., 2015).

Members of the genus *Miscanthus* exhibit high genetic diversity (Clark et al., 2019) and heterogeneity in genome structural organization, with ploidy levels ranging from 2 to 6 and variable chromosome numbers (for example, in the species reviewed here – *M. sacchariflorus* (*Msa*), *M. lutarioriparius* (*Mlu*), *M. floridulus* (*Mfl*), and *M. sinensis* (*Msi*) – the chromosome number is 19, whereas in Asian species such as *M. fuscus*, *M. nepalensis*, and *M. nudipes*, and African species such as *M. ecklonii*, *M. junceus*, *M. sorghum*, and *M. violaceus*, chromosome numbers may be 10 or 15) (Hodkinson et al., 2015). *Msi* is a genetic diploid ($2n = 2x = 38$) with a genome size of $1C = 2.4–2.6$ Gb (Rayburn et al., 2009); the closely related *Msa* occurs in both diploid ($2n = 2x = 38$) and tetraploid ($2n = 4x = 76$) forms (Mitros et al., 2020). The high genetic diversity within the genus *Miscanthus* is further enhanced by their ability to readily form hybrids, even between species of different ploidy levels (Clark et al., 2015, 2019).

Miscanthus species are among the most promising crops for biofuel production, owing to several key traits relevant to industrial applications: high photosynthetic efficiency, effective nutrient and water use, high biomass yield (Wang et al., 2021), and broad adaptability to diverse climatic conditions and soil types (Clifton-Brown et al., 2017). These characteristics enable the use of *Miscanthus* as an important industrial crop across various business activities, including construction, manufacturing, and agriculture (Mironova et al., 2023) (see the Figure). For bioenergy and biomass feedstock purposes, the most important species are *Msa* and *Msi*, as well as their interspecific hybrid *M. × giganteus* (*M × g*), classified as a nothospecies (Trieu et al., 2022).

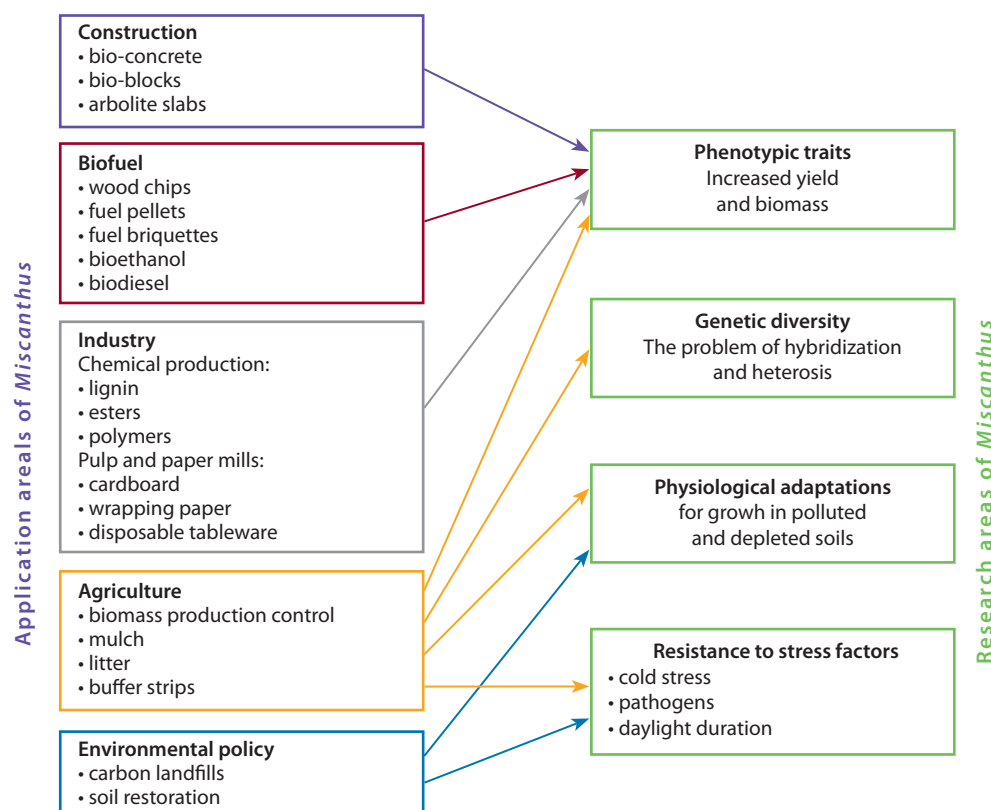
In construction, *Miscanthus* serves as a source of biofiber which, when added to concrete, improves its mechanical properties as well as its sound and thermally insulating performance (Chen Y.X., Yu, 2024). The biomass yield of *Miscanthus* is comparable to that of maize (Sanford et al., 2016), higher than that of several other crops – including switchgrass (*Panicum virgatum*), poplar (*Populus* sp.), and willow (*Salix* sp.) – and exceeded only by eucalyptus (*Eucalyptus* sp.) (Li W. et al., 2018).

Miscanthus biomass comprises cellulose, hemicellulose, lignin, and various associated compounds (Schäfer et al., 2019). This composition underpins the active use of *Miscanthus* as a raw material source for the pulp and paper industry and for the production of eco-friendly packaging (Cappelletto et al., 2000). Furthermore, *Miscanthus* biomass serves as a feedstock for producing a wide range of reagents and polymers, including cellulose derivatives, ethanol, 5-hydroxymethylfurfural, furfural, and various phenolic compounds (Shavyrkina et al., 2023) (see the Figure).

In agriculture, *Miscanthus* is used as a source of biochar to improve soil condition and enhance fertility (Nagel et al., 2019). *Miscanthus* straw treated with *Trichoderma* can serve as a peat substitute for strawberry cultivation (Debode et al., 2018). It is also employed as a substrate for the anaerobic digestion of cattle and swine manure in biogas fermenters (Jury et al., 2022) and as bedding for cattle, which reduces bedding costs (Van Weyenberg et al., 2016).

One of the most important characteristics of *Miscanthus* is its low requirement for cultivation conditions. This enables the use of this crop for soil remediation (Grzegórska et al., 2023), particularly in areas contaminated with pollutants such as organochlorine pesticides (Mamirova et al., 2021), heavy metal salts (Andrić et al., 2025), and polycyclic aromatic hydrocarbons (Técher et al., 2012). The ability of *Miscanthus* to accumulate toxins in its roots is currently under close scientific scrutiny and is being actively applied in practice (Nsanganwimana et al., 2021).

The primary objectives in *Miscanthus* breeding can be summarized as follows: 1) overcoming the low genetic diversity of commercial hybrids (such as *M × g*); 2) preventing invasiveness through reproductive isolation (sterility) and specific containment measures for vegetatively propagated, long-rhizome



Miscanthus in human economic activities and key research areas for harnessing its genetic potential as a novel biotechnological crop.

Miscanthus species, such as *Msa*; 3) developing new hybrids with novel beneficial traits through interspecific crosses (e. g., *Msi* × *Msa*, *Msi* × *Mfl*); and 4) leveraging modern biotechnologies, including genetic modification and *in vitro* regeneration.

The efficiency of classical breeding methods can be substantially enhanced through the application of modern genetic approaches and omics technologies (Yang et al., 2021). In this review, we examine the use of genomic technologies to study genetic diversity in *Miscanthus*, reconstruct genomes, and identify target genes for plant breeding programs.

Studying genetic diversity in *Miscanthus*

Studying diversity primarily helps to link genetic relatedness between various taxa of this genus with their geographical distribution. The main goal of such studies is to identify patterns of distribution of various *Miscanthus* species, identify centers of origin and genetic diversity, and determine the influence of climatic factors on evolutionary mechanisms of speciation. One of the main questions is also establishing the relationships between species of the genus *Miscanthus*, since despite intensive research, their taxonomic position and the evolution of species *Msi*, *Mfl*, *Msa*, and *Mlu* remain not fully understood (Sun et al., 2010; Sang, Zhu, 2011; Hodkinson et al., 2015).

In the study (Cichorz et al., 2014), an analysis of genetic diversity of samples of four *Miscanthus* species from Poland was conducted: *M* × *g*, *Msi*, *Msa*, and *Mfl*. ISSR (inter-simple sequence repeats) and RAPD (random amplified polymorphic

DNA) markers were used for the analysis. For *M* × *g*, low genetic diversity was shown, consistent with its origin from a single clone. In contrast, *Msi* is characterized by high genetic diversity, which is confirmed in the study of *Msi* populations from Southwest China using SRAP (sequence-related amplified polymorphism) markers (Nie et al., 2014).

A large-scale study of *Msi* populations in China based on the analysis of 459 samples from various geographical regions of the country using SSR markers revealed high genetic diversity, indicating the great potential of *Msi* as a genetic resource for breeding (Zhao et al., 2013). Interestingly, *Msi* chloroplast DNA demonstrates comparatively low genetic variability, which is explained by greater susceptibility to random drift and a smaller effective population size for chloroplast DNA compared to nuclear DNA (Yan et al., 2015).

In the study (Li S.-S. et al., 2019), 100 populations of four *Miscanthus* species, *Msi*, *Mlu*, *Mfl*, and *Msa*, growing in China from northern to southern regions were investigated. Two genetically homogeneous groups were identified. The first included plants of species *Msa* and *Mlu*, the second – *Mfl* and *Msi*. Based on genetic diversity data, possible routes of *Miscanthus* population evolution were modeled. Results showed that species from the first group underwent a range contraction during the Last Glacial Maximum (26–19 thousand years ago) and then its gradual expansion, whereas species of the second group have gradual range expansion from the interglacial period (129–116 thousand years ago) to the present, expanding to the southern China regions. The authors of this study

proposed a hypothesis on the origin of *Mfl*, according to which *Mfl* species originated from *Msi* ancestors in southeastern China (Li S.-S. et al., 2019). This is confirmed by the shared geographical distribution of these species, as well as their positions on the phylogenetic tree, where the *Mfl* population cluster is located within the *Msi* cluster.

A large-scale research was dedicated to investigating the *Msa* and *Mlu* population, which included 764 samples from Russia, China, South Korea, and Japan (Clark et al., 2019). The aim of the work was to study the diversity, origin, and evolution of representatives of these *Miscanthus* species. DNA analysis was based on RAD-seq (restriction site-associated DNA sequencing) technology, and additionally, ploidy analysis was performed using flow cytometry. As a result, six genetic groups were identified: three groups were represented by diploid varieties growing predominantly in Northeast China, Korea, and Russia; the remaining three groups turned out to be tetraploids growing in Northern China, Korea, and Japan.

According to the researchers, the modern *Msa* population formed during the last glacial period from an ancestral population, whose habitat was located in Eastern China, in an area now covered by the Yellow and East China Seas. Also, this work showed that polyploidization events occurred independently for the Japanese and mainland *Msa* groups. Analysis of plastid haplotypes revealed 56 unique variants between *Msa* and *Msi*. Nine of them are intermediate between the most common haplotypes of the compared species (Clark et al., 2019).

The results of this work also confirmed the assumption that *Mlu* is a subspecies of *Msa*, originating in the Yangtze River basin area, and not a separate *Miscanthus* species. The study (Clark et al., 2019) demonstrated higher genetic diversity of the species *Msa* compared to *Msi*, indicating a difference in the population history of these two species.

In the study (Chen et al., 2022), using SLAF-seq (specific-locus amplified fragment sequencing) technology, genetic differences between plants belonging to species *Msi*, *Mlu*, and *Msa* were investigated. Analyzing 50 samples of each of these three species, the authors showed that the genetic distance between *Mlu* and *Msi* turned out to be smaller than the genetic differences within the *Mlu* population itself. Their results are consistent with the results of the study L.V. Clark et al. (2019).

A large-scale population analysis of various *Miscanthus* samples, including about 2,800 samples from China, Korea, Russia, and Japan, was conducted using GBS (genotyping-by-sequencing) in the study T. Mitros et al. (2020) along with the assembly of the *Msi* genome. Results showed high genetic diversity among representatives of this genus. The analysis also allowed clarifying taxonomic relationships within it. According to this work, the populations of species *Msi* split into two large clusters, one of which includes representatives growing in Japan, and the other – in China and Korea.

Plants of species *Msa* represent a separate distinct group, which demonstrate less diversity. Representatives of the $M \times g$ population occupied an intermediate position between *Msi* and *Msa*, combining genetic components of both species. The genetic diversity of *Msi* significantly exceeds that of a number of other species, *M. transmorrissonensis* and *Mfl*. This

allowed the authors to conclude that the latter two taxa can be considered as subtypes of *Msi*, belonging to the subpopulation from Japan (Mitros et al., 2020).

In the study (Zhang et al., 2021), aimed at sequencing and analyzing the *Mfl* genome, genetic diversity was evaluated for 75 accessions of *Mfl*, $M \times g$, *Msa*, *Mlu*, *Msi* and a number of hybrid plants obtained by crossing these species. Analysis of these populations showed that *Mfl* and *Msi* represent separate groups genetically distant from other representatives of the genus, and plants *Msa* and *Mlu* form a separate cluster, confirming the status of *Mlu* as a subspecies of *Msa*.

Thus, analysis of genetic diversity based on the use of whole-genome sequencing of DNA fragments allowed establishing the population history of *Miscanthus* species and clarifying taxonomic relationships within representatives of this genus: classifying the taxon *Mlu* as a subspecies of *Msa*, and taxa *M. transmorrissonensis* and *Mfl* as subtypes of *Msi*. The wide genetic variability of *Miscanthus* genus revealed makes it possible to conduct targeted breeding for agricultural and industrial purposes (Hodkinson et al., 2015; Trieu et al., 2022).

Genetic and geographical data serve as a basis for creating breeding programs aimed at developing varieties adapted to different climatic zones. In a recent study (Dorogina et al., 2025), an investigation was conducted on the possibility of cultivating *Msi* samples from monsoon climate regions under condition of the forest-steppe zone in Western Siberia. During the trials, two samples, S1 and S2, were selected, which formed viable seeds during the short growing season characteristic of the Siberian continental climate, from which two generations, G1 and G2, were obtained. These samples also had accelerated rates of seasonal development and a more compact habitus. They formed less vegetative mass and began to form generative organs earlier than typical representatives of *Msi*.

Study of genetic variability in generation G1 showed complete uniformity of genotypes. In generation G2, on the contrary, diversity was observed with the identification of five different genotype variants. As the authors state, the phenotypic and genetic variability they discovered in *Msi* will allow selecting forms with various economically valuable traits for further genetic improvement and development of a variety with desired traits.

Genomic studies of genus *Miscanthus*

To date, several whole-genome assemblies have been generated for four *Miscanthus* species: *Msa* (De Vega et al., 2021), *Mlu* (Miao et al., 2021), *Mfl* (Zhang et al., 2021), and *Msi* (Mitros et al., 2020), differing in quality and completeness (see the Table).

The Table shows the general assembly parameters: sequencing technologies used, assembly size, N50 metric, guanine and cytosine content (GC content), genome and transcriptome quality assessment using BUSCO (Benchmarking Universal Single-Copy Orthologs) (Simão et al., 2015), as well as the proportion of mobile elements in the genomic sequence. Assemblies were performed using different sequencing technologies and assembly methods, which affects their accuracy and suitability for downstream analysis. Below, we present the results of genomic studies for these four *Miscanthus* species.

The *Msa* genome

Samples from the diploid cultivar Robustus 297 were used for sequencing of the *Msa* genome (De Vega et al., 2021). Short-read libraries with small inserts generated using Illumina HiSeq 2500 platform were employed (~5.86 billion reads, genome coverage ~50×). For scaffold assembly, a paired-end read library with a read length of 150 bp and a large insert size (7 kb) was used, totaling 141 million read pairs.

The primary contig assembly (Msac_v2) was performed using the ABySS software (Simpson et al., 2009), yielding 17 million contigs with a total length of 3.27 Gb. As a result of scaffolding, ~589 thousand sequences were obtained with a total length of 2.54 Gb and an N50 metric of approximately 10.2 kb (see the Table). To improve assembly quality, sequences shorter than 2 kb were filtered out, and further scaffolding was performed for the remaining longer scaffolds using paired libraries from *Msi* data from the study (Mitros et al., 2020). Ultimately, the “Msac_v3” assembly was obtained with a length of 2.074 Gb (see the Table), consisting of 137,916 scaffolds, with the N50 increasing to 25.6 kb.

The final assembly (Msac_v3) was achieved by ordering scaffolds into chromosomes based on alignment with *Msi* chromosomes. Genome annotation was performed using the AUGUSTUS software (Stanke et al., 2004). The final annotation covers approximately 81 thousand genes for the Msac_v2 assembly and approximately 68 thousand for the Msac_v3 assembly. The proportion of mobile elements for the obtained assembly was 38.81 % (see the Table). For the *Msa* genome assembly Msac_v3, completeness based on BUSCO analysis was 55.5–59.8 % (De Vega et al., 2021).

The *Mlu* genome

The *Mlu* genome was assembled using multiple sequencing and assembly technologies, which ensured high-quality representation at the chromosomal level (Miao et al., 2021). The Oxford Nanopore PromethION platform was applied for sequencing yielding long DNA reads with a total size of 307.71 Gb and N50 parameter of 32.21 kb. Additionally, three Illumina libraries with different insert sizes were used, providing reads totaling 205.74 Gb, of which 172.52 Gb remained after quality filtering. Hi-C (High-throughput chromosome conformation capture) technology was implemented to order the assembled genomic sequence fragments into chromosomes.

The primary genome assembly was performed using SMARTdenovo software (Liu et al., 2021) and based solely on long Nanopore reads. The size of the primary assembly was 2.25 Gb with an N50 parameter of 1.71 Mb. At the final

assembly stage, reads from all libraries and Hi-C data were incorporated, resulting in 919 scaffolds with a total size of approximately 2.074 Gb (see the Table), of which 94.3 % (approximately 1.96 Gb) were placed and oriented on 19 pseudochromosomes ranging in size from 61.78 to 150.81 Mb.

Comparison of contig sequences with BAC (bacterial artificial chromosome) sequences obtained by Sanger sequencing showed very high similarity at the 99 % level. Alignment of Illumina reads to the assembled genome indicated its high completeness and accuracy (99.8 %). The GC content of the genome was approximately 45.5 %, and mobile elements (MEs) accounted for 64.4 % of the genome (Miao et al., 2021). The total number of predicted genes was 68,328. BUSCO analysis of protein-coding gene content demonstrated a completeness of 97.4 % (see the Table). The authors performed a synteny analysis of the obtained *Mlu* assembly with the sorghum genome, which allowed the identification of conserved genomic regions.

The *Mfl* genome

For the *Mfl* genome assembly, sequencing was performed using Illumina, PacBio, and 10x Genomics platforms, and a Hi-C library was used for scaffold ordering (Zhang et al., 2021). The final assembly comprises 19 pseudochromosomes covering approximately 2.44 Gb (91 % of the genome), with a total genome size of 2.700 Gb and an N50 of 143 Mb (see the Table). The quality of the final assembly was confirmed by high coverage with short reads and high BUSCO completeness scores for protein-coding genes (96 %).

The *Mfl* genome is characterized by a high proportion of mobile elements (63.6 %, predominantly LTR retrotransposons). Genome analysis identified 76,913 genes, of which 71,637 (91.1 %) were functionally annotated. Comparison of *Mfl* with other crops, *Zea mays*, *Oryza sativa*, *Sorghum bicolor*, and *S. spontaneum*, revealed that a set of more than 13 thousand gene families was common to these cereal genomes. In turn, 2,219 gene families were unique to *Mfl*.

The *Msi* genome

The *Msi* genome was assembled to the chromosome level (Mitros et al., 2020). The total size of the diploid *Msi* genome is approximately 2.4–2.6 Gb. The *Msi* genome assembly was performed using an integrated approach involving multiple sequencing methods. The starting material was a dihaploid (DH1) *Msi* line, which is homozygous for genetic markers. Sequencing was performed using Illumina HiSeq 2000 technology, as well as Nextera long-insert paired-end reads and fosmid libraries on the Illumina MiSeq platform. To improve

Comparative characteristics of genome assemblies in four *Miscanthus* species

Species	Sequencing technologies	Assembly size, billion bp	N50, billion bp	GC-content	Number of protein-coding genes	BUSCO, %	Proportion of mobile elements, %
<i>Msa</i> (“Msac_v3”)	Illumina, Nextera	2.074	0.0256	46.0	68 578	59.8	38.81
<i>Mlu</i>	PromethION, Illumina, Hi-C	2.074	113.5	45.5	68 328	97.4	64.39
<i>Mfl</i>	PacBio, Illumina, 10x Genomics, Hi-C	2.700	143	45.5	76 913	96.0	63.6
<i>Msi</i>	Illumina, Nextera	2.079	88.5	45.8	67 967	97.6	72.4

the genome assembly, read libraries obtained using Hi-C technology were employed, enabling the establishment of three-dimensional contacts between chromosome fragments within the nucleus. Ultimately, a sequence of 2.079 Gb was obtained for the nuclear genome, with an assembly N50 parameter of 88.5 Mb (see the Table). In addition to the nuclear genome, the assembly of chloroplast and mitochondrial genomes was performed within this study.

For protein-coding gene prediction, bioinformatic *ab initio* methods and RNA-seq transcriptome library data were used. In this study, transcriptome data included 57 growth time points for three *Msi* DH1 tissues; additionally, transcriptome data for the $M \times g$ hybrid were used. In total, this allowed the prediction of 67,967 protein-coding genes (94.3 % localized on chromosomes), of which more than 50,000 were confirmed by RNA-sequencing data. The proportion of repetitive DNA for the obtained assembly was 72.4 % (see the Table). Assessment of annotation completeness using the BUSCO method (Simão et al., 2015) showed that 97.6 % of *Embryophyta* genes are represented completely in the assembly (see the Table), another 1 % are fragmented, and 64.3 % of genes are duplicated, which is consistent with the paleotetraploid nature of representatives of the genus *Miscanthus* (Mitros et al., 2020).

The available genome assemblies of *Miscanthus* species have enabled in-depth studies of the evolution of their genome structures, allowing the identification of conserved and unique fragments through comparisons of genome sequences both with other species and among themselves.

Comparative analysis of the sequenced *Msi* and sorghum genomes (Mitros et al., 2020) revealed that each sorghum chromosome corresponds to a pair of *Msi* chromosomes, with the exception of a single *Miscanthus* chromosome. Specifically, this chromosome corresponds to the fusion of two sorghum chromosomes. Analysis of the *Msi* genome revealed synteny between *Miscanthus* and *S. bicolor* genome sequences. The *Msi* DNA appears to contain DNA similar to that of the ancestor of modern sorghum, the amount of which was doubled during chromosome fusion events; consequently, the genus *Miscanthus* comprises paleoallopolyploids (Ma et al., 2012; Mitros et al., 2020; Trieu et al., 2022).

Thus, during their evolution prior to diversification, but after diverging from the closely related sugarcane clade, species of the genus *Miscanthus* underwent a stage of ancestral genome tetraploidization and subsequent chromosome fusion (Swaminathan et al., 2012). As a result of diploidization, *Miscanthus* spp. possess two homeologous chromosomes that are highly syntenic, potentially providing functional redundancy between homeologous gene pairs (Trieu et al., 2022).

Studying the structure of target genes based on whole-genome assemblies

Whole-genome assembly enables complete and highly accurate identification of the set of protein-coding genes in *Miscanthus*, including genes that are targets for breeding based on the most important biotechnological traits.

In the study dedicated to the whole-genome assembly of *Msi* (Mitros et al., 2020), an analysis of seasonal dynamics of gene expression in leaves, stems, and rhizomes of plants was con-

ducted. The authors showed that variations in gene expression levels between different organs exceed seasonal variations. Nevertheless, for genes within a single tissue, seasonal changes significantly influenced expression levels. Expression data allowed the identification of several functional gene clusters related to important functions: nitrogen metabolism; amino acid metabolism; hormonal regulation and signaling; genes associated with pathogen response; and genes involved in starch and sucrose biosynthesis. Among these gene groups, key genes were highlighted based on expression data. A network of seasonal nutrient uptake and utilization was reconstructed.

In the study on the *Msa* genome assembly (De Vega et al., 2021), reconstructed sequences of protein-coding genes enabled the search for orthologs in *Msi*, as well as in other plant species (foxtail millet *Setaria italica*, sorghum *S. bicolor*, maize *Z. mays*, switchgrass *P. virgatum*).

In the study on the *Mfl* genome reconstruction (Zhang et al., 2021), considerable attention was devoted to investigating the structure of cellulose biosynthesis genes – cellulose synthases. Within this group of enzymes, based on ortholog searches considering synteny of genomes of both *Miscanthus* and rice, sorghum, and sugarcane, several functional clusters were identified, whose origin is associated with duplication and expansion events of this family in *Miscanthus*.

In the study on the *Mlu* genome assembly (Miao et al., 2021), the focus of protein-coding gene research was on functional annotation analysis and patterns of expansion and contraction of gene families. The proportion of gene families subject to expansion exceeded the number of families with gene losses by nearly threefold (9,509 versus 3,228). A total of 211 gene families demonstrated rapid expansion in *Mlu*; these included pathogen resistance genes (containing the NB-ARC domain), genes associated with xylanase inhibition function, salt stress response genes, cytochromes P450, terpene synthases, and several others.

A separate analysis was dedicated to genes of the C4 photosynthesis pathway. These were identified based on orthology with sorghum genes, totaling 55. Most of these genes in *Mlu* underwent expansion through proximal, tandem, and segmental genomic duplications. As a result, their number was almost twice that found in sorghum.

Conclusion

The high biotechnological potential of perennial grasses of the genus *Miscanthus* for human economic activities, their broad adaptive capabilities, and their ability to maintain high biomass productivity under unfavorable abiotic and biotic stress conditions – all of these factors make them a valuable target for breeding and industrial applications. The adaptive potential of *Miscanthus* species (*Msi*, *Msa*, *Mlu*, and *Mfl*) is underpinned by conserved and species-specific features of genome organization. At the same time, the hybrid $M \times g$, derived from crossing *Msa* and *Msi*, is characterized by seed sterility and extremely low genetic diversity, rendering it vulnerable to diseases. To overcome this limitation in *Miscanthus* breeding, working with diverse hybrids can be recommended.

Modern technologies for whole-genome genotyping, sequencing, genome assembly, and bioinformatics provide

opportunities for in-depth investigation of genetic diversity within the genus *Miscanthus*, as well as genome structure and gene repertoire. This enables the identification of genetic control mechanisms underlying traits that are key to the industrial potential of this crop. The complete gene set of *Miscanthus* allows reconstruction of gene networks and metabolic pathways, and study of the impact of genetic variation on the functioning of key molecular and physiological processes in the plant. Such research, utilizing an expanded set of sequenced genomes, will form the foundation for targeted *Miscanthus* breeding programs.

References

- Andrić S., Knežević G., Maletić S., Rončević S., Isakovski M.K., Zermeski T., Beljin J. An experimental assessment of *Miscanthus* × *giganteus* for landfill leachate treatment: a case study of the grebača landfill in obrenovac. *Processes*. 2025;13(3):768. doi 10.3390/pr13030768
- Brosse N., Dufour A., Meng X.Z., Sun Q.N., Ragauskas A. *Miscanthus*: a fast-growing crop for biofuels and chemicals production. *Biofuels Bioprod Biorefin*. 2012;6(5):580-598. doi 10.1002/bbb.1353
- Cappelletto P., Mongardini F., Barberi B., Sannibale M., Brizzi M., Pignatelli V. Papermaking pulps from the fibrous fraction of *Miscanthus* × *giganteus*. *Ind Crop Prod*. 2000;11(2-3):205-210. doi 10.1016/S0926-6690(99)00051-5
- Chen Y.X., Yu Q. Surface modification of miscanthus fiber with hydrophobic silica aerogel for high performance bio-lightweight concrete. *Constr Build Mater*. 2024;411:134478. doi 10.1016/j.conbuildmat.2023.134478
- Chen Z., He Y., Iqbal Y., Shi Y., Huang H., Yi Z. Investigation of genetic relationships within three *Miscanthus* species using SNP markers identified with SLAF-seq. *BMC Genomics*. 2022;23(1):43. doi 10.1186/s12864-021-08277-8
- Cichorz S., Goška M., Litwiniec A. *Miscanthus*: genetic diversity and genotype identification using ISSR and RAPD markers. *Mol Biotechnol*. 2014;56(10):911-924. doi 10.1007/s12033-014-9770-0
- Clark L.V., Stewart J.R., Nishiwaki A., Toma Y., Kjeldsen J.B., Jørgensen U., Zhao H., Peng J., Yoo J.H., Heo K., Yu C.Y., Yamada T., Sacks E.J. Genetic structure of *Miscanthus sinensis* and *Miscanthus sacchariflorus* in Japan indicates a gradient of bidirectional but asymmetric introgression. *J Exp Bot*. 2015;66(14):4213-4225. doi 10.1093/jxb/eru511
- Clark L.V., Jin X., Petersen K.K., Anzoua K.G., Bagmet L., Chebukin P., Deuter M., ... Yamada T., Yoo J.H., Yu C.Y., Long S.P., Sacks E.J. Population structure of *Miscanthus sacchariflorus* reveals two major polyploidization events, tetraploid-mediated unidirectional introgression from diploid *M. sinensis*, and diversity centred around the Yellow Sea. *Ann Bot*. 2019;124(4):731-748. doi 10.1093/aob/mcy161
- Clifton-Brown J., Hastings A., Mos M., McCalmont J.P., Ashman C., Awtycarroll D., Flavell R. Progress in upscaling *Miscanthus* biomass production for the European bio-economy with seed-based hybrids. *Glob Change Biol Bioenergy*. 2017;9(1):6-17. doi 10.1111/gcbb.12357
- Debode J., De Tender C., Cremelie P., Lee A.S., Kyndt T., Muylle H., De Swaef T., Vandecasteele B. Trichoderma-inoculated miscanthus straw can replace peat in strawberry cultivation, with beneficial effects on disease control. *Front Plant Sci*. 2018;9:213. doi 10.3389/fpls.2018.00213
- De Vega J., Donnison I., Dyer S., Farrar K. Draft genome assembly of the biofuel grass crop *Miscanthus sacchariflorus* [version 1; peer review: 2 approved]. *F1000Res*. 2021;10:29. doi 10.12688/f1000research.44714.1
- Dorogina O.V., Kuban I.N., Zueva G.A., Zhmud E.V., Vasilyeva O.Yu. Morphological variability and genetic structure of *Miscanthus sinensis* (Poaceae) cultivated in the forest-steppe of Western Siberia. *Vavilov J Genet Breed*. 2025;29(2):210-218. doi 10.18699/vjgb-25-24
- Grzegórska A., Czaplicka N., Antonkiewicz J., Rybarczyk P., Baran A., Dobrzyński K., Zabrocki D., Rogala A. Remediation of soils on municipal rendering plant territories using *Miscanthus* × *giganteus*. *Environ Sci Pollut Res Int*. 2023;30(9):22305-22318. doi 10.1007/s11356-022-23724-z
- Hodkinson T.R., Klaas M., Jones M.B., Prickett R., Barth S. *Miscanthus*: a case study for the utilization of natural genetic variation. *Plant Genet Resour*. 2015;13(3):219-237. doi 10.1017/S147926211400094X
- Jury C., Thomas H.L., Carrère H. Life cycle assessment of two alkaline pretreatments of sorghum and miscanthus and of their batch co-digestion with cow manure. *BioEnergy Res*. 2022;15:810-833. doi 10.1007/s12155-021-10369-y
- Li S.-S., Zhou H.-F., Chen W.-L., Yan J., Cai Z., Wei R.-X., Chen C.-H., Han B., Li J.-Q., Sang T., Ge S. Population genetics and evolutionary history of *Miscanthus* species in China. *J Syst Evol*. 2019;57(5):530-542. doi 10.1111/jse.12497
- Li W., Ciais P., Makowski D., Peng S. A global yield dataset for major lignocellulosic bioenergy crops based on field measurements. *Sci Data*. 2018;5:180169. doi 10.1038/sdata.2018.169
- Liu H., Wu S., Li A., Ruan J. SMARTdenovo: a de novo assembler using long noisy reads. *GigaByte*. 2021;2021:gigabyte15. doi 10.46471/gigabyte.15
- Ma X.-F., Jensen E., Alexandrov N., Troukhan M., Zhang L., Thomas-Jones S., Farrar K., Clifton-Brown J., Donnison I., Swaller T., Flavell R. High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid *Miscanthus sinensis*. *PLoS One*. 2012;7(3):e33821. doi 10.1371/journal.pone.0033821
- Mamirova A., Pidlisnyuk V., Amirbekov A., Ševců A., Nurzhanova A. Phytoremediation potential of *Miscanthus sinensis* And. in organochlorine pesticides contaminated soil amended by Tween 20 and Activated carbon. *Environ Sci Pollut Res Int*. 2021;28(13):16092-16106. doi 10.1007/s11356-020-11609-y
- Miao J., Feng Q., Li Y., Zhao Q., Zhou C., Lu H., Fan D., ... Huang T., Li L., Huang X., Sang T., Han B. Chromosome-scale assembly and analysis of biomass crop *Miscanthus lutarioriparius* genome. *Nat Commun*. 2021;12(1):2458. doi 10.1038/s41467-021-22738-4
- Mironova G.F., Budaeva V.V., Skiba E.A., Gismatulina Y.A., Kashcheyeva E.I., Sakovich G.V. Recent advances in *Miscanthus* macromolecule conversion: a brief overview. *Int J Mol Sci*. 2023;24(16):13001. doi 10.3390/ijms241613001
- Mitros T., Session A.M., James B.T., Wu G.A., Belaffif M.B., Clark L.V., Shu S., ... Hudson M.E., Sacks E.J., Moose S.P., Swaminathan K., Rokhsar D.S. Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. *Nat Commun*. 2020;11(1):5442. doi 10.1038/s41467-020-18923-6
- Nagel K., Hoilett N.O., Mottaleb M.A., Meziani M.J., Wistrom J., Belamy M. Physicochemical characteristics of biochars derived from corn, hardwood, miscanthus, and horse manure biomasses. *Commun Soil Sci Plant Anal*. 2019;50(8):987-1002. doi 10.1080/00103624.2019.1594881
- Nie G., Zhang X.Q., Huang L.K., Xu W.Z., Wang J.P., Zhang Y.W., Ma X., Yan Y.H., Yan H.D. Genetic variability and population structure of the potential bioenergy crop *Miscanthus sinensis* (Poaceae) in southwest china based on SRAP markers. *Molecules*. 2014;19(8):12881-12897. doi 10.3390/molecules190812881
- Nsanganwimana F., Al Souki K.S., Waterlot C., Douay F., Pelfrène A., Ridošková A., Louvel B., Pourrut B. Potentials of *Miscanthus* × *giganteus* for phytostabilization of trace element-contaminated soils: ex situ experiment. *Ecotoxicol Environ Saf*. 2021;214:112125. doi 10.1016/j.ecoenv.2021.112125
- Rayburn A.L., Crawford J., Rayburn C.M., Juvik J.A. Genome size of three *Miscanthus* species. *Plant Mol Biol Rep*. 2009;27:184-188. doi 10.1007/s11105-008-0070-3
- Sanford G.R., Oates L.G., Jasrotia P., Thelen K.D., Robertson G.P., Jackson R.D. Comparative productivity of alternative cellulosic bioenergy cropping systems in the North Central USA. *Agric Ecosyst Environ*. 2016;216:344-355. doi 10.1016/j.agee.2015.10.018

- Sang T., Zhu W. China's bioenergy potential. *Glob Change Biol Bioenergy*. 2011;3(2):79-90. doi 10.1111/j.1757-1707.2010.01064.x
- Schäfer J., Sattler M., Iqbal Y., Lewandowski I., Bunzel M. Characterization of *Miscanthus* cell wall polymers. *Glob Change Biol Bioenergy*. 2019;11(1):191-205. doi 10.1111/gcbb.12538
- Shavyrkina N.A., Budaeva V.V., Skiba E.A., Gismatulina Y.A., Sakovich G.V. Review of current prospects for using miscanthus-based polymers. *Polymers*. 2023;15(14):3097. doi 10.3390/polym15143097
- Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210-3212. doi 10.1093/bioinformatics/btv351
- Simpson J.T., Wong K., Jackman S.D., Schein J.E., Jones S.J.M., Birol I. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 2009;19(6):1117-1123. doi 10.1101/gr.089532.108
- Stanke M., Steinkamp R., Waack S., Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res*. 2004;32:W309-W312. doi 10.1093/nar/gkh379
- Sun Q., Lin Q., Yi Z.L., Yang Z.R., Zhou F.S. A taxonomic revision of *Miscanthus* s.l. (Poaceae) from China. *Bot J Linn Soc*. 2010;164(2):178-220. doi 10.1111/j.1095-8339.2010.01082.x
- Swaminathan K., Chae W.B., Mitros T., Varala K., Xie L., Barling A., Glowacka K., Hall M., Jezowski S., Ming R., Hudson M., Juvik J.A., Rokhsar D.S., Moose S.P. A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy. *BMC Genomics*. 2012;13:142. doi 10.1186/1471-2164-13-142
- Técher D., Martinez-Chois C., Laval-Gilly P., Henry S., Bennasroune A., D'Innocenzo M., Falla J. Assessment of *Miscanthus × giganteus* for rhizoremediation of long term PAH contaminated soils. *Appl Soil Ecol*. 2012;62:42-49. doi 10.1016/j.apsoil.2012.07.009
- Trieu A., Belaffif M.B., Hirannaiah P., Manjunatha S., Wood R., Bathula Y., Billingsley R.L., Arpan A., Sacks E.J., Clemente T.E., Moose S.P., Reichert N.A., Swaminathan K. Transformation and gene editing in the bioenergy grass *Miscanthus*. *Biotechnol Biofuels Bioprod*. 2022;15:148. doi 10.1186/s13068-022-02241-8
- Van Weyenberg S., Ulens T., de Reu K., Zwervaegher I., Demeyer P., Pluym L. Feasibility of *Miscanthus* as alternative bedding for dairy cows. *Vet Med*. 2016;60(3):121-132. doi 10.17221/8059-VETMED
- Wang C., Kong Y., Hu R., Zhou G. *Miscanthus*: a fast-growing crop for environmental remediation and biofuel production. *Glob Change Biol Bioenergy*. 2021;13(1):58-69. doi 10.1111/gcbb.12761
- Yan H., Zhang X., Fu C., Huang L., Yin G., Nie G., Liu W., Chen T., Lee S., Ma X., Peng Y., Yan Y., Ling Y., Liu W. Chloroplast DNA variation and genetic structure of *Miscanthus sinensis* in southwest China. *Biochem Syst Ecol*. 2015;58:132-138. doi 10.1016/j.bse.2014.11.007
- Yang Y., Saand M.A., Huang L., Abdelaal W.B., Zhang J., Wu Y., Li J., Sirohi M.H., Wang F. Applications of multi-omics technologies for crop improvement. *Front Plant Sci*. 2021;12:563953. doi 10.3389/fpls.2021.563953
- Zhang G., Ge C., Xu P., Wang S., Cheng S., Han Y., Wang Y., ... Ming R., Dong S., Ma J., Zhang X., Chen C. The reference genome of *Miscanthus floridulus* illuminates the evolution of Saccharinae. *Nat Plants*. 2021;7(5):608-618. doi 10.1038/s41477-021-00908-y
- Zhao H., Wang B., He J., Yang J., Pan L., Sun D., Peng J. Genetic diversity and population structure of *Miscanthus sinensis* germplasm in China. *PLoS One*. 2013;8(10):e75672. doi 10.1371/journal.pone.0075672

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

Received November 17, 2025. Revised January 15, 2026. Accepted February 16, 2026.

doi 10.18699/vjgb-26-61

Development of potato (*Solanum tuberosum* L.) plants with inactivated *StPain-1* gene using CRISPR/Cas9

V.D. Karlov , M.K. Volkov , A.D. Antipov , Yu.S. Monahova , A.S. Trofimov, L.N. Konovalova ,
A.V. Babakov, R.A. Komakhin , V.V. Taranov 

All-Russia Research Institute of Agricultural Biotechnology, Moscow, Russia

 v4slvk@yandex.ru

Abstract. Storage of potato tubers intended for further processing is complicated by their cold-induced sweetening (CIS). Enzymatic hydrolysis of sucrose into glucose and fructose occurs at low temperatures. The resulting high hexose content adversely affects the quality of processed potato products such as chips and fries and promotes the formation of acrylamide, which is a neurotoxin and carcinogen. During CIS, sucrose hydrolysis is catalyzed by vacuolar invertase encoded by the *StPain-1* gene. Previous studies have shown that suppression of the enzyme activity confers potato resistance to CIS without reducing the nutritional value of tubers. In this study, CRISPR/Cas9 technology was used to generate *Solanum tuberosum* L. cv. Fritella plants with a knockout of *StPain-1*. Two binary vectors based on pKSE401 were constructed (Vector A and Vector B), each carrying two gRNAs targeting exon 1 (sgRNA-P1.A or sgRNA-P1.B) and exon 3 (sgRNA-P3.A or sgRNA-P3.B). Editing efficiency with each gRNA was evaluated through next-generation sequencing (NGS). Transformation with Vector A produced 48 transformants, 22 of which carried knockouts in all *StPain-1* alleles. Transformation with Vector B yielded 26 transformants, including 10 plants with complete *StPain-1* knockout. Chips made from tubers of nine edited Fritella plants demonstrated reduced vacuolar invertase activity: chips from *StPain-1* knockout lines were lighter compared to the non-edited control sample. Quantitative assessment of glucose, fructose, and sucrose levels, as well as *StPain-1* mRNA expression in tubers of four selected transformants (two per vector), confirmed enzyme inactivation. The resulting plants exhibit increased resistance to cold-induced sweetening and can be used as a promising source of nonfunctional *StPain-1* alleles for breeding new potato varieties.

Key words: vacuolar invertase; cold-induced sweetening; sucrose; glucose; fructose; potato; CRISPR/Cas9; guide RNA

For citation: Karlov V.D., Volkov M.K., Antipov A.D., Monahova Yu.S., Trofimov A.S., Konovalova L.N., Babakov A.V., Komakhin R.A., Taranov V.V. Development of potato (*Solanum tuberosum* L.) plants with inactivated *StPain-1* gene using CRISPR/Cas9. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed.* 2026;30(4):601-611. doi 10.18699/vjgb-26-611

Funding. The work was supported by the Ministry of Science and Higher Education of the Russian Federation, project FGUM-2026-0008.

Создание растений картофеля (*Solanum tuberosum* L.) с инактивированным геном *StPain-1* с помощью CRISPR/Cas9

В.Д. Карлов , М.К. Волков , А.Д. Антипов , Ю.С. Монахова , А.С. Трофимов,
Л.Н. Коновалова , А.В. Бабаков, Р.А. Комахин , В.В. Таранов 

Всероссийский научно-исследовательский институт сельскохозяйственной биотехнологии, Москва, Россия

 v4slvk@yandex.ru

Аннотация. Холодовое хранение клубней картофеля для последующей переработки осложнено развитием холодового осахаривания. При пониженной температуре происходит реакция ферментативного гидролиза сахарозы на глюкозу и фруктозу. Как следствие, высокое содержание гексоз негативно сказывается на качестве продуктов переработки картофеля – чипсов и фри, а также способствует образованию нейротоксина и канцерогена акриламида. Гидролиз сахарозы в процессе холодового осахаривания катализирует вакуолярная инвертаза, кодируемая геном *StPain-1*. Ранее было показано, что подавление активности фермента придает картофелю устойчивость к холодовому осахариванию, не снижая при этом питательных качеств клубней. В настоящей работе с помощью CRISPR/Cas9 были созданы растения культурного картофеля *Solanum tuberosum* L. сорта Фрителла с нокаутом гена *StPain-1*. Для редактирования целевого гена были собраны две генетические конструкции на основе вектора pKSE401 (Вектор А и Вектор В), каждая из которых несла две гидовые РНК: на участки в первом (sgRNA-P1.A и sgRNA-P1.B) и третьем (sgRNA-P3.A и sgRNA-P3.B) экзонах. Эффективность редактирования каждой gRNA была охарактеризована с применением высокопроизводительного секвенирования. С конструкцией Вектор А получено 48 трансформантов, среди которых 22 содержали нокаут всех аллелей гена *StPain-1*; с конструкцией Вектор В – 26 трансформантов, из

них 10 были с нокаутом. Путем обжарки чипсов из клубней девяти отредактированных растений картофеля сорта Фрителла было наглядно продемонстрировано снижение активности вакуолярной инвертазы. Чипсы из образцов с нокаутом гена *StPain-1* имели более светлую окраску, чем чипсы из контрольного образца без редактирования. По результатам проведенной работы отобрано четыре растения с нокаутом гена *StPain-1*, по два для каждой конструкции, для которых количественная оценка глюкозы, фруктозы и сахарозы, а также измерение уровня мРНК подтвердили инактивацию фермента. Полученные растения обладают повышенной устойчивостью к холодовому осахариванию и являются перспективными источниками нефункциональных аллелей гена *StPain-1* для селекции новых сортов картофеля.

Ключевые слова: вакуолярная инвертаза; холодовое осахаривание; сахароза; глюкоза; фруктоза; картофель; CRISPR/Cas9; гидовая РНК

Introduction

Potato (*Solanum tuberosum* L.) is one of the world's most important food crops. It ranks third among crops consumed worldwide and second in Russia (Osipov, Zeldner, 2023). In countries with temperate climate, potatoes are harvested once a year. In order to preserve commercial qualities, harvested tubers are stored at temperatures of 4–10 °C (Wustman, Struik, 2007). At these storage temperatures, the process called cold-induced sweetening (CIS) occurs in the tubers. It involves excessive accumulation of glucose and fructose (Zrenner et al., 1996). Tubers having undergone CIS are unsuitable for further processing into chips or fries, as they turn dark and bitter (Murata, 2021).

Vacuolar invertase (β -fructofuranosidase; EC 3.2.1.26) plays a key role in CIS (Davies et al., 1989; Richardson et al., 1990; Blenkinsop et al., 2003). It catalyzes the hydrolysis of sucrose into hexoses: glucose and fructose (Zrenner et al., 1996; Zhu et al., 2024). During high-temperature processing, the Maillard reaction occurs, whereby the hexoses interact with the amino groups of proteins and free amino acids to produce melanoidin pigments. Another product of this reaction is acrylamide, a neurotoxin and carcinogen that poses a hazard to human health (Tareke et al., 2002; Siaw et al., 2018).

The glucose and fructose contents largely determine the color, taste, and safety of roasted potato products (Siaw et al., 2018). Use of tubers with hexose content greater than 1 mg per 1 g of fresh weight (mg/g FTW) is not recommended due to high accumulation of acrylamide in the final product (Biedermann-Brem et al., 2003).

Several experimental approaches have been developed to suppress the enzymatic activity of vacuolar invertase in order to reduce hexose levels and increase the resistance of potato tubers to CIS. These approaches include the overexpression of genes for natural vacuolar invertase inhibitors (Greiner et al., 1999; McKenzie et al., 2013), the suppression of the endogenous *StPain-1* gene expression via RNA silencing (Zrenner et al., 1996; Zhang et al., 2008; Bhaskar et al., 2010; Wu et al., 2011; Zhu et

al., 2014; Toufiq et al., 2025), and the inactivation of the *StPain-1* gene using genome editing techniques (Clasen et al., 2016; Yasmeen et al., 2022; Ly et al., 2023; Teper-Bamnolker et al., 2023; Shumbe et al., 2024; Egorova et al., 2025; Massa et al., 2025).

Genome editing appears to be the most promising method for raising potato varieties resistant to CIS, as it enables the creation of new, nonfunctional variants of the *StPain-1* gene. These nonfunctional *StPain-1* alleles can then be transferred to new potato varieties through crossbreeding. Furthermore, as demonstrated in (Yasmeen et al., 2022), it is sufficient to inactivate only part of the alleles to effectively reduce vacuolar invertase activity.

Previously, traditional breeding methods were used to develop potato varieties intended for processing into chips and fries, such as Queen Anne, Lady Claire, Barin, Vympel, and others. Tubers of the aforementioned potato varieties accumulate less glucose and fructose during cold storage. However, hexose content is sufficient to reduce the quality of the final product (Kulakova et al., 2022; Bruno et al., 2025). The growing demand for potato products means a need for developing new varieties that would be more resistant to CIS.

Although several potato genotypes resistant to CIS have been created based on the Atlantic, Desiree, and Symphony varieties using genome editing techniques, legislative restrictions in a number of countries prevent the free use of such plants for industrial purposes due to the presence of a transgene (Ly et al., 2023; Teper-Bamnolker et al., 2023; Egorova et al., 2025; Massa et al., 2025).

The aim of this study was to knock out the vacuolar invertase gene in Fritella potato in order to obtain a plant source of nonfunctional *StPain-1* alleles for the breeding of new varieties (Simakov et al., 2017). Fritella variety was chosen due to high tuber suitability for making French fries and chips. The plants developed in this study show promise as donors for the selection of new nontransgenic potato varieties with increased resistance to CIS.

Materials and methods

Plant material. Aseptic tetraploid potato plants (*S. tuberosum* L.) of the Fritella cv. (provided by the A.G. Lorch Research Institute of Potato Farming, Moscow) were cultivated on MS medium (Murashige, Skoog, 1962) at 20–22 °C, with the 16-hour light/8-hour dark schedule, and a photosynthetic photon flux density (PPFD) of 80 $\mu\text{mol m}^{-2} \text{s}^{-1}$. The absence of viral contamination in the aseptic plants was confirmed by real-time PCR analysis with a commercially available kit “PHY-TOSCREEN Potato Virus X, Y, M, L, S, A, PSTVd-RT” in expert molecular study No. 57/2023 dated July 26, 2023 by Syntol LLC (Russia). Agrobacterium-mediated transformation of aseptic leaf explants was performed according to a previously published method (Vetchinkina et al., 2016). Potato plants were grown in vessels in a greenhouse at 20–21 °C, with the 16-hour light/8-hour dark schedule and PPFD of 150 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Tubers were stored in the dark at 4 °C.

Agrobacterium strain. *Agrobacterium tumefaciens* bacteria of AGL0 strain were used to transform potato explants.

gRNA selection and cloning. Genomic DNA was extracted from Fritella potato leaves using cetyl trimethylammonium bromide (CTAB). Coding sequences of the first and third exons of the *StPain-1* gene were analyzed using CRISPOR (<https://crispor.org>) to choose gRNAs (Concordet, Haeussler, 2018). Four sgRNAs were used in the study: two gRNAs targeting exon 1 (sgRNA-P1.A and sgRNA-P1.B), and the other two targeting exon 3 (sgRNA-P3.A and sgRNA-P3.B). The sequence of sgRNA-P3.A was taken from (Yasmeen et al., 2022). gRNAs were cloned in pairs into the pKSE401 binary vector for genome editing according to the recommendations (Xing et al., 2014). This resulted in the creation of Vector A (sgRNA-P1.A and sgRNA-P3.A) and Vector B (sgRNA-P1.B and sgRNA-P3.B).

Genotyping of transgenic plants. Genomic DNA was extracted from leaves of reproductive T0 generation plants using CTAB. Regions of exons 1 and 3 of the *StPain-1* gene were amplified using the oligonucleotides listed in Supplementary Table S1¹.

For PCR, a mixture containing 1× Taq polymerase buffer, 0.2 mM dNTP, 0.5 μM of each primer, 1.25 U of Taq DNA polymerase (Eurogen, Moscow), and 20 ng of genomic DNA was prepared and adjusted to a final volume of 25 μL . PCR was performed under the following conditions: predenaturation at 94 °C for 2 min followed by 30 cycles of 10 s at 94 °C, 15 s at 58–62 °C, 40 s at 72 °C, and then the final elongation step at 72 °C for

5 min. The amplicons were resolved in 1.5 % agarose gel and purified with a ColGen kit (Syntol, Moscow). The amplicon sequencing was performed on the Illumina MiSeq platform in the Biotechnology Shared Access Center at the All-Russia Research Institute of Agricultural Biotechnology. Sequencing quality was assessed with the MultiQC tool ver. 1.22.2 (Ewels et al., 2016). Editing efficiencies for each target region in each plant were determined as a percentage of the number of reads containing mutations relative to the total number of reads, using CRISPResso2 ver. 2.3.1 software (Clement et al., 2019).

Sugar content assessment. Tuber tissue (100 mg) was homogenized with a pestle in 500 μL of water. The volume of the homogenate was adjusted to 1,000 μL with ddH₂O and mixed by vortexing. After centrifugation at 13,000g for 2 min the debris was removed. To measure the sugar content, 100 μL of the extract was taken. The glucose, fructose, and sucrose contents were quantified with commercially available Enzytec Liquid D-Glucose/D-Fructose and Enzytec Liquid Sucrose/D-Glucose kits (R-Biopharm, Germany), following the manufacturer’s recommendations. The glucose, fructose, and sucrose contents are given in mg/g FTW.

Starch content assessment. A piece of tuber (50 mg) was homogenized with a pestle in 500 μL of a 4:1 mixture of dimethyl sulfoxide (DMSO) and hydrochloric acid (HCl). The suspension volume was adjusted to 1.25 mL using the same mixture, mixed by vortexing, and incubated at 60 °C for 30 min. The suspension was diluted with 2.5 mL of ddH₂O and titrated with 5 M NaOH until pH 4–5. The mixture was centrifuged at 13,000g for 2 min, and 100 μL of the supernatant was collected for analysis. Quantitative assessment was performed using a commercially available Enzytec Liquid Starch kit (R-Biopharm, Germany), following the manufacturer’s recommendations. The starch content is given in mg/g FTW).

RNA extraction. To extract RNA, 100 mg of tuber tissue was crushed with a pestle in a mortar containing liquid nitrogen. RNA was then extracted by the column method using a SKYprep RNA Pure Plant Plus Kit (SkyGen, Russia). The RNA quality was checked by electrophoresis in 1.5 % agarose gel.

Reverse transcription. Reverse transcription was performed using MMLV reverse transcriptase (Eurogen, Russia). To perform the reaction, a mixture of 1× First strand buffer, 0.2 mM dNTP mix, 2 uM oligo(dT)₁₅ primer, 1 mg of isolated RNA, 2 mM DTT, and 100 U of MMLV-RT was prepared with a final volume of 20 μL according to the manufacturer’s recommendations. The reaction was carried out for 60 min at 37 °C, followed

¹ Supplementary Tables S1–S7 are available at:

https://vavilov.elpub.ru/jour/manager/files/Suppl_Karlov_Engl_30_4.xlsx

by 10-min incubation at 70 °C. The synthesized cDNA was then used as a template for quantitative PCR.

Quantitative PCR analysis. The reaction mixture was prepared using 5× qPCRmix-HS SYBR (Evrogen, Russia). The oligonucleotides used are listed in Table S1. The reaction was carried out under the following conditions: 3 min at 95 °C, followed by 50 cycles of 10 s at 95 °C, 30 s at 60 °C, and 30 s at 72 °C. The *Tubulin* and *EF-1a* genes were used as reference sequences (Table S1) (Nicot et al., 2005; Zhu et al., 2014). The relative expression of the *StPain-1* gene was calculated using the $2^{-\Delta\Delta C_t}$ method (Vandesompele et al., 2002).

Frying chips from tubers. Tubers harvested from the T0 vegetative generation after 2 months of storage at 4 °C were cut into ~1–2 mm slices and fried in vegetable oil at 170 °C for 2 min.

Statistical analysis. All experiments were performed in triplicate. The resulting data were processed by ANOVA. The studied samples were compared with the control group using the Dunnett test. Mean values and their standard deviations were calculated to construct glucose, fructose, sucrose and starch content and relative

expression charts. All the calculations were performed in GraphPad Prism 8.

Results

Choice of guide RNAs to introduce knockout mutations into the vacuolar invertase gene *StPain-1*

Vacuolar invertase gene *StPain-1* comprises seven exons and six introns (Abbas et al., 2022). Exons 1 and 3 were selected as targets for editing since introducing mutations into an upstream part of a gene increases knockout probability. Amplicons of exons 1 and 3 of the *StPain-1* gene obtained from the genomic DNA of the Fritella cv. were sequenced so as to choose gRNAs afterwards. Alignment of the nucleotide sequences of the first exon amplicons to the reference sequence of the *StPain-1* gene (NCBI: HQ110080.1) revealed several single nucleotide polymorphisms at the 32nd, 174th, and 288th positions from the start codon. Among these, the substitution of C for T at the 32nd position (c.32T>C) resulted in the amino acid substitution Leu>Pro in three of the four alleles (Fig. 1a, b). The nucleotide sequences

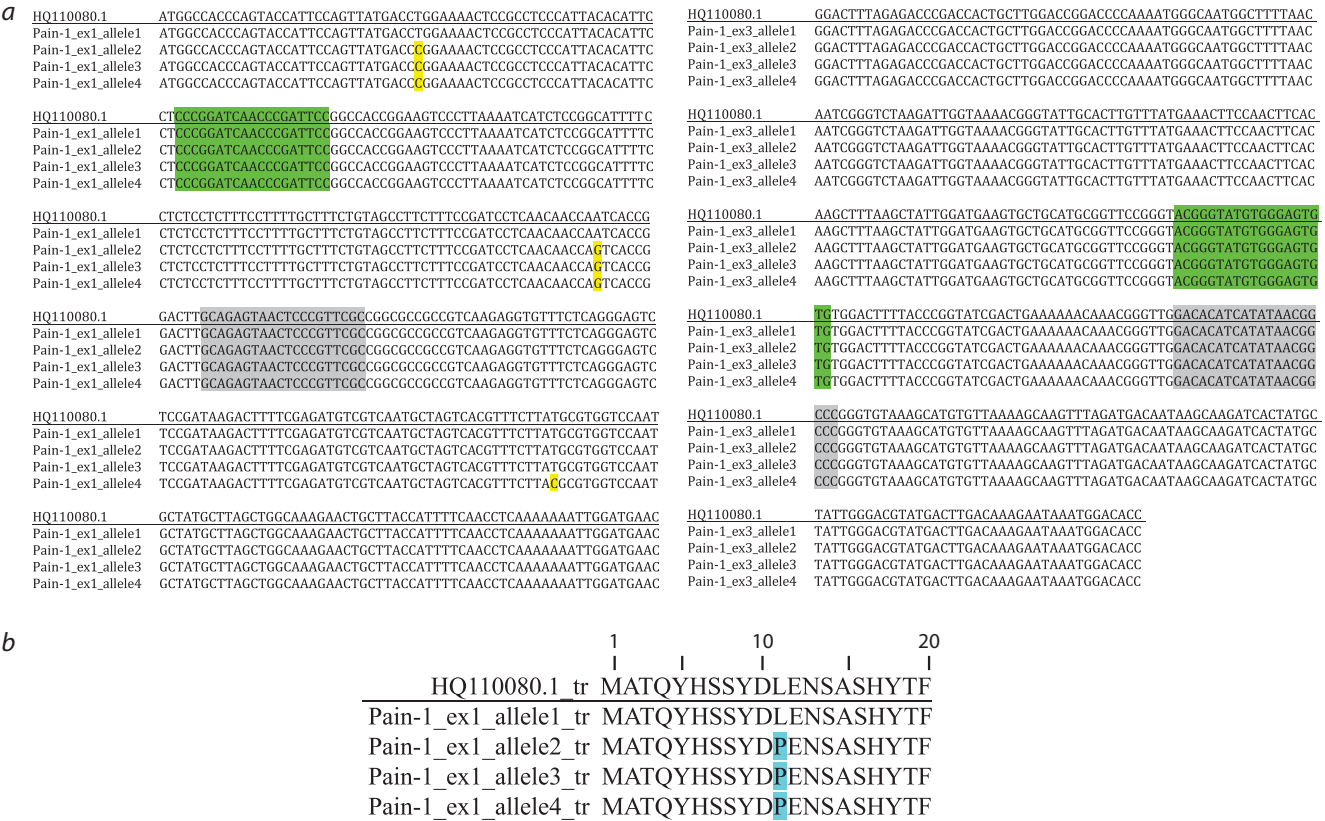


Fig. 1. The similarity in the *StPain-1* gene regions across four alleles.
a – alignment of the nucleotide sequences of exon 1 and fragment of exon 3 of four identified *StPain-1* gene alleles to the reference sequence (Madeira et al., 2024). Polymorphisms are shaded yellow. Editing sites are shaded green for Vector A and gray for Vector B; b – translation of the first 20 codons of the reference *StPain-1* sequence and the sequences of the four alleles found in the genome of the Fritella potato cultivar. Polymorphisms in the amino acid sequences of the alleles are shaded blue.

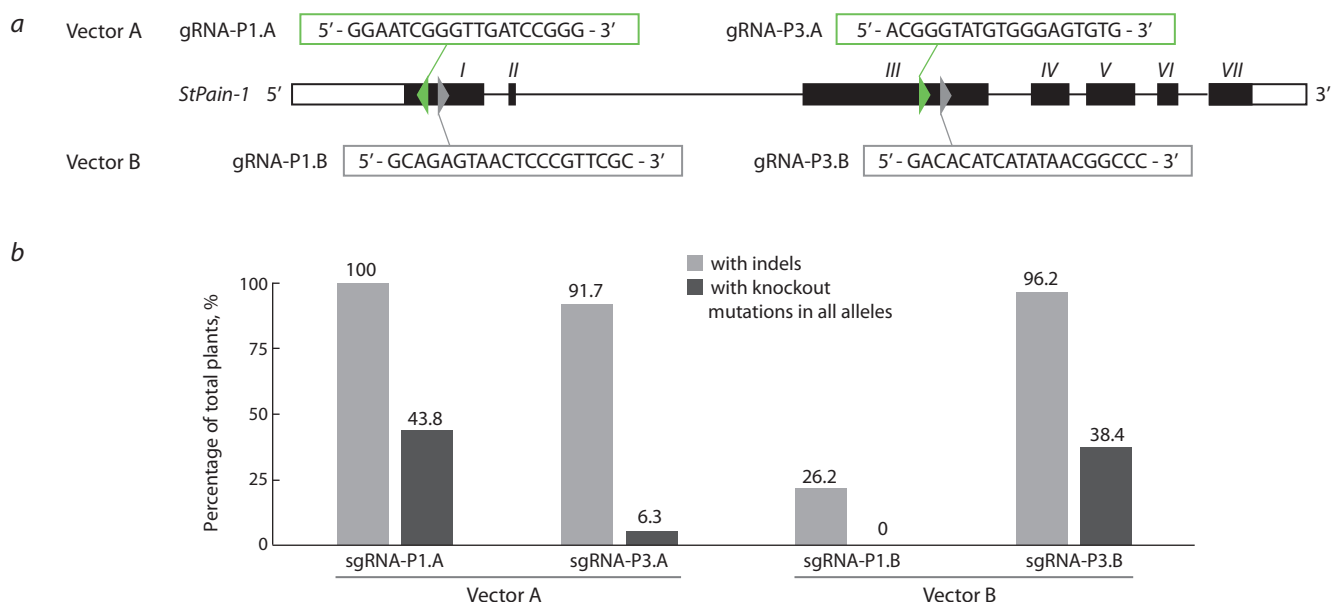


Fig. 2. Efficiency of introducing mutations into the vacuolar invertase gene *StPain-1* of Fritella plants using different guide RNAs.

a – the exon-intron structure of the *StPain-1* gene and the location of the guide RNAs for Vector A (highlighted in green) and Vector B (highlighted in gray). Arrows indicate the directions of the guide RNAs; *b* – percentages of plants with indel mutations (shown in gray) and with knockout mutations in all alleles (shown in black) of the *StPain-1* gene among transformants. For Vector A, *n* = 48 transformants; for Vector B, *n* = 26 transformants.

of the exon 3 fragment in all alleles were identical to the reference sequence (Fig. 1*a*).

Four gRNAs were selected to edit the most conserved regions of the vacuolar invertase gene: two gRNAs targeting exon 1 (sgRNA-P1.A and sgRNA-P1.B), and the other two targeting exon 3 (sgRNA-P3.A and sgRNA-P3.B) (Fig. 1*a* and 2*a*). The sequences of three gRNAs (sgRNA-P1.A, sgRNA-P1.B, and sgRNA-P3.B) were selected using the CRISPOR tool, and one guide RNA (sgRNA-P3.A) was taken from an earlier work (Yasmeen et al., 2022). Two pairs of gRNAs were cloned into two genetic constructs, designated Vector A (sgRNA-P1.A and sgRNA-P3.A) and Vector B (sgRNA-P1.B and sgRNA-P3.B) (Fig. 2*a*). The simultaneous use of two guide RNAs in a single genetic construct was required to increase the probability of gene knockout due to an additive effect in case of low editing efficiency.

The efficiency of *StPain-1* gene editing process depends on the selection of gRNA

The *Agrobacterium*-mediated transformation of potato leaf explants with Vector A and with Vector B yielded 48 and 26 regenerants, respectively. The editing efficiency was evaluated using targeted high-throughput sequencing on the Illumina platform. The average number of reads was 1,447, ranging from a minimum of 128 to a maximum of 4,251. Genotyping of the transformants revealed that both genetic constructs introduced indel mutations with high efficiency (Fig. 2*b*).

The highest mutation efficiency was achieved using the guide RNAs sgRNA-P1.A and sgRNA-P3.B, which created indels in 100 and 96.2 % of transformants, respectively. However, knockout of all *StPain-1* alleles was achieved in 43.75 % of transformants using sgRNA-P1.A and in 38.4 % of transformants using sgRNA-P3.B. Despite the high efficiency of introducing indels (91.7 % of transformants) with sgRNA-P3.A, knockout of all *StPain-1* alleles was detected in only 6.25 % of transformants (3 out of 48). The sgRNA-P1.B gRNA showed the lowest efficiency in introducing indels, which were observed in only 26.9 % of plants (7 out of 26). Also, no transformants with knockout of all *StPain-1* alleles were obtained using sgRNA-1.B. The percentages of indels and knockout mutations in transformants as calculated from the NGS results are shown in Tables S2 and S3.

The diversity of mutant variants depends on the editing site

Targeted sequencing on the Illumina platform allowed characterization of the diversity of types of indels obtained with each of the four gRNAs (Fig. 3). The numbers of mutation variants correlated with gRNA efficiency: the most efficient guide RNAs generated a greater diversity of indels. The vast majority of mutations were deletions of 1–10 nucleotides.

StPain-1 gene editing with Vector A resulted in the detection of 17 different indels in exon 1. The most prevalent of these were deletions of 4 nucleotides (in

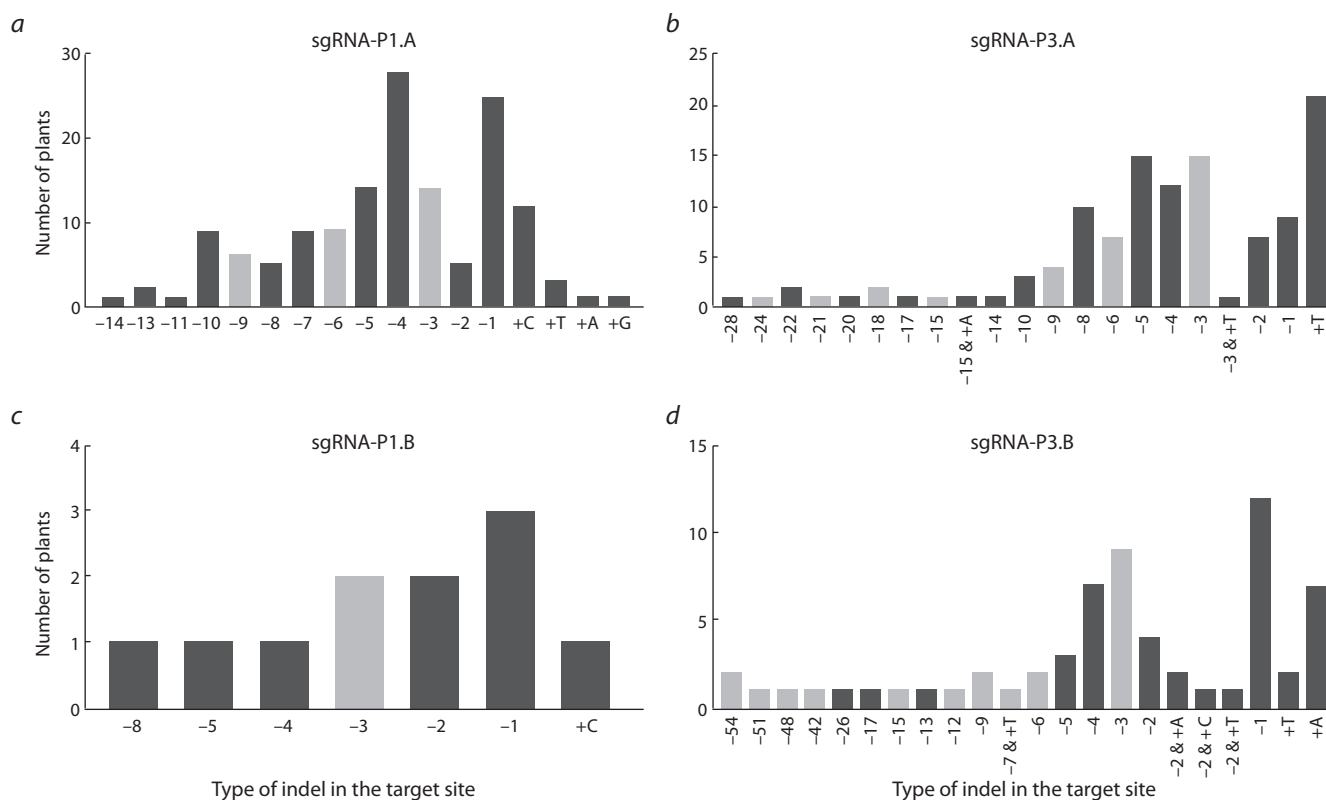


Fig. 3. Diversity of indel mutations causing (black bars) and not causing (gray bars) open reading frame shifts in the *StPain-1* gene in Fritella transformants created using the Vector A construct with sgRNA-P1.A (a) and sgRNA-P3.A (b) or Vector B with sgRNA-P1.B (c) and sgRNA-P3.B (d) guides

28 plants) and 1 nucleotide (in 24 plants). The largest deletion included 14 nucleotides. Three indel types did not cause open reading frame shift: deletions of 3, 6, and 9 nucleotides (Fig. 3a). Twenty-one types of indels were found in exon 3. The most common of them were insertions of 1 nucleotide with a thymine base (in 21 plants) and deletions of 5 and 3 nucleotides (in 15 plants each). The largest deletion in exon 3 included 28 nucleotides (Fig. 3b).

Use of the Vector B construct in exon 1 resulted in the detection of only 7 mutation variants: deletions of 8, 5, 4, 3, 2, and 1 nucleotides and an insertion of 1 nucleotide with a cytosine base (Fig. 3c). Exon 3 exhibited the greatest diversity of indels (22 types). The most common was a deletion of 1 nucleotide (12 plants). The largest deletion included 54 nucleotides (two plants). Almost half of the detected indel mutations were deletions of multiples of three nucleotides (Fig. 3d). It can therefore be concluded that a high level of gRNA efficiency is associated with the generation of a wide variety of indel mutations. Furthermore, the diversity of these mutations is determined by the specific editing site. The vast majority of indel mutations are deletions of 1 to 10 nucleotides (see Tables S2–S7 for details).

Evaluation of plants with *StPain-1* gene knockout for resistance to cold-induced sweetening

For phenotypic evaluation of transformant resistance to CIS by frying chips, tubers were taken from nine plants obtained using Vector A (5 - 2, 11 - 2, 12 - 1, 13 - 2, 26 - 1, 36 - 1, 45 - 1, 49 - 1, 57 - 2) and from five plants obtained using Vector B (16 - 1, 20 - 1, 42 - 2, 50 - 2, 56 - 2). Tubers from transgenic plant 27 - 3 (Vector B) that did not contain indels in any of the *StPain-1* alleles were used as a control (see the Table).

Chip frying showed that samples with edited vacuolar invertase gene variants were lighter in color than the control sample. Sample 11 - 2 (Vector A) showed a clear discrepancy between the genotype and phenotype: the chips were almost completely dark despite the knockout being confirmed by NGS (see the “Discussion” section for details).

The experiment also included several samples from plants containing both edited and wild-type alleles (26 - 1 – Vector A and 20 - 1, 50 - 2 – Vector B). The color of the chips from sample 50 - 2 was noticeably lighter than that of the control but darker than the knockout samples. Meanwhile, samples 26 - 1 and 20 - 1 had enough inactivated alleles to keep the chips light in color.

Characteristics of Fritella transformants based on the knockout mutation content in exons 1 and 3 of the *StPain-1* gene and the reaction of reducing sugars in tuber slices upon frying

Genetic consrtuct	Sample	Reads with knockout, %		Chips
		in exon 1	in exon 3	
Vector A	<u>27-3</u>	0	0	
	5-2	100	29	
	11-2	100	86	
	12-1	100	75	
	13-2	100	100	
	26-1	46	4	
	<u>36-1</u>	100	100	
	<u>45-1</u>	100	50	
	49-1	100	0	
	57-2	100	28	
Vector	<u>16-1</u>	20	100	
	20-1	0	47.5	
	42-2	0	100	
	50-2	24	64.7	
	<u>56-2</u>	0	100	

Note. Samples taken for the assessment of starch and sugar content, as well as *StPain-1* mRNA levels, are underlined.

A similar effect was previously observed by A. Yasmeen et al. (2022), where two of the four *StPain-1* alleles were inactivated, ensuring CIS resistance in edited plants (see the “Discussion” section for details). Tubers of four transformants were selected for subsequent measurement of *StPain-1* mRNA levels and quantification of sugars and starch: 36 - 1 and 45 - 1 (Vector A), 16 - 1 and 56 - 2 (Vector B).
The introduction of indel mutations into the coding sequence of a gene leads to the synthesis of aberrant mRNA, which is destroyed by cellular mRNA quality control systems. When measuring *StPain-1* expression in the tubers of selected plants with a knockout of the

target gene, a significant decrease in the level of the corresponding mRNA was observed relative to the *Tubulin* and *EF-1a* genes (Fig. 4a).
To confirm the inactivation of the vacuolar invertase enzyme, the sugar and starch contents were assayed. The glucose contents in the knockout samples ranged from 0.2 to 0.6 mg/g FTW (Fig. 4b), whereas the fructose content ranged from 0.1 to 0.2 mg/g FTW (Fig. 4c). These values were significantly lower than those observed in the control sample. Conversely, the sucrose content of the *StPain-1* knockout samples ranged from 18.1 to 24.5 mg/g FTW, which was significantly higher than in the control sample (Fig. 4d). As expected, no differences

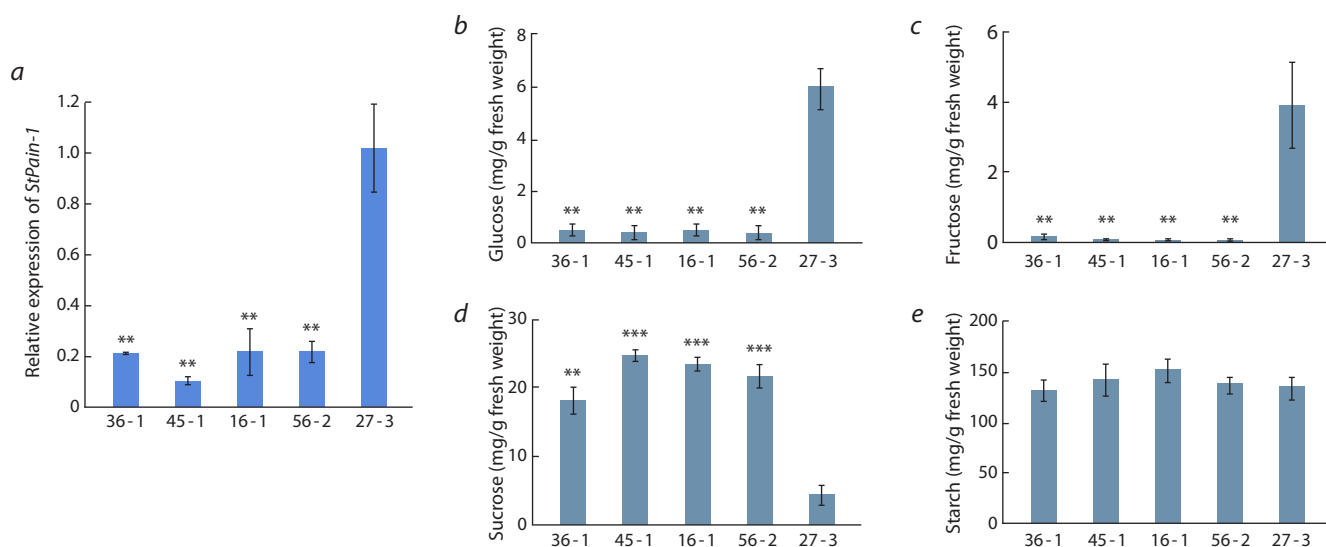


Fig. 4. Assessment of carbohydrate contents and levels of *StPain-1* mRNA in tubers of some Fritella potato transformants with knockout of all alleles of the *StPain-1* gene.

a – comparison of *StPain-1* mRNA levels relative to the reference genes *Tubulin* and *EF-1a* in potatoes by real-time PCR. Carbohydrate contents: b – glucose, c – fructose, d – sucrose, e – starch, measured in mg per g of fresh tuber weight. The graphs show the mean values ($n = 3$) and their standard errors. The significance of the difference calculated using Dunnett's multiple comparison test is indicated with asterisks (*) $P \leq 0.05$; (**) $P \leq 0.01$; (***) $P \leq 0.001$.

were observed between the samples when the starch content in the tubers was measured (Fig. 4e).

This work resulted in the development and characterization of Fritella plants with a knockout of the *StPain-1* gene. Glucose and fructose levels in tubers of the obtained plants were below the recommended amount of 1 mg of hexoses per 1 g FTW (Biedermann-Brem et al., 2003). These plants are a promising source of nonfunctional *StPain-1* alleles for breeding new potato varieties resistant to CIS.

Discussion

Modern approaches to targeted genome mutagenesis have accelerated the development of new crop varieties significantly. The key objectives in modern breeding programs are to enhance nutritional quality and increase resistance to abiotic and biotic stresses, ensuring more efficient usage of natural resources and reducing economic costs in agriculture. This study focuses on an economically important issue in food industry: CIS of potato tubers, which is caused by vacuolar invertase.

The CRISPR/Cas9 tool was used to inactivate the vacuolar invertase gene *StPain-1*. Based on the pKSE401 vector (Xing et al., 2014), two constructs containing a pair of guide RNAs each were created: Vector A and Vector B. In each construct, one gRNA targeted exon 1 and the other targeted exon 3. Both genetic constructs demonstrated high efficiency in editing the target gene. According to NGS data, Vector A resulted in indel muta-

tions in at least one allele of the target gene in all 48 regenerants, whereas Vector B produced indel mutations in 25 out of 26 regenerants. Moreover, Vector A knocked out all *StPain-1* alleles in 22 out of the 48 regenerants, whereas Vector B, in only 10 out of the 26. The high efficiency of target gene editing was largely ensured by one of the two cloned gRNAs: sgRNA-P1.A for Vector A and sgRNA-P3.B for Vector B.

This work clearly demonstrates the critical importance of choosing right guide RNAs for gene editing in crops. Targeted amplicon sequencing was carried out to compare the efficiency with which four guide RNAs introduced indels. The sgRNA-P1.A, sgRNA-P1.B, and sgRNA-P3.B sequences were chosen for their predicted high efficiency and low potential for side effects, as determined by the CRISPOR and CRISPR-P algorithms (Lei et al., 2014; Concordet, Haeussler, 2018).

The sgRNA-P3.A sequence was taken from the literature (Yasmeen et al., 2022). In practice, however, sgRNA-P1.B exhibited the lowest editing efficiency. It produced indel mutations in only 7 out of 26 plants and not in all alleles. The outcome of plant genome editing depends on various factors, including the delivery method of the editing system, the stability of the gRNA-Cas9 complex, the spatial organization of the genome and the accessibility of the target site to the gRNA-Cas9 complex, the functioning of genome repair systems, etc. In (Egorova et al., 2025), three guide RNAs (gRNA1, gRNA2, and gRNA3) were used within a single construct

to edit the *StPain-1* gene. Two of these (gRNA1 and gRNA2) targeted sites in the first exon. Upon sequencing the resulting plant sample, it was found that editing with gRNA2 was much more effective than with gRNA1. Conversely, despite its high predicted activity, gRNA3 exhibited the lowest efficiency among all the guide RNAs used in the experiment. Furthermore, the editing efficiency may depend on the cultivar. For example, the *StPain-1* gene knockout was more efficient in the Atlantic cv. than in Spunta when the same guide RNAs were employed (Massa et al., 2025).

The CRISPR/Cas9 system is renowned for its high specificity. However, when editing organisms, an off-target effect often occurs, whereby indel mutations are introduced into non-target regions of the genome with a partially complementary gRNA sequence. These additional mutations can affect the phenotype of the plant. Obviously, sets of off-target sites are different with different guide RNAs. Therefore, plants obtained using different guide RNAs will have different sets of off-target mutations. In this study, the usage of two different genetic constructs enabled us to monitor the potential impact of off-target editing in the assessment of the phenotypes of edited plants.

Editing the *StPain-1* gene produced a variety of mutant alleles, the majority of which exhibited deletions of 1, 3, or 4 nucleotides. It should be noted that some differences in the formation of indels were observed when editing different exons and different regions of the same exon. Exon 1 is 361 bp in size, and exon 3 is 860 bp. The maximum deletion size in exon 1 was 14 nucleotides, whereas in exon 3 a deletion size reached 54 nucleotides. Among all +1 indel mutations, insertion of C was prevalent in exon 1, whereas insertion of T was prevalent in exon 3. Additionally, indels that did not cause an open reading frame shift (i. e. multiples of three nucleotides) occurred more often in exon 3.

The differences in editing various sites within the same exon were primarily related to the efficiency of indel mutation generation. Knockout efficiency was six times lower with sgRNA-P3.A than with sgRNA-P3.B. Notably, the distance between the Cas9 target sites in exon 3 was only 61 bp. The difference in the knockout efficiency was even greater for guide RNAs sgRNA-P1.A (43.8 %) and sgRNA-P1.B (0 %), which were 137 bp apart.

To visually confirm the inactivation of vacuolar invertase in the Fritella regenerants, an experiment involving frying chips was conducted. Of the 14 regenerants obtained with Vectors A and B, 11 demonstrated the knockout of all gene alleles. When slices of tubers from knockout sample 11 - 2 (Vector A) were fried, the chips exhibited dark coloring with small light segments.

This effect may have been due to number of causes. For instance, dark chip color resulting from high hexose levels may have been an individual response of the transformant to altered sucrose metabolism as a consequence of the *StPain-1* gene knockout. Accumulation of hexoses in this case was probably caused by another sucrose-hydrolyzing enzyme, such as sucrose synthases or homologous invertases.

Alternatively, as a result of *Agrobacterium*-mediated transformation, T-DNA may have been inserted into the plant genome, often in euchromatic regions. This genome modification could result in altered expression of certain genes, which could subsequently cause changes in sucrose metabolism. Another cause of the observed effect could be the chimerism of the regenerated plant. The presence of cells with different genotypes within a single plant is a well-known phenomenon that occurs during the *Agrobacterium*-mediated transformation of callus. A notable example of this effect can be seen in a plant obtained by inactivating the *StLEAFY* gene (Lebedeva et al., 2022).

Three other regenerants whose tubers were used to make chips retained one or two wild-type alleles (26 - 1, 20 - 1, 50 - 2). However, dark coloration was only observed in the chips from sample 50 - 2. Indeed, a decreased content of reducing sugars had previously been observed when the *StPain-1* gene was knocked down (Yasmeen et al., 2022).

Several hypotheses can be proposed to explain this phenomenon. Tetraploid potatoes are known for their high degree of heterozygosity. Different *StPain-1* alleles of the gene may contain nucleotide polymorphisms that affect the enzymatic activity of vacuolar invertase (Draffehn et al., 2010; Slugina, Kochieva, 2014). In this regard, inactivation of the allele encoding the most active enzyme can lead to a significant decrease in CIS and vice versa. Also, the reaction rate can be influenced by the amount of enzyme, which is determined by systems that regulate cellular gene expression. Moreover, mRNA synthesis most often occurs in an allele-specific manner (Pham et al., 2017). Therefore, introducing mutations into the intensely transcribed *StPain-1* alleles could also affect the glucose and fructose contents in tubers of the transformants under study.

Conclusion

Plants with increased resistance to CIS were obtained using the genome editing method, based on the Russian-bred Fritella potato variety intended for industrial processing. These plants show great promise for breeding new potato varieties and can be used as donors of non-functional alleles of the vacuolar invertase gene.

References

- Abbas A., Shah A.N., Shah A.A., Nadeem M.A., Alsaleh A., Javed T., Alotaibi S.S., Abdelsalam N.R. Genome-wide analysis of invertase gene family, and expression profiling under abiotic stress conditions in potato. *Biology (Basel)*. 2022;11(4):539. doi 10.3390/biology11040539
- Bhaskar P.B., Wu L., Busse J.S., Whitty B.R., Hamernik A.J., Jan-sky S.H., Buell C.R., Bethke P.C., Jiang J. Suppression of the vacuolar invertase gene prevents cold-induced sweetening in potato. *Plant Physiol.* 2010;154(2):939-948. doi 10.1104/pp.110.162545
- Biedermann-Brem S., Noti A., Grob K., Imhof D., Bazzocco D., Pfeifferle A. How much reducing sugar may potatoes contain to avoid excessive acrylamide formation during roasting and baking? *Eur Food Res Technol.* 2003;217:369-373. doi 10.1007/s00217-003-0779-z
- Blenkinsop R.W., Yada R.Y., Marangoni A.G. Metabolic control of low-temperature sweetening in potato tubers during postharvest storage. *Hortic Rev.* 2003;30:317-354. doi 10.1002/9780470650837.ch8
- Bruno F., Hein I., Jorat M.E., Ledbetter M., Harrower B., Davies B., Sturrock K., Montague G., Knott M., McNamara G., Fiore A. Effects of agronomical practices on potato growth, nutritional profile, and suitability for frying. *J Sci Food Agric.* 2025;105(7):3983-3992. doi 10.1002/jsfa.14147
- Clasen B.M., Stoddard T.J., Luo S., Demorest Z.L., Li J., Cedrone F., Tibebu R., ... Haun W., Baltes N.J., Mathis L., Voytas D.F., Zhang F. Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnol J.* 2016;14:169-176. doi 10.1111/pbi.12370
- Clement K., Rees H., Canver M.C., Gehrke J.M., Farouni R., Hsu J.Y., Cole M.A., Liu D.R., Joung J.K., Bauer D.E., Pinello L. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol.* 2019;37(3):224-226. doi 10.1038/s41587-019-0032-3
- Concordet J.-P., Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* 2018;46(W1):W242-W245. doi 10.1093/nar/gky354
- Davies H.V., Jefferies R.A., Scobie L. Hexose accumulation in cold-stored tubers of potato (*Solanum tuberosum* L.): the effects of water stress. *J Plant Physiol.* 1989;134(4):471-475. doi 10.1016/S0176-1617(89)80012-4
- Draffehn A.M., Meller S., Li L., Gebhardt C. Natural diversity of potato (*Solanum tuberosum*) invertases. *BMC Plant Biol.* 2010;10:271. doi 10.1186/1471-2229-10-271
- Egorova A.A., Zykova T.E., Kostina N.E., Saboiev I.A., Koloshina K.A., Filipenko E.A., Hoffee I., Hertig C., Hiekel S., Kumlehn J., Kochetov A.V., Gerasimova S.V. Reduction in cold-induced sweetening by Cas9 endonuclease-mediated knockout of the *POTATO VACUOLAR INVERTASE 1* gene in the cultivar 'Symfonia.' *Potato Res.* 2025;68:1571-1593. doi 10.1007/s11540-024-09800-6
- Ewels P., Magnusson M., Lundin S., Källér M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016;32:3047-3048. doi 10.1093/bioinformatics/btw354
- Greiner S., Rausch T., Sonnewald U., Herbers K. Ectopic expression of a tobacco invertase inhibitor homolog prevents cold-induced sweetening of potato tubers. *Nat Biotechnol.* 1999;17:708-711. doi 10.1038/10924
- Kulakova A.V., Efremov G.I., Shchennikova A.V., Kochieva E.Z. Dependence of the content of starch and reducing sugars on the level of expression of the genes of β -amylases *StBAM1* and *StBAM9* and the amylase inhibitor *StAI* during long-term low-temperature storage of potato tubers. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2022;26(6):507-514. doi 10.18699/VJGB-22-62
- Lebedeva M., Komakhin R., Konovalova L., Ivanova L., Taranov V., Monakhova Y., Babakov A., Klepikova A., Zlobin N. Development of potato (*Solanum tuberosum* L.) plants with *StLEAFY* knockout. *Planta.* 2022;256:116. doi 10.1007/s00425-022-04032-9
- Lei Y., Lu L., Liu H.-Y., Li S., Xing F., Chen L.-L. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol Plant.* 2014;7(9):1494-1496. doi 10.1093/mp/ssu044
- Ly D.N.P., Iqbal S., Fosu-Nyarko J., Milroy S., Jones M.G. Multiplex CRISPR-Cas9 gene-editing can deliver potato cultivars with reduced browning and acrylamide. *Plants (Basel)*. 2023;12(2):379. doi 10.3390/plants12020379
- Madeira F., Madhusoodanan N., Lee J., Eusebi A., Niewielska A., Tivey A.R.N., Lopez R., Butcher S. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* 2024;52(W1):W521-W525. doi 10.1093/nar/gkae241
- Massa G.A., Décima Oneto C.A., González M.N., Poulsen Hornum A., Arizmendi A., Sucar S., Divito S.B., Feingold S.E. CRISPR/Cas9-mediated development of potato varieties with long-term cold storage and bruising resistance. *Biology (Basel)*. 2025;14(4):445. doi 10.3390/biology14040445
- McKenzie M.J., Chen R.K.Y., Harris J.C., Ashworth M.J., Brummell D.A. Post-translational regulation of acid invertase activity by vacuolar invertase inhibitor affects resistance to cold-induced sweetening of potato tubers. *Plant Cell Environ.* 2013;36(1):176-185. doi 10.1111/j.1365-3040.2012.02565.x
- Murashige T., Skoog F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol Plant.* 1962;15(3):473-497. doi 10.1111/j.1399-3054.1962.tb08052.x
- Murata M. Browning and pigmentation in food through the Maillard reaction. *Glycoconj J.* 2021;38:283-292. doi 10.1007/s10719-020-09943-x
- Nicot N., Hausman J.-F., Hoffmann L., Evers D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot.* 2005;56(421):2907-2914. doi 10.1093/jxb/eri285
- Osipov V.S., Zeldner A.G. Potato production in Russia and Ukraine. In: Çalışkan M.E., Bakhsh A., Jabran K. (Eds) *Potato Production Worldwide*. Academic Press, 2023;341-363. doi 10.1016/B978-0-12-822925-5.00023-2
- Pham G.M., Newton L., Wiegert-Rininger K., Vaillancourt B., Douches D.S., Buell C.R. Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *Plant J.* 2017;92(4):624-637. doi 10.1111/tbj.13706
- Richardson D.L., Davies H.V., Ross H.A., Mackay G.R. Invertase activity and its relation to hexose accumulation in potato tubers. *J Exp Bot.* 1990;41(1):95-99. doi 10.1093/jxb/41.1.95
- Shumbe L., Soares E., Muhovski Y., Smit I., Vanderschuren H. Mutation of the *Vinv* 5' UTR regulatory region reduces acrylamide levels in processed potato to reach EU food-safety standards. *Plant Biotechnol J.* 2024;22(10):2738-2740. doi 10.1111/pbi.14400
- Siauw M.O., Ofori I.W., Lutterodt H.E., Ankar-Brewoo G.M. Acrylamide exposure and risks in most frequently consumed foods in a total diet study. *Am J Food Sci Technol.* 2018;6(4):123-137. doi 10.12691/ajfst-6-4-1
- Simakov E.A., Anisimov B.V., Mityushkin A.V., Zhuravlev A.A. Varietal resources for intended potato growing. *Kartofel' i Ovoshchi = Potato and Vegetables.* 2017;11:24-26 (in Russian)
- Slugina M.A., Kochieva E.Z. Variability of a fragment of the acid vacuolar invertase *Pain-1* gene in potato cultivars. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2014;18(4/1):718-723 (in Russian)
- Tareke E., Rydberg P., Karlsson P., Eriksson S., Törnqvist M. Analysis of acrylamide, a carcinogen formed in heated foodstuffs. *J Agric Food Chem.* 2002;50(17):4998-5006. doi 10.1021/jf020302f
- Teper-Bamnolker P., Roitman M., Katar O., Peleg N., Aruchamy K., Suher S., Doron-Faigenboim A., ... Fält A., Volpin H., Hofvander P., Gal-On A., Eshel D. An alternative pathway to plant cold tolerance in the absence of vacuolar invertase activity. *Plant J.* 2023;113(2):327-341. doi 10.1111/tbj.16049

- Toufiq N., Adeyinka O.S., Khan A., Shafique S., Jahan N., Bhatti M.U., Khalid R., Naeem A., Abbas Q., Shafique S., Tabassum B. Multiple transgenic strategies positively regulate cold-induced sweetening in low temperature stored potato tubers. *Potato Res.* 2025;68:69-88. doi 10.1007/s11540-024-09733-0
- Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A., Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002;3(7):research0034. doi 10.1186/gb-2002-3-7-research0034
- Vetchinkina E.M., Komakhina V.V., Vysotskii D.A., Zaitsev D.V., Smirnov A.N., Babakov A.V., Komakhin R.A. Expression of plant antimicrobial peptide *pro-SmAMP2* gene increases resistance of transgenic potato plants to *Alternaria* and *Fusarium* pathogens. *Russ J Genet.* 2016;52:939-951. doi 10.1134/S1022795416080147
- Wu L., Bhaskar P.B., Busse J.S., Zhang R., Bethke P.C., Jiang J. Developing cold-chipping potato varieties by silencing the vacuolar invertase gene. *Crop Sci.* 2011;51(3):981-990. doi 10.2135/cropsci.2010.08.0473
- Wustman R., Struik P.C. The canon of potato science: 35. Seed and ware potato storage. *Potato Res.* 2007;50:351-355. doi 10.1007/s11540-008-9079-0
- Xing H.-L., Dong L., Wang Z.-P., Zhang H.-Y., Han C.-Y., Liu B., Wang X.-C., Chen Q.-J. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* 2014;14:327. doi 10.1186/s12870-014-0327-y
- Yasmeen A., Shakoor S., Azam S., Bakhsh A., Shahid N., Latif A., Shahid A.A., Husnain T., Rao A.Q. CRISPR/Cas-mediated knock-down of vacuolar invertase gene expression lowers the cold-induced sweetening in potatoes. *Planta.* 2022;256:107. doi 10.1007/s00425-022-04022-x
- Zhang C., Xie C., Song B., Liu X., Liu J. RNAi effects on regulation of endogenous acid invertase activity in potato (*Solanum tuberosum* L.) tubers. *Chin J Agric Biotechnol.* 2008;5(2):107-112. doi 10.1017/S1479236208002192
- Zhu X., Richael C., Chamberlain P., Busse J.S., Bussan A.J., Jiang J., Bethke P.C. Vacuolar invertase gene silencing in potato (*Solanum tuberosum* L.) improves processing quality by decreasing the frequency of sugar-end defects. *PLoS One.* 2014;9:e93381. doi 10.1371/journal.pone.0093381
- Zhu X., Chen A., Butler N.M., Zeng Z., Xin H., Wang L., Lv Z., Eshel D., Douches D.S., Jiang J. Molecular dissection of an intronic enhancer governing cold-induced expression of the vacuolar invertase gene in potato. *Plant Cell.* 2024;36(5):1985-1999. doi 10.1093/plcell/koae050
- Zrenner R., Schöler K., Sonnewald U. Soluble acid invertase determines the hexose-to-sucrose ratio in cold-stored potato tubers. *Planta.* 1996;198(2):246-252. doi 10.1007/BF00206250

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

Received November 7, 2025. Revised February 16, 2026. Accepted February 17, 2026.

doi 10.18699/vjgb-26-62

Fungal pathogens of plants: deciphering the mechanisms of Fusarium wilt

M.P. Bankin ¹, A.A. Samsonova ¹, T.A. Rozhmina², M.G. Samsonova  ¹ Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia² Federal Research Center for Bast Fiber Crops, Torzhok, Russia m.g.samsonova@gmail.com

Abstract. *Fusarium oxysporum* (Fo) is among the most dangerous soilborne pathogens, causing Fusarium wilt and root rots in over 100 plant species worldwide. Some pathogen strains can also infect immunocompromised animals and humans. Consequently, studying the molecular mechanisms associated with pathogen virulence and the plant immune response at different stages of disease development is of paramount importance. The process of host recognition by the pathogen and all stages of the infection process involve a wide repertoire of specific signaling molecules, effector proteins, receptor complexes, as well as interconnected and overlapping signaling pathways. Plants, in turn, have evolved a complex defense system to counter this attack: they also possess intricate molecular-level mechanisms that, triggered by pathogen assault, transmit signals to activate a defensive response. In this review, we examine the main currently known molecular mechanisms of Fo-host interaction within the plant-pathogen system: from plant detection and directed hyphal growth driven by chemotropism, to complex interactions at the level of immune response and specific fungal tactics for its suppression. The review includes sections dedicated to the dynamics of plant infection, pathogen genome organization and its genomic diversity, plant immune response and pathogen suppression tactics, as well as an analysis of the main known effector molecules of the pathogen and associated transcription factors. Special emphasis is put on the special form of Fo that infects flax (*Linum usitatissimum* L.).

Key words: *Fusarium oxysporum*; host-pathogen interaction; plant immunity; effectors

For citation: Bankin M.P., Samsonova A.A., Rozhmina T.A., Samsonova M.G. Fungal pathogens of plants: deciphering the mechanisms of Fusarium wilt. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov J Genet Breed*. 2026;30(4):612-624. doi 10.18699/vjgb-26-62

Funding. This work was supported by the Russian Science Foundation grant 23-16-00037.

Грибные патогены растений: расшифровка механизмов фузариозного увядания

М.П. Банкин ¹, А.А. Самсонова ¹, Т.А. Рожмина², М.Г. Самсонова  ¹ Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия² Федеральный научный центр лубяных культур, Торжок, Россия m.g.samsonova@gmail.com

Аннотация. Несовершенные грибы вида *Fusarium oxysporum* (Fo) относятся к числу самых опасных почвенных патогенов, вызывая фузариозное увядание и корневые гнили более чем у 100 видов растений по всему миру. Некоторые штаммы патогена способны также поражать организмы животных и человека с ослабленным иммунитетом. В связи с этим особую важность представляет изучение молекулярных механизмов, ассоциированных с вирулентностью патогена и иммунным ответом растения на разных стадиях развития заболевания. Процесс обнаружения потенциального хозяина патогеном и все стадии развития инфекционного процесса включают в себя широкий репертуар специфических сигнальных молекул, эффекторных белков, рецепторных комплексов, а также взаимосвязанных и перекрывающихся сигнальных путей. Растения, в свою очередь, выработали сложную систему защиты, чтобы противостоять этой атаке: они тоже обладают комплексными механизмами на молекулярном уровне, которые, будучи запущенными в результате атаки патогена, передают сигналы для активации защитного ответа. В данном обзоре мы рассматриваем основные из известных на сегодняшний день молекулярные механизмы взаимодействия Fo с хозяином в системе «растение–патоген»: с момента обнаружения растения и направленного роста гиф патогена, обусловленного реакцией хемотропизма, и до сложных взаимодействий на уровне иммунного ответа и специфических тактик его подавления со стороны гриба. Обзор включает разделы, посвященные динамике заражения растений Fo,

рассмотрению особенностей организации генома патогена и его геномного разнообразия, иммунному ответу растений и тактике его подавления патогеном, а также анализ основных известных эффекторных молекул патогена и связанных с ними транскрипционных факторов. Отдельное внимание уделено специальной форме *Fo*, поражающей растения льна (*Linum usitatissimum* L.).

Ключевые слова: *Fusarium oxysporum*; взаимодействие растение–патоген; иммунитет растений; эффекторы

Introduction

Fusarium wilt is a dangerous plant disease caused by pathogenic strains of the fungus *Fusarium oxysporum* (*Fo*) (Gordon, 2017; Zhang, Ma, 2017). It poses a serious threat to agriculture, and *Fo* ranks among the top ten most destructive fungal plant pathogens worldwide (Kommedahl et al., 1970; Dean et al., 2012; Rozhmina et al., 2022). *Fo* infection can lead to catastrophic crop losses, a prime example being Fusarium wilt of banana (Widinugraheni et al., 2018; Zhang et al., 2024). In addition to Fusarium wilt, some *Fo* strains can also cause root rot (Gargouri Jbir et al., 2024; Ma K. et al., 2024).

The pathogenicity of *Fo* is host-dependent, as strains infecting one plant species typically do not cause disease in others (Kistler, 1997; O'Donnell et al., 1998; Edel-Herrmann, Lecomte, 2019). Based on this host specificity, pathogenic *Fo* strains have been classified into so-called formae speciales (ff. spp.), of which over 100 have been described to date (Dean et al., 2012; Edel-Herrmann, Lecomte, 2019). Besides infecting plants, forms capable of infecting animals and humans are also known (Zhang et al., 2020). *Fo* exhibits a polymorphic lifestyle, varying among genotypes, ranging from soil saprophytes and endophytic strains to specialized parasites (Dean et al., 2012). *Fo* endophytes inhabit tissues of living plants without causing any negative effects on their function and development (de Lamo, Takken, 2020). Another peculiarity is that sexual reproduction has never been observed in *Fo*, despite the presence of conserved mating-type genes (*MAT1-1* or *MAT1-2*) characteristic of sexually reproducing species (Yun et al., 2000; Fayyaz et al., 2023; Zhang et al., 2024). For instance, most pathogenic *Fo* strains infecting flax studied by us contained the *MAT1-2* mating-type locus idiomorph, corresponding to mating type *a*, and only one strain contained the *MAT1-1* idiomorph, corresponding to mating type *alpha* (Logachev et al., 2024). Since *Fo* reproduces asexually, the very fact of the conservation of these genes in the pathogen's genome remains a puzzle.

Host perception and virulence in *Fo* involve several receptors and signaling cascades. The membrane protein Msb2 regulates the Fmk1 signaling pathway, which plays a key role in *Fo* virulence (Pérez-Nadales, Di Pietro, 2011). Members of the Velvet complex (LaeA/VeA/VelB/VelC), Ras proteins, G-protein components,

and the cAMP pathway can also regulate *Fo* virulence (Husaini et al., 2018). The four proteins of the Velvet complex play a critical role in ensuring chromatin accessibility and expression of the beauvericin biosynthesis gene cluster – a depsipeptide mycotoxin. The VeA and LaeA proteins were necessary for full virulence of *Fo* on tomato plants (López-Berges et al., 2013). The Rho1 GTPase of *Fo* plays an important role in maintaining hyphal architecture and infecting plants, but not animals (Martínez-Rocha et al., 2008). Genes for the alpha and beta subunits of G-proteins, *FGA2* and *FGB1*, are involved in the pathogenicity of *Fo* f. sp. *cubense*. *FGA2* directly regulates fungal virulence, whereas *FGB1* is involved in both virulence and development through multiple pathways, including the cAMP-PKA pathway (Guo et al., 2016).

Dynamics of plant infection by *Fo*

The process of plant colonization by the pathogen begins with the detection of the root by *Fo* spores or hyphae. Host plant identification is based on chemotropism – the directed growth of the pathogen along the concentration gradient of chemicals secreted by the plant (Nordzieke et al., 2019). Plant root exudates contain carbohydrates that induce germination of chlamydospores and microconidia of the tomato pathogens *Fo* f. sp. *lycopersici* (*Fol*) and *Fo* f. sp. *radicis-lycopersici* (*Forl*) (Steinkeller et al., 2005; Turrà et al., 2015). Additionally, hyphal chemotropism towards tomato roots is triggered by secreted plant peroxidases (Turrà et al., 2015; Nordzieke et al., 2019; Sridhar et al., 2020).

After spore germination, both endophytic and pathogenic *Fo* strains colonize the plant root system. Fungal hyphae penetrate roots through wounds, cracks in the epidermis, lateral root emergence points, or by direct penetration into the root tip, depending on the *Fo* strain and plant species (Zhou et al., 2020). Hyphae reach the xylem vessels via the root cortex apoplast. Both pathogenic and non-pathogenic strains colonize the cortex, but although the initial colonization pattern is similar, the extent and nature of colonization differ at later stages (Validov et al., 2010). Typically, only pathogenic strains can efficiently penetrate into the xylem vessels, from which they colonize the aboveground tissues (Mes et al., 2000; Benhamou, Garand, 2001; Van der Does et

al., 2019). A possible explanation for this was recently provided by Ayukawa et al. (2021). It was found that colonization of *Arabidopsis* stems by the *Fo* f. sp. *conglutinans* (*Focn*) strain Cong:1-1 is suppressed by the activity of the *CYP79B2/CYP79B3* genes, which control the synthesis of tryptophan-derived secondary metabolites. One *Focn* chromosome, ChrSC10/SC20, containing the linked genes *Six8* and *PSE1*, enabled stem colonization beyond the root xylem vessels in double mutants of these genes.

Fo genome and genetic diversity

To date, dozens of genomes of many *Fo* formae speciales infecting economically important crops have been sequenced: e. g., *fragariae* (Henry et al., 2021), *cepaе* (Armitage et al., 2018), *vasinfectum* (Seo et al., 2020), *ciceris* (Fayyaz et al., 2023), *pisi* (Williams et al., 2016; Jenkins et al., 2021), as well as endophytic strains (Wang B. et al., 2020). All these genomes are compart-

mentalized and exhibit a mosaic architecture, comprising, on the one hand, gene-rich, transposon-poor regions containing highly conserved housekeeping genes, and on the other hand, gene-poor, repeat-rich regions containing rapidly evolving genes associated with virulence (e. g., effector genes) (Dong et al., 2015). Additional characteristics of the variable part of the *Fo* genome can include AT-rich isochores, specific chromatin modifications, and physical organization as accessory chromosomes, which enhances genome compartmentalization and facilitates its rapid evolution (Frantzeskakis et al., 2019; Torres et al., 2020).

Most known *Fo* genomes have around 11 chromosomes belonging to the core genome. The variable part of the genome can be entirely embedded within core chromosomes as extended blocks, making all chromosomes chimeric (*Fo* f. sp. *cubense*) (Zhang et al., 2024), or partially incorporated into chimeric chromosomes and partially represented as separate accessory chromosomes

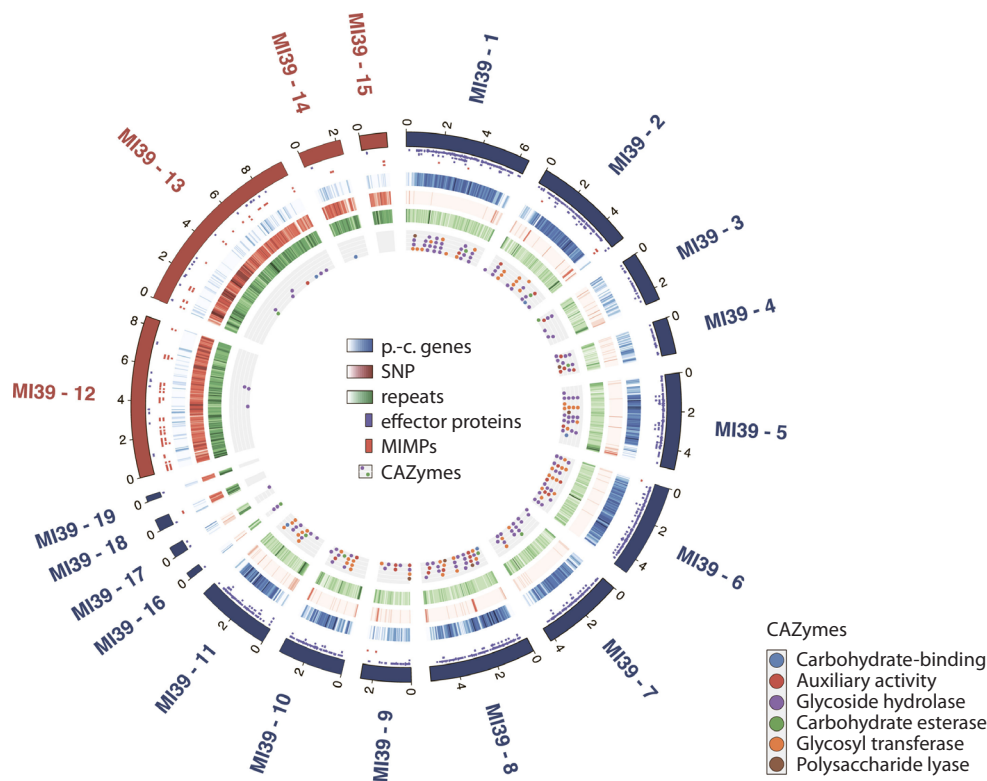


Fig. 1. Organization of the *Fo* f. sp. *lini* genome.

The *Folini* reference genome was assembled from Illumina short and PacBio long DNA reads of the highly virulent strain MI39 (Samsonova et al., 2021). In the Circos diagram, the outer circle depicts the chromosome ideogram, where chromosomes are assigned to the core (blue) or variable (red) genome compartments. The next two rings (purple and red ticks) show genes encoding effectors (necessary for virulence) and genomic positions of MIMP transposons, respectively. The density of protein-coding genes (p-c. genes), single nucleotide polymorphisms, and repetitive elements along the chromosomes is represented as blue, red, and green tracks. The color intensity gradient reflects changes in density. The darkest shade corresponds to the maximum density values. The innermost track (grey with circular glyphs) shows the location of enzymes involved in carbohydrate synthesis and degradation (CAZymes). CAZyme types are denoted as follows: blue – carbohydrate-binding, red – auxiliary activity, purple – glycoside hydrolases, green – carbohydrate esterases, orange – glycosyltransferases, brown – polysaccharide lyases.

(in *Fo* f. sp. *lycopersici*, *Fo* f. sp. *conglutinans*) (Ma L.-J. et al., 2010; Wang Y. et al., 2024). There are genomes lacking chimeric chromosomes, such as *Fo* f. sp. *lini* (Fig. 1) or the endophytic strain Fo47 (Kanapin et al., 2020; Wang B. et al., 2020; Samsonova et al., 2021). The number of accessory chromosomes varies depending on the strain and forma specialis, ranging from 1 to 9. Accordingly, *Fo* genome sizes and the number of protein-coding genes vary within 50–70 Mb and 14,000–21,000, respectively (Ma L.-J. et al., 2010; Williams et al., 2016; Armitage et al., 2018; Kanapin et al., 2020; Seo et al., 2020; Wang B. et al., 2020; Samsonova et al., 2021; Fayyaz et al., 2023; Wang Y. et al., 2024).

Analysis of individual formae speciales genomes revealed high collinearity of core chromosomes and a lack of collinearity between homologous accessory chromosomes (Fig. 2), indicating that accessory chromosomes in individual genomes undergo multiple structural rearrangements (Bates et al., 2024; Logachev et al., 2024). Individual comparisons also demonstrated a high level of variability in the repertoire of individual genes. For example, only 54 % of genes of the entire *Folini* gene repertoire were present in the genomes of 13 isolates of this forma specialis. Moreover, different *Folini* isolates differ in their repertoire of genes encoding effectors and other secreted proteins necessary for infection, allowing them to effectively bypass plant defense mechanisms (Logachev et al., 2024).

Phylogenetic analysis of *Fo* reveals polyphyletic separation, where isolates infecting the same host plant can fall into different clades. For instance, on the maximum likelihood tree based on the EF-1 alpha gene sequence for a group of 50 isolates, including 13 *Folini* genomes and genomes of other formae speciales, isolates infecting flax are found in different clades (Logachev et al., 2024). This means they evolved independently on different genetic backgrounds during evolution. Since *Fo* lacks sexual reproduction, the mechanism for this was elusive for a long time, but it was recently shown that horizontal chromosome transfer is responsible for the emergence of new pathogenic lineages.

Horizontal transfer of accessory or lineage-specific chromosomes has been described in several plant-pathogenic filamentous fungi (Vlaardingerbroek et al., 2016). The possibility of horizontal chromosome transfer in *Fo* was first demonstrated for the tomato-infecting forma specialis, *Fol*. As a result of co-incubation of microconidia from two strains – the pathogenic strain Fol007 and the endophytic strain Fo47, differing in antibiotic resistance, – strains resistant to both antibiotics were selected (Ma L.-J. et al., 2010). These strains were pathogenic when infecting tomato, and contour-clamped homogeneous electric field electrophoresis showed they

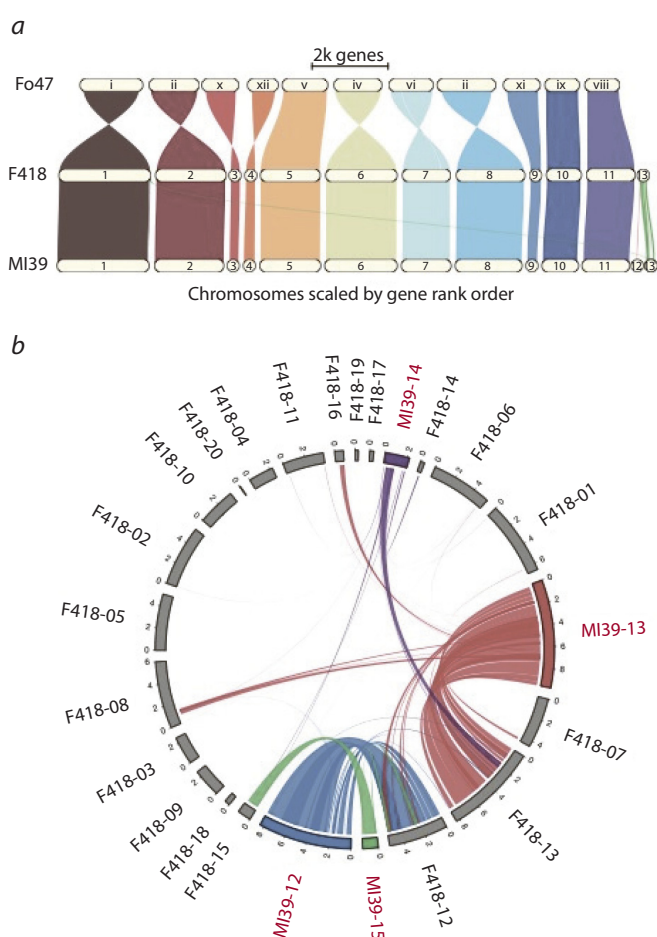


Fig. 2. Comparison of the MI39 and F418 genomic assemblies.

a, Syntenic map for the core genome regions of two *Folini* isolates MI39 and F418, as well as the endophytic strain Fo47. *b*, Circos plot (Krzywinski et al., 2009) showing syntenic regions between the accessory chromosomes of these same strains.

all contained, besides chromosomes from the endophytic strain Fo47, one or two new chromosomes. One of these chromosomes corresponds to accessory chromosome 14 of *Fol*. Horizontal chromosome transfer has now also been demonstrated for another *Fo* forma specialis: *radicis cucumerinum* (*Forc*), pathogenic on cucumber, melon, and watermelon (Vlaardingerbroek et al., 2016; van Dam et al., 2017; Li et al., 2020).

Mechanisms of plant immune response

Plant defense mechanisms include several types of immune responses (Jones, Dangl, 2006; Kourcelis, van der Hoorn, 2018). The first, called PTI (Pattern-Triggered Immunity), is activated by Pathogen-Associated Molecular Patterns (PAMPs) or Damage-Associated Molecular Patterns (DAMPs), which are typically small molecules produced by the pathogen or resulting from plant cell wall destruction (Balint-Kurti, 2019). PAMPs are recognized by Pattern Recognition Receptors (PRRs)

(Kunstler et al., 2016), which are membrane-bound proteins. In response, several signaling events occur, such as activation of the MAP kinase cascade, Ca^{2+} influx into the cytosol, and production of Reactive Oxygen Species (ROS). Consequently, triggering a series of signaling reactions leads to the synthesis of antimicrobial compounds and activation of defense genes.

The second line of defense is initiated by effectors – proteins that can modify cellular targets to suppress PTI (Tsuda, Katagiri, 2010). Effector-Triggered Immunity (ETI) is activated when a plant R-gene protein recognizes an effector encoded by a pathogen avirulence (Avr) gene (Kunstler et al., 2016). This type of interaction, the so-called “gene-for-gene” relationship, was first discovered by Flor (Flor, 1971) when studying the flax/flax rust host-pathogen interaction. Effector interactions with R-gene products elicit a strong defense response called the Hypersensitive Response (HR), often culminating

in localized cell death. PTI and ETI are mutually linked and together potentiate the immune response (Ngou et al., 2021; Nguyen et al., 2021).

Proven intracellular functions of fungal effectors include manipulation of metabolic processes or transcriptional regulators (Djamei et al., 2011; Plett et al., 2014; Tanaka et al., 2014; Weßling et al., 2014). It was recently shown that fungal effectors can also suppress intracellular PTI signaling (Di et al., 2017; Irieda et al., 2019; Navarrete et al., 2022), although the generality of this mechanism for fungi remains to be proven (Tintor et al., 2020).

Tactics of plant immune response suppression by *Fo*

Recently, many new *Fo* effector genes have been identified, and several generalizations have been made regarding their genomic location, presence of their products in the plant, and their mechanism of action (Fig. 3). First

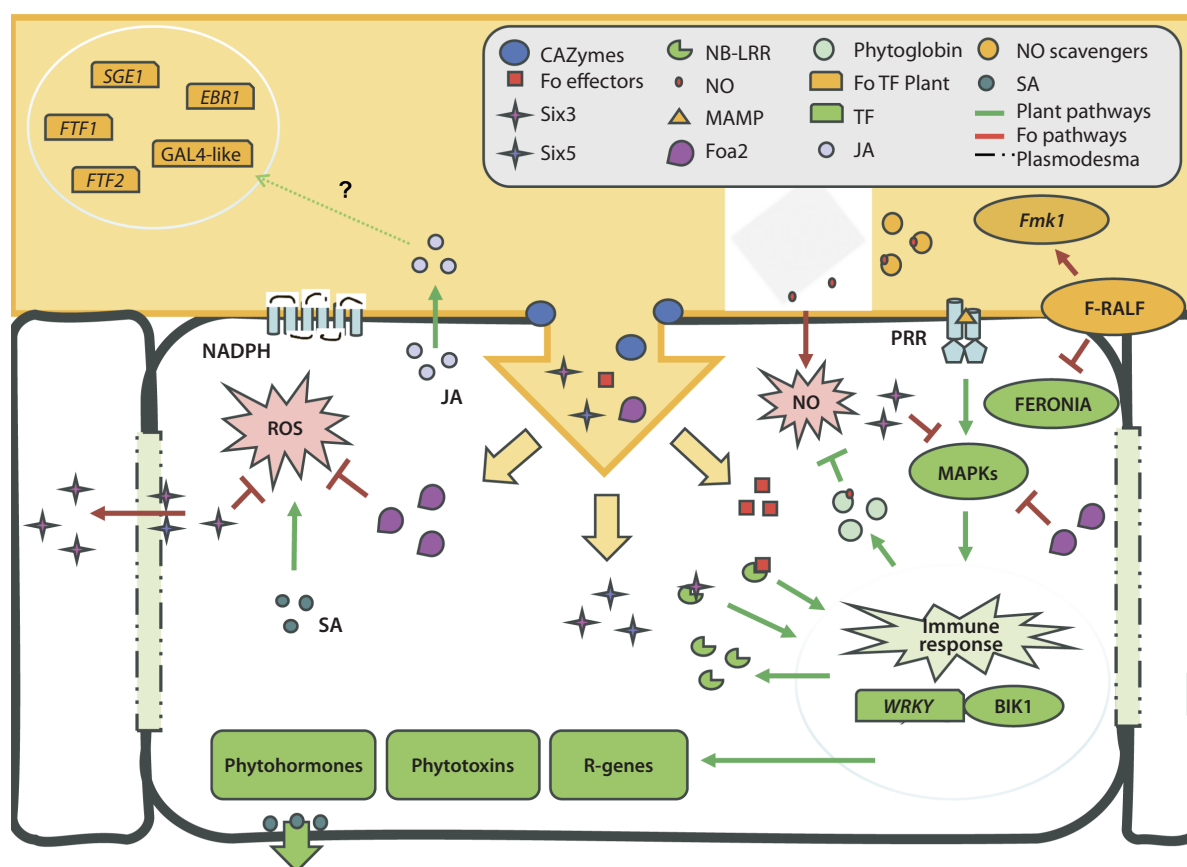


Fig. 3. Schematic representation of *Fo* effector interaction with key components of the plant immune system.

In the figure, red lines indicate the main pathways of pathogen action on the plant, green lines indicate pathways of the plant's response to pathogen invasion. Abbreviations in the legend: CAZymes – enzymes actively degrading and synthesizing carbohydrates; Six3/Six5 – small, cysteine-rich *Fo* effector proteins; NB-LRR/PRR – cytoplasmic and transmembrane proteins encoded by R genes; NO – nitric oxide; MAMP – microbe-associated molecular patterns; Foa2 – *Fo* effector; phytooglobins – plant proteins involved in NO inactivation; Fo TF Plant – *Fo* transcription factors; TF – transcription factors; JA – jasmonic acid; NO scavengers – *Fo* proteins involved in NO inactivation; SA – salicylic acid.

and foremost, it became evident that effector genes are located not only on accessory chromosomes (Sun et al., 2022). Studies of *Folini* genomes showed that most effector genes localize to the core part of the genome and that about 40 % of such genes are not present in the genomes of all strains. This means different *Folini* isolates have different effector repertoires (Logachev et al., 2024).

SIX proteins are the most well-studied class of effectors in *Fo*. These small, cysteine-rich proteins were isolated from the xylem sap of tomato seedlings infected with *Fo* f. sp. *lycopersici* (Rep et al., 2002). *SIX* genes belong to 14 gene families and are located on variable chromosomes. Each forma specialis has a characteristic profile of SIX proteins and gene sequences (De Sain, Rep, 2015). Endophytic *Fo* strains usually have fewer predicted effector genes than pathogenic strains (van Dam et al., 2017; de Lamo, Takken, 2020; Constantin et al., 2021). Interestingly, in our studies, we were unable to detect *SIX* genes in two low-virulence *Folini* strains, F365 and F482 (Logachev et al., 2024). Likely, *Folini* isolates can infect plants even in the absence of *SIX* genes. Nevertheless, other *Folini* genomes, regardless of their pathogenicity status, displayed identical sets of *SIX* gene families, namely *SIX1*, *SIX7*, *SIX10*, *SIX12*, and *SIX13*. Thus, it can be assumed that the *SIX* gene repertoire is conserved in the flax-infecting forma specialis (Logachev et al., 2024).

Loss of *SIX* genes reduces *Fo* pathogenicity. Such results have been obtained for genes *SIX1*, *SIX3*, *SIX5*, *SIX6*, and *SIX8*, primarily in the tomato-infecting *Fo* f. sp. *lycopersici*, and for the *SIX6* gene in *Forc*, the cucumber-infecting forma specialis (Rep et al., 2004; Houterman et al., 2009; Gawehns et al., 2014, 2015; Ma et al., 2015; Ayukawa et al., 2021).

Despite their apparent importance for host plant infection, the functions of most SIX proteins remain unknown (Jangir et al., 2021). The only exception is the *SIX3* (*AVR2*) gene (Houterman et al., 2009), which manipulates plant immunity at the PTI level (Fig. 3). As a virulence factor, *Avr2* suppresses several PTI responses, including ROS accumulation, callose deposition, and MAPK activation. Recent studies in *Arabidopsis thaliana* established that the mechanism of *Avr2* action involves the receptor-like cytoplasmic kinase BOTRYTIS-INDUCED KINASE1 (BIK1). Bacterial flagellin (and its derivative, flg22 peptide) and fungal chitin are two well-studied PAMPs recognized by Pattern Recognition Receptors (PRRs) of *A. thaliana* – FLAGELLIN SENSING 2 (FLS2) and CHITIN ELICITOR RECEPTOR KINASE 1 (CERK1)/LYSINE MOTIF RECEPTOR KINASE 5 (LYK5), respectively. Upon binding flg22, FLS2 forms a signaling complex with its co-receptor BAK1. BAK1 is a leucine-rich repeat receptor kinase.

Within the FLS2-BAK1 complex, BAK1 phosphorylates BIK1, a receptor-like cytoplasmic kinase, at several sites (Lin et al., 2014). BIK1 is then monoubiquitinated by E3 ligases RHA3A and RHA3B, allowing it to dissociate from the FLS2-BAK1 complex into the cytosol (Blekemolen et al., 2023).

The immunoregulatory function of BIK1 depends on both nuclear and cytosolic localization. Nuclear localization of BIK1 is necessary for activating defense gene expression, e. g. by phosphorylating WRKY transcription factors involved in regulating defense hormones – salicylic and jasmonic acids. Cytoplasmic localization of BIK1 upon dissociation from the immune receptor complex on the plasma membrane is necessary, for example, for triggering ROS-mediated signaling through phosphorylation of the RBOHD complex localized on the plasma membrane (Fu et al., 2024). *Avr2* disrupts monoubiquitination of BIK1, and since monoubiquitination of BIK1 is necessary for its internalization, in the presence of *Avr2*, BIK1 remains on the plasma membrane and cannot perform its immunoregulatory function (Blekemolen et al., 2023; Li et al., 2024).

In the tomato pathogen, the effector genes *Six3* and *Six5* share a common promoter located upstream of their coding regions. It was shown that *Six3* and *Six5* localize to both the cytoplasm and nucleus of cells, but their colocalization is detected only at plasmodesmata (Cao et al., 2018). The interaction of *Six5* with *Six3* alters the physical properties of plasmodesmata, allowing *Six3* to move into neighboring cells. In a susceptible plant, this benefits the pathogen and promotes infection spread.

Besides receptor-like kinases, targets of *Fo* effectors at the first level of defense can be the MAP kinase signaling cascade and oxidative stress (Fig. 3). It was shown (Tintor et al., 2020) that stable expression of the effector gene *Foa2* from Fo5176, infecting *A. thaliana*, in cells of this plant blocks accumulation of both ROS and phosphorylated MAP kinases in response to PAMP stimulation.

In *Fo*, nitric oxide (NO) is a key signaling molecule involved in the interaction of the fungus with the host plant (Terrón-Camero et al., 2023) (Fig. 3). *Fusarium oxysporum* f. sp. *cubense* (*Foc*) race R1 of the banana-infecting forma specialis devastated trade in Gros Michel bananas (*Musa acuminata*), and now the tropical race TR4 threatens global production of its replacement – Cavendish bananas (Viljoen et al., 2020). Upon infection with race TR4, plants showed unique induction of genes localized in mitochondria involved in nitric oxide (NO) biosynthesis and detoxification pathways (Zhang et al., 2024). Methyl jasmonate stimulation led to a NO burst in TR4 isolates. Knocking out two highly inducible TR4 genes involved in the NO biosynthesis pathway sig-

nificantly reduced NO production and fungal virulence, indicating that inducing the NO burst directly affects TR4 infectivity. It was suggested that the ability to produce the NO burst, which simultaneously disarms host defense and protects the fungus from toxic environmental impact, was acquired by race TR4 through the emergence of race-specific genes on accessory chromosomal regions that participate in NO production and fungal virulence.

F-RALF proteins are also effectors that suppress host immunity by alkalinizing the environment surrounding plant roots (Fig. 3). During plant infection, *Fo* causes an increase in extracellular pH, leading to activation of the MAPK signaling cascade Fmk1, thereby promoting invasive hyphal growth and virulence (Masachis et al., 2016). Mutants of *Fo* f. sp. *lycopersici* in the *FOXG_21151* gene, encoding an F-RALF protein with pH-raising function, showed reduced ability to colonize the host plant and caused significantly less mortality than the wild type or complemented strain, as well as activation of plant defense reactions (induction of defense genes, accumulation of reactive oxygen species, and callose deposition). In tomato plants, F-RALF action is mediated by interaction with the receptor-like kinase FERONIA, which also interacts with endogenous plant RALF proteins and functions as a negative regulator of immune response (Stegmann et al., 2017).

Some effectors are highly conserved and apparently present in all fungi interacting with plants (Redkar et al., 2022a). Deletion of genes for such effectors, named ERC (Effectors Required for Colonization), in the genome of the pathogenic tomato fungus *Fo4787* reduces its virulence and leads to rapid activation of immune response genes, and in the genome of the endophytic strain *Fo47* it reduces root colonization extent and ability to act as a biocontrol agent. ERC effectors are also involved in infection of the liverwort *Marchantia polymorpha* by strain *Fol4287*.

Besides effectors, genes encoding secondary metabolites (Ito et al., 2004; Coleman et al., 2011), lipid metabolism (He, Dung, 2020) and CAZymes (Carbohydrate-Active enZymes) (Menna et al., 2021), as well as genes controlling fungal penetration (Liu et al., 2019) also play a role in *Fo* virulence.

Transcription factors and effector activity

The transcription factor SGE1 (SIX Gene Expression 1) regulates expression of *SIX* genes in *Fol* *in vivo* (Michielse et al., 2009) (Fig. 3). In other *Fo* formae speciales, SGE1 is necessary for *SIX* gene expression, as well as for genes, the products of which are associated with secondary metabolism (Jonkers et al., 2012; Brown et al., 2014; Zhao S. et al., 2020). The *SGE1* gene is an ortholog of the conserved fungal transcription factor Wor1 from *Candida albicans* and *Histoplasma capsu-*

latum, which regulates morphological transition and is associated with virulence towards humans (Zhao S. et al., 2020).

In addition to *SGE1*, located in the core genome, a group of genes named *FTF* (*Fusarium Transcription Factor*) (Fig. 3) is found on both core and accessory chromosomes of *Fol*. In *Fo* f. sp. *phaseoli*, expression of *FTF1* increases during infection of bean plants and is necessary for fungi pathogenicity (de Vega-Bartol et al., 2011). Knockout of *FTF* genes suggested they regulate pathogenicity mainly by controlling effector expression (Niño-Sánchez et al., 2015). Expression profile analysis showed that transcript levels of *SGE1* and *FTF1* increase during infection processes, and constitutive expression of *FTF1*, *FTF2*, or *SGE1* induces expression of a large overlapping set of known effector genes in *Fol*, indicating interaction between these transcription factors (van der Does et al., 2016).

Another transcription factor in *Fo*, *EBR1* (Enhanced Branching 1), is involved in virulence, similar to its ortholog in *F. graminearum* (Zhao C. et al., 2011; Jonkers et al., 2014). *EBR1* is located on chromosome 7 of *Fo* f. sp. *lycopersici* and is thus part of the core genome, while other *EBR* paralogs are located on accessory chromosomes. *EBR1* mutants showed slower growth in culture and reduced virulence against tomato plants (Niño-Sánchez et al., 2015).

Promising methods for combating Fusarium infection

The development of *Fo* control strategies is complicated by the fact that different isolates of the same forma specialis may use different mechanisms to suppress host immunity (Logachev et al., 2024), as well as by the high evolutionary rate of the accessory genome compartment, allowing pathogens to easily and quickly bypass developed defense mechanisms.

Among modern methods of combating Fusarium wilt, special attention should be paid to biological control methods as the most environmentally friendly, as well as to the use of disease prediction models based on analysis of microbial community composition characteristic of specific soils.

The soil microbiome represents one of the most complex systems directly linked to plant development and health (Raaijmakers, Mazzola, 2016). Bacterial and fungal communities have the ability to influence hormonal signaling in plants, thereby manipulating defense mechanisms (Eichmann et al., 2021), as well as agricultural productivity, forming key ecosystem functions such as nutrient cycling and resistance to plant pathogens (Yuan et al., 2020).

These properties of microbial communities can be used as means to combat Fusarium wilt. For instance, Bubic

et al. (2019) demonstrated the possibility of controlling Fusarium wilt in field conditions with efficacy up to 79 % using *Pseudomonas* spp. strains and up to 70 % using several endophytes and *Trichoderma* spp. strains. Lower biocontrol efficacy (42–55 %) was observed with arbuscular mycorrhizal fungi, *Bacillus* spp., and non-pathogenic *Fusarium* strains (Bubici et al., 2019).

Use of the EGB strain of the myxobacterium *Coral-lococcus* sp. reduced cucumber Fusarium wilt incidence by 9.6 % in greenhouse conditions, and by 66.0 and 53.9 % in field conditions in 2015 and 2016, respectively (Ye et al., 2020). Meanwhile, the rhizosphere microbiota formed in a hydroponic system with multiple parallel mineralization was able to control *Fo* population dynamics but not eliminate the fungal pathogen. The surviving *Fo* in the hydroponic system formed chlamydospores upon contact with the rhizosphere microbiota. The authors suggested that the microbiota suppresses *Fo* spread by controlling pathogen morphogenesis and creating an ecosystem allowing coexistence with *Fo* (Fujiwara et al., 2013). Maximum reduction in pea root rot severity (80 %) under greenhouse conditions was achieved by a synergistic triple treatment consisting of arbuscular mycorrhizal fungi, *Trichoderma harzianum*, and *Pseudomonas fluorescens* (El-Sharkawy et al., 2021).

Application of arbuscular mycorrhiza and *T. harzianum* on tomato plants also revealed their effectiveness in mitigating Fusarium wilt by 45.14 and 44.91 %, respectively, compared to untreated infected plants (Meddad-Hamza et al., 2023). *P. fluorescens* mitigates peanut root rot symptoms even under field conditions and significantly inhibits *Fo* growth (Ren et al., 2024). Studies on the interaction of *Fo* f. sp. *lycopersici* with tomato plants showed that fusaric acid production by the pathogen causes systemic changes in the rhizosphere microbiota and directly influences the recruitment of specific disease-suppressive taxa (Jin et al., 2024).

To date, the most well-known ways to limit the risk of flax Fusarium wilt are the creation of resistant cultivars and suitable crop rotations. However, the emergence of new pathogenic strains requires the development of alternative methods to reduce disease, primarily biological control methods. For example, Planchon et al. (2021) discovered a *Bacillus subtilis* strain possessing biocontrol activity. Using thermogravimetry, the authors showed that this strain, acting cultivar-specifically, promoted the strengthening of stem cell walls.

Mathematical models can be used to predict the potential risk of Fusarium outbreak by identifying key biological indicators and features characteristic of the microbiome of diseased soil. For instance, machine learning can classify *Fo*-infected and non-infected soil samples based on their bacterial and fungal communities (Yuan

et al., 2020). It turned out that the healthy soil microbiome has a higher abundance of *Streptomyces mirabilis*, *Bradyrhizobiaceae*, *Comamonadaceae*, *Mortierella*, and non-pathogenic *Fusarium* strains. The random forest method identified 45 bacterial and 40 fungal OTUs that classified soil health status with over 80 % accuracy (Yuan et al., 2020).

Conclusions

The lack of effective methods to combat Fusarium wilt threatens the production of economically important crops worldwide, including flax. Upon infection, *Fo* penetrates plants directly through the roots via infectious hyphae and then colonizes and spreads within the xylem vessels (Yadeta, Thomma, 2013). Perception of a potential host in *Fo* begins with directed chemotropic growth towards the plant roots (Nordzieke et al., 2019), involving both root exudates inducing germination of *Fo* chlamydospores and microconidia (Steinkeller et al., 2005; Turrà et al., 2015), and secreted plant peroxidases (Turrà et al., 2015; Nordzieke et al., 2019; Sridhar et al., 2020). Many molecular aspects of this phenomenon require further study, for example, concerning the functions and interactions of receptors involved in the early stages of infection (Jiang et al., 2019).

During co-evolution with their hosts, phytopathogens have developed molecular mechanisms allowing them to effectively overcome plant immune responses. To date, significant material has accumulated on signaling cascades activated in response to *Fo* infection (Fig. 3). However, our knowledge of the regulation of this pathogen's interaction with the plant is limited to transcriptome response analysis (van der Does et al., 2016), and the role of DNA methylation and other epigenetic mechanisms in this process remains to be investigated.

A recent meta-analysis of plant transcriptional response to *Fo* infection (Cai et al., 2022) identified hub genes encoding several CAZymes, including xyloglucanase and β -glucosidase. Further research is needed for a deeper understanding of molecular mechanisms in *Fo* concerning carbohydrate metabolism and the enzymes involved. As part of our work, it was shown that carbohydrate-degrading proteins constitute a significant part of the secretome of *Folini* isolates (Logachev et al., 2024).

It is evident that genes encoding important virulence factors are regulated by more than one transcription factor; for example, FTF1 and SGE1 jointly participate in regulating *SIX* family genes (van der Does et al., 2016). *SIX* genes are specifically expressed during infection, but exactly how FTF1 and SGE1 promote *SIX* gene expression during infection is still unknown; thus, a detailed functional analysis of these important virulence-associated transcription factors is necessary.

The question of mechanisms determining host specificity of different *Fo* formae speciales deserves special attention. At first glance, it might seem that since there are many effector proteins, the control of this process should be polygenic. However, it turned out this is not the case (Li et al., 2020). For example, strain Forc016 *Fo* f. sp. *radicis-cucumerinum* can cause disease in cucumber, melon, and watermelon, while strain Fom005 *Fo* f. sp. *melonis* (*Fom*) can only infect melon plants. Forc016 was transformed with the effector gene *g14035*, located on the pathogenic chromosome of Fom005. Transformants had greatly reduced or absent pathogenicity towards cucumber while retaining pathogenicity towards melon and watermelon. This suggests that the protein encoded by *g14035* is recognized by an immune receptor in cucumber plants.

Another result indicating that host switching can be determined by a single gene was obtained by transforming the flax-infecting forma specialis with the *FoPDA1* gene from *Fo* f. sp. *pisi*. This gene encodes a demethylase that detoxifies the phytoalexin pisatin in pea plants (Coleman et al., 2011).

Endophytic interactions of *Fo* with plants are much more common than pathogenic ones (de Lamo, Takken, 2020; de Lamo et al., 2021; Redkar et al., 2022b), yet such interactions are much less studied at the molecular level, and the genetic basis underlying endophytic and pathogenic behavior is unknown.

Currently, no biocontrol agent provides a 100 % level of protection against the harmful effects of *Fo*. The same picture was observed in breeding resistant agricultural cultivars: 100 % resistance is usually provided by a combination of genes. So, improvements of biological protection tools should also move towards creating communities including several bacteria/fungi, each reducing the harmful effect of *Fo*. This imperative, however, elevates the problem of developing biocontrol agents to a new level of complexity, as it requires accounting for and analyzing interactions within a complex multi-component system composed of community members, the plant, the pathogen, other rhizosphere microbes, and the environment.

Thus, despite the significant current knowledge on the molecular mechanisms of *Fo* interaction with plants, numerous questions requiring further study remain.

References

- Armitage A.D., Taylor A., Sobczyk M.K., Baxter L., Greenfield B.P.J., Bates H.J., Wilson F., Jackson A.C., Ott S., Harrison R.J., Clarkson J.P. Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. *Sci Rep*. 2018;8(1):13530. doi 10.1038/s41598-018-30335-7
- Ayukawa Y., Asai S., Gan P., Tsushima A., Ichihashi Y., Shibata A., Komatsu K., Houterman P.M., Rep M., Shirasu K., Arie T. A pair of effectors encoded on a conditionally dispensable chromosome of *Fusarium oxysporum* suppress host-specific immunity. *Commun Biol*. 2021;4(1):707. doi 10.1038/s42003-021-02245-4
- Balint-Kurti P. The plant hypersensitive response: concepts, control and consequences. *Mol Plant Pathol*. 2019;20(8):1163-1178. doi 10.1111/mpp.12821
- Bates H.J., Pike J., Price R.J., Jenkins S., Connell J., Legg A., Armitage A., Harrison R.J., Clarkson J.P. Comparative genomics and transcriptomics reveal differences in effector complement and expression between races of *Fusarium oxysporum* f. sp. *lactucae*. *Front Plant Sci*. 2024;15:1415534. doi 10.3389/fpls.2024.1415534
- Benhamou N., Garand C. Cytological analysis of defense-related mechanisms induced in pea root tissues in response to colonization by nonpathogenic *Fusarium oxysporum* Fo47. *Phytopathology*. 2001;91(8):730-740. doi 10.1094/PHYTO.2001.91.8.730
- Blekemolen M.C., Liu Z., Stegman M., Zipfel C., Shan L., Takken F.L.W. The PTI-suppressing Avr2 effector from *Fusarium oxysporum* suppresses mono-ubiquitination and plasma membrane dissociation of BIK1. *Mol Plant Pathol*. 2023;24(10):1273-1286. doi 10.1111/mpp.13369
- Brown D.W., Busman M., Proctor R.H. *Fusarium verticillioides* SGE1 is required for full virulence and regulates expression of protein effector and secondary metabolite biosynthetic genes. *Mol Plant Microbe Interact*. 2014;27(8):809-823. doi 10.1094/MPMI-09-13-0281-R
- Bubici G., Kaushal M., Prigigallo M.I., Gómez-Lama Cabanás C., Mercado-Blanco J. Biological control agents against *Fusarium* wilt of banana. *Front Microbiol*. 2019;10:616. doi 10.3389/fmicb.2019.00616
- Cai H., Yu N., Liu Y., Wei X., Guo C. Meta-analysis of fungal plant pathogen *Fusarium oxysporum* infection-related gene profiles using transcriptome datasets. *Front Microbiol*. 2022;13:970477. doi 10.3389/fmicb.2022.970477
- Cao L., Blekemolen M.C., Tintor N., Cornelissen B.J.C., Takken F.L.W. The *Fusarium oxysporum* Avr2-Six5 effector pair alters plasmodesmatal exclusion selectivity to facilitate cell-to-cell movement of Avr2. *Mol Plant*. 2018;11(5):691-705. doi 10.1016/j.molp.2018.02.011
- Coleman J.J., Wasmann C.C., Usami T., White G.J., Temporini E.D., McCluskey K., VanEtten H.D. Characterization of the gene encoding pisatin demethylase (*FoPDA1*) in *Fusarium oxysporum*. *Mol Plant Microbe Interact*. 2011;24(12):1482-1491. doi 10.1094/MPMI-05-11-0119
- Constantin M.E., Fokkens L., De Sain M., Takken F.L.W., Rep M. Number of candidate effector genes in accessory genomes differentiates pathogenic from endophytic *Fusarium oxysporum* strains. *Front Plant Sci*. 2021;12:761740. doi 10.3389/fpls.2021.761740
- de Lamo F.J., Takken F.L.W. Biocontrol by *Fusarium oxysporum* using endophyte-mediated resistance. *Front Plant Sci*. 2020;11:37. doi 10.3389/fpls.2020.00037
- de Lamo F.J., Šimkovicová M., Fresno D.H., De Groot T., Tintor N., Rep M., Takken F.L.W. Pattern-triggered immunity restricts host colonization by endophytic fusaria, but does not affect endophyte-mediated resistance. *Mol Plant Pathol*. 2021;22(2):204-215. doi 10.1111/mpp.13018
- De Sain M., Rep M. The role of pathogen-secreted proteins in fungal vascular wilt diseases. *Int J Mol Sci*. 2015;16(10):23970-23993. doi 10.3390/ijms161023970
- De Vega-Bartol J.J., Martín-Dominguez R., Ramos B., García-Sánchez M.-A., Díaz-Minguez J.M. New virulence groups in *Fusarium oxysporum* f. sp. *phaseoli*: the expression of the gene coding for the transcription factor *ftf1* correlates with virulence. *Phytopathology*. 2011;101(4):470-479. doi 10.1094/PHYTO-09-10-0252
- Dean R., Van Kan J.A.L., Pretorius Z.A., Hammond-Kosack K.E., Di Pietro A., Spanu P.D., Rudd J.J., Dickman M., Kahmann R., Ellis J., Foster G.D. The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol*. 2012;13(4):414-430. doi 10.1111/j.1364-3703.2011.00783.x

- Di X., Cao L., Hughes R.K., Tintor N., Banfield M.J., Takken F.L.W. Structure-function analysis of the *Fusarium oxysporum* Avr2 effector allows uncoupling of its immune-suppressing activity from recognition. *New Phytologist*. 2017;216(3):897-914. doi 10.1111/nph.14733
- Djamei A., Schipper K., Rabe F., Ghosh A., Vincon V., Kahnt J., Osorio S., ... Stierhof Y.-D., Schwarz H., Macek B., Mann M., Kahmann R. Metabolic priming by a secreted fungal effector. *Nature*. 2011;478(7369):395-398. doi 10.1038/nature10454
- Dong S., Raffaele S., Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev*. 2015;35:57-65. doi 10.1016/j.gde.2015.09.001
- Edel-Hermann V., Lecomte C. Current status of *Fusarium oxysporum* formae speciales and races. *Phytopathology*. 2019;109(4):512-530. doi 10.1094/PHYTO-08-18-0320-RVW
- Eichmann R., Richards L., Schäfer P. Hormones as go-betweens in plant microbiome assembly. *Plant J*. 2021;105(2):518-541. doi 10.1111/tpl.15135
- El-Sharkawy H.H.A., Abbas M.S., Soliman A.S., Ibrahim S.A., El-Nady I.A.I. Synergistic effect of growth-promoting microorganisms on bio-control of *Fusarium oxysporum* f. sp. *pisi*, growth, yield, physiological and anatomical characteristics of pea plants. *Pestic Biochem Physiol*. 2021;178:104939. doi 10.1016/j.pestbp.2021.104939
- Fayyaz A., Robinson G., Chang P.L., Bekele D., Yimer S., Carrasquilla-Garcia N., Negash K., Surendrarao A., Von Wettberg E.J.B., Kemal S.-A., Tesfaye K., Fikre A., Farmer A.D., Cook D.R. Hiding in plain sight: genome-wide recombination and a dynamic accessory genome drive diversity in *Fusarium oxysporum* f. sp. *ciceris*. *Proc Natl Acad Sci USA*. 2023;120(27):e2220570120. doi 10.1073/pnas.2220570120
- Flor H.H. Current status of the gene-for-gene concept. *Annu Rev Phytopathol*. 1971;9(1):275-296. doi 10.1146/annurev.py.09.090171.001423
- Frantzeskakis L., Kusch S., Panstruga R. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Mol Plant Pathol*. 2019;20(1):3-7. doi 10.1111/mpp.12738
- Fu J., Wang H., Chen Y., Zhang C., Zou Y. The multifaceted ubiquitination of BIK1 during plant immunity in *Arabidopsis thaliana*. *Int J Mol Sci*. 2024;25(22):12187. doi 10.3390/ijms252212187
- Fujiwara K., Iida Y., Iwai T., Aoyama C., Inukai R., Ando A., Ogawa J., Ohnishi J., Terami F., Takano M., Shinohara M. The rhizosphere microbial community in a multiple parallel mineralization system suppresses the pathogenic fungus *Fusarium oxysporum*. *MicrobiologyOpen*. 2013;2(6):997-1009. doi 10.1002/mbo3.140
- Gargouri Jbir T., Zitnick-Anderson K., Pasche J.S., Kalil A.K. Characterization of *Fusarium oxysporum* f. sp. *pisi* associated with root rot of field pea in North Dakota and the effects of temperature on aggressiveness. *Plant Dis*. 2024;108(2):365-374. doi 10.1094/PDIS-05-23-0908-RE
- Gawehns F., Houterman P.M., Ichou F.A., Michiels C.B., Hijdra M., Cornelissen B.J.C., Rep M., Takken F.L.W. The *Fusarium oxysporum* effector Six6 contributes to virulence and suppresses I-2-mediated cell death. *Mol Plant Microbe Interact*. 2014;27(4):336-348. doi 10.1094/MPMI-11-13-0330-R
- Gawehns F., Ma L., Bruning O., Houterman P.M., Boeren S., Cornelissen B.J.C., Rep M., Takken F.L.W. The effector repertoire of *Fusarium oxysporum* determines the tomato xylem proteome composition following infection. *Front Plant Sci*. 2015;6:967. doi 10.3389/fpls.2015.00967
- Gordon T.R. *Fusarium oxysporum* and the *Fusarium* wilt syndrome. *Annu Rev Phytopathol*. 2017;55(1):23-39. doi 10.1146/annurev-phyto-080615-095919
- Guo L., Yang L., Liang C., Wang J., Liu L., Huang J. The G-protein subunits FGA2 and FGB1 play distinct roles in development and pathogenicity in the banana fungal pathogen *Fusarium oxysporum* f. sp. *cubense*. *Physiol Mol Plant Pathol*. 2016;93:29-38. doi 10.1016/j.pmp.2015.12.003
- He M., Ding N.-Z. Plant unsaturated fatty acids: multiple roles in stress response. *Front Plant Sci*. 2020;11:562785. doi 10.3389/fpls.2020.562785
- Henry P.M., Pincot D.D.A., Jenner B.N., Borrero C., Avilés M., Nam M.H., Epstein L., Knapp S.J., Gordon T.R. Horizontal chromosome transfer and independent evolution drive diversification in *Fusarium oxysporum* f. sp. *fragariae*. *New Phytol*. 2021;230:327-340. doi 10.1111/nph.17141. <https://pubmed.ncbi.nlm.nih.gov/33616938/>
- Houterman P.M., Ma L., Van Ooijen G., De Vroomen M.J., Cornelissen B.J.C., Takken F.L.W., Rep M. The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *Plant J*. 2009;58(6):970-978. doi 10.1111/j.1365-3113X.2009.03838.x
- Husaini A.M., Morimoto K., Chandrasekar B., Kelly S., Kaschani F., Palmero D., Jiang J., Kaiser M., Ahrazem O., Overkleeft H.S., Van Der Hoorn R.A.L. Multiplex fluorescent, activity-based protein profiling identifies active α -glycosidases and other hydrolases in plants. *Plant Physiol*. 2018;177(1):24-37. doi 10.1104/pp.18.00250
- Irieda H., Inoue Y., Mori M., Yamada K., Oshikawa Y., Saitoh H., Uemura A., Terauchi R., Kitakura S., Kosaka A., Singkaravanit-Ogawa S., Takano Y. Conserved fungal effector suppresses PAMP-triggered immunity by targeting plant immune kinases. *Proc Natl Acad Sci USA*. 2019;116(2):496-505. doi 10.1073/pnas.1807297116
- Ito S., Eto T., Tanaka S., Yamauchi N., Takahara H., Ikeda T. Tomatidine and lycotetraose, hydrolysis products of α -tomatine by *Fusarium oxysporum* tomatinase, suppress induced defense responses in tomato cells. *FEBS Lett*. 2004;571(1-3):31-34. doi 10.1016/j.febslet.2004.06.053
- Jangir P., Mehra N., Sharma K., Singh N., Rani M., Kapoor R. Secreted in xylem genes: drivers of host adaptation in *Fusarium oxysporum*. *Front Plant Sci*. 2021;12:628611. doi 10.3389/fpls.2021.628611
- Jenkins S., Taylor A., Jackson A.C., Armitage A.D., Bates H.J., Mead A., Harrison R.J., Clarkson J.P. Identification and expression of *Secreted In Xylem* pathogenicity genes in *Fusarium oxysporum* f. sp. *pisi*. *Front Microbiol*. 2021;12:593140. doi 10.3389/fmicb.2021.593140
- Jiang C., Cao S., Wang Z., Xu H., Liang J., Liu H., Wang G., Ding M., Wang Q., Gong C., Feng C., Hao C., Xu J.-R. An expanded subfamily of G-protein-coupled receptor genes in *Fusarium graminearum* required for wheat infection. *Nat Microbiol*. 2019;4(9):1582-1591. doi 10.1038/s41564-019-0468-8
- Jin X., Jia H., Ran L., Wu F., Liu J., Schlaeppi K., Dini-Andreote F., Wei Z., Zhou X. Fusaric acid mediates the assembly of disease-suppressive rhizosphere microbiota via induced shifts in plant root exudates. *Nat Commun*. 2024;15(1):5125. doi 10.1038/s41467-024-49218-9
- Jones J.D.G., Dangl J.L. The plant immune system. *Nature*. 2006;444(7117):323-329. doi 10.1038/nature05286
- Jonkers W., Dong Y., Broz K., Corby Kistler H. The Wor1-like protein Fgpl regulates pathogenicity, toxin synthesis and reproduction in the phytopathogenic fungus *Fusarium graminearum*. *PLoS Pathog*. 2012;8(5):e1002724. doi 10.1371/journal.ppat.1002724
- Jonkers W., Xayamongkhon H., Haas M., Olivain C., Van Der Does H.C., Broz K., Rep M., Alabouvette C., Steinberg C., Kistler H.C. *EBR1* genomic expansion and its role in virulence of *Fusarium* species. *Environ Microbiol*. 2014;16(7):1982-2003. doi 10.1111/1462-2920.12331
- Kanapin A., Samsonova A., Rozhmina T., Bankin M., Logachev A., Samsonova M. The genome sequence of five highly pathogenic isolates of *Fusarium oxysporum* f. sp. *lini*. *Mol Plant Microbe Interact*. 2020;33(9):1112-1115. doi 10.1094/MPMI-05-20-0130-SC
- Kistler H.C. Genetic diversity in the plant-pathogenic fungus *Fusarium oxysporum*. *Phytopathology*. 1997;87(4):474-479. doi 10.1094/PHYTO.1997.87.4.474
- Kommedahl Thor, Christensen J.J., Frederiksen R.A. A half century of research in Minnesota on flax wilt caused by *Fusarium oxysporum*. *Tech Bull Agric Exp Stn*. 1970;273. <https://www.cabidigitallibrary.org/doi/full/10.5555/19711608337>

- Kourelis J., van der Hoorn R.A.L. Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell*. 2018;30(2):285-299. doi 10.1105/tpc.17.00579
- Krzywinski M., Schein J., Birol I., Connors J., Gascoyne R., Horsman D., Jones S.J., Marra M.A. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639-1645. doi 10.1101/gr.092759.109
- Kunstler G., Falster D., Coomes D.A., Hui F., Kooyman R.M., Laughlin D.C., Poorter L., ... Zavala M.A., Zeng H., Zimmerman J.K., Zimmermann N.E., Westoby M. Plant functional traits have globally consistent effects on competition. *Nature*. 2016;529(7585):204-207. doi 10.1038/nature16476
- Li J., Fokkens L., Conneely L.J., Rep M. Partial pathogenicity chromosomes in *Fusarium oxysporum* are sufficient to cause disease and can be horizontally transferred. *Environ Microbiol*. 2020;22(12):4985-5004. doi 10.1111/1462-2920.15095
- Li J., Ai M., Hou J., Zhu P., Cui X., Yang Q. Plant-pathogen interaction with root rot of *Panax notoginseng* as a model: insight into pathogen pathogenesis, plant defence response and biological control. *Mol Plant Pathol*. 2024;25(2):e13427. doi 10.1111/mpp.13427
- Lin W., Li B., Lu D., Chen S., Zhu N., He P., Shan L. Tyrosine phosphorylation of protein kinase complex BAK1/BIK1 mediates *Arabidopsis* innate immunity. *Proc Natl Acad Sci USA*. 2014;111(9):3632-3637. doi 10.1073/pnas.1318817111
- Liu S., Wu B., Yang J., Bi F., Dong T., Yang Q., Hu C., ... Shao C., Chen Y., Yi G., Li C., Guo X. A cerato-platanin family protein FocCP1 is essential for the penetration and virulence of *Fusarium oxysporum* f. sp. *cubense* tropical race 4. *Int J Mol Sci*. 2019;20(15):3785. doi 10.3390/ijms20153785
- Logachev A., Kanapin A., Rozhmina T., Stanin V., Bankin M., Samsonova A., Orlova E., Samsonova M. Pangenomics of flax fungal parasite *Fusarium oxysporum* f. sp. *lini*. *Front Plant Sci*. 2024;15:1383914. doi 10.3389/fpls.2024.1383914
- López-Berges M.S., Hera C., Sulyok M., Schäfer K., Capilla J., Guarro J., Di Pietro A. The velvet complex governs mycotoxin production and virulence of *Fusarium oxysporum* on plant and mammalian hosts. *Mol Microbiol*. 2013;87(1):49-65. doi 10.1111/mmi.12082
- Ma K., Cai L., Wang R., Wang J., Zhan H., Ni H., Lu B., Zhang Y., Gao J. Complete genome sequence of a novel mitovirus isolated from the fungus *Fusarium oxysporum* f. sp. *ginseng* causing ginseng root rot. *Arch Virol*. 2024;169(3):53. doi 10.1007/s00705-024-05962-3
- Ma L., Houterman P.M., Gawehns F., Cao L., Sillo F., Richter H., Clavijo-Ortiz M.J., Schmidt S.M., Boeren S., Vervoort J., Cornelissen B.J.C., Rep M., Takken F.L.W. The *AVR2-SIX5* gene pair is required to activate I-2-mediated immunity in tomato. *New Phytol*. 2015;208(2):507-518. doi 10.1111/nph.13455
- Ma L.-J., Van Der Does H.C., Borkovich K.A., Coleman J.J., Daboussi M.-J., Di Pietro A., Dufresne M., ... Zhou S., Galagan J., Cuomo C.A., Kistler H.C., Rep M. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*. 2010;464(7287):367-373. doi 10.1038/nature08850
- Martínez-Rocha A.L., Roncero M.I.G., López-Ramírez A., Mariné M., Guarro J., Martínez-Cadena G., Di Pietro A. Rho1 has distinct functions in morphogenesis, cell wall biosynthesis and virulence of *Fusarium oxysporum*. *Cell Microbiol*. 2008;10(6):1339-1351. doi 10.1111/j.1462-5822.2008.01130.x
- Masachis S., Segorbe D., Turrà D., Leon-Ruiz M., Fürst U., El Ghalid M., Leonard G., López-Berges M.S., Richards T.A., Felix G., Di Pietro A. A fungal pathogen secretes plant alkalizing peptides to increase infection. *Nat Microbiol*. 2016;1(6):16043. doi 10.1038/nmicrobiol.2016.43
- Meddad-Hamza A., Benzina F., Meddad C., Hamza N., Reghmit A., Ziane H., Ksentini H. Biological control of arbuscular mycorrhizal fungi and *Trichoderma harzianum* against *Fusarium oxysporum* and *Verticillium dahliae* induced wilt in tomato plants. *Egypt J Biol Pest Control*. 2023;33(1):91. doi 10.1186/s41938-023-00737-5
- Menna A., Dora S., Sancho-Andrés G., Kashyap A., Meena M.K., Sklodowski K., Gasperini D., Coll N.S., Sánchez-Rodríguez C. A primary cell wall cellulose-dependent defense mechanism against vascular pathogens revealed by time-resolved dual transcriptomics. *BMC Biol*. 2021;19(1):161. doi 10.1186/s12915-021-01100-6
- Mes J.J., Van Doorn A.A., Wijbrandi J., Simons G., Cornelissen B.J.C., Haring M.A. Expression of the *Fusarium* resistance gene I-2 co-localizes with the site of fungal containment. *Plant J*. 2000;23(2):183-193. doi 10.1046/j.1365-3113x.2000.00765.x
- Michielse C.B., Van Wijk R., Reijnen L., Manders E.M.M., Boas S., Olivain C., Alabouvette C., Rep M. The nuclear protein Sge1 of *Fusarium oxysporum* is required for parasitic growth. *PLoS Pathog*. 2009;5(10):e1000637. doi 10.1371/journal.ppat.1000637
- Navarrete F., Gallei M., Kornienko A.E., Saado I., Khan M., Chia K.-S., Darino M.A., Bindics J., Djamei A. TOPLESS promotes plant immunity by repressing auxin signaling and is targeted by the fungal effector Naked1. *Plant Commun*. 2022;3(2):100269. doi 10.1016/j.xplc.2021.100269
- Ngou B.P.M., Ahn H.-K., Ding P., Jones J.D.G. Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature*. 2021;592(7852):110-115. doi 10.1038/s41586-021-03315-7
- Nguyen Q.-M., Iswanto A.B.B., Son G.H., Kim S.H. Recent advances in effector-triggered immunity in plants: new pieces in the puzzle create a different paradigm. *Int J Mol Sci*. 2021;22(9):4709. doi 10.3390/ijms22094709
- Niño-Sánchez J., Tello V., Casado-Del Castillo V., Thon M.R., Benito E.P., Díaz-Minguez J.M. Gene expression patterns and dynamics of the colonization of common bean (*Phaseolus vulgaris* L.) by highly virulent and weakly virulent strains of *Fusarium oxysporum*. *Front Microbiol*. 2015;6:234. doi 10.3389/fmicb.2015.00234
- Nordzickie D.E., Fernandes T.R., El Ghalid M., Turrà D., Di Pietro A. NADPH oxidase regulates chemotropic growth of the fungal pathogen *Fusarium oxysporum* towards the host plant. *New Phytol*. 2019;224(4):1600-1612. doi 10.1111/nph.16085
- O'Donnell K., Kistler H.C., Cigelnik E., Ploetz R.C. Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. *Proc Natl Acad Sci USA*. 1998;95(5):2044-2049. doi 10.1073/pnas.95.5.2044
- Pérez-Nadales E., Di Pietro A. The membrane mucin Msb2 regulates invasive growth and plant infection in *Fusarium oxysporum*. *Plant Cell*. 2011;23(3):1171-1185. doi 10.1105/tpc.110.075093
- Planchon A., Durambur G., Besnier J.-B., Plasson C., Gügi B., Bernard S., Méricau A., Trouvé J.-P., Dubois C., Laval K., Driouich A., Mollet J.-C., Gattin R. Effect of a *Bacillus subtilis* strain on flax protection against *Fusarium oxysporum* and its impact on the root and stem cell walls. *Plant Cell Environ*. 2021;44(1):304-322. doi 10.1111/pce.13882
- Plett J.M., Daguerre Y., Wittulsky S., Vayssières A., Deveau A., Melton S.J., Kohler A., Morrell-Falvey J.L., Brun A., Veneault-Fourrey C., Martin F. Effector MiSSP7 of the mutualistic fungus *Laccaria bicolor* stabilizes the *Populus* JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proc Natl Acad Sci USA*. 2014;111(22):8299-8304. doi 10.1073/pnas.1322671111
- Raaijmakers J.M., Mazzola M. ECOLOGY. Soil immune responses. *Science*. 2016;352(6292):1392-1393. doi 10.1126/science.aaf3252
- Redkar A., Sabale M., Schudoma C., Zechmann B., Gupta Y.K., López-Berges M.S., Venturini G., Gimenez-Ibanez S., Turrà D., Solano R., Di Pietro A. Conserved secreted effectors contribute to endophytic growth and multihost plant compatibility in a vascular wilt fungus. *Plant Cell*. 2022a;34(9):3214-3232. doi 10.1093/plcell/koac174
- Redkar A., Sabale M., Zuccaro A., Di Pietro A. Determinants of endophytic and pathogenic lifestyle in root colonizing fungi. *Curr Opin Plant Biol*. 2022b;67:102226. doi 10.1016/j.pbi.2022.102226
- Ren J., Cao T., Zang X., Liu J., Yang D. Antifungal mechanisms and characteristics of *Pseudomonas fluorescens*: promoting peanut growth and combating *Fusarium oxysporum*-induced root rot. *Plant Physiol Biochem*. 2024;216:109092. doi 10.1016/j.plaphy.2024.109092

- Rep M., Dekker H.L., Vossen J.H., De Boer A.D., Houterman P.M., Speijer D., Back J.W., De Koster C.G., Cornelissen B.J.C. Mass spectrometric identification of isoforms of PR proteins in xylem sap of fungus-infected tomato. *Plant Physiol.* 2002;130(2):904-917. doi 10.1104/pp.007427
- Rep M., Van Der Does H.C., Meijer M., Van Wijk R., Houterman P.M., Dekker H.L., De Koster C.G., Cornelissen B.J.C. A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for I-3-mediated resistance in tomato. *Mol Microbiol.* 2004;53(5):1373-1383. doi 10.1111/j.1365-2958.2004.04177.x
- Rozhmina T., Samsonova A., Kanapin A., Samsonova M. An account of *Fusarium* wilt resistance in flax *Linum usitatissimum*: the disease severity data. *Data Brief.* 2022;41:107869. doi 10.1016/j.dib.2022.107869
- Samsonova A., Kanapin A., Bankin M., Logachev A., Gretsova M., Rozhmina T., Samsonova M. A genomic blueprint of flax fungal parasite *Fusarium oxysporum* f. sp. *lini*. *Int J Mol Sci.* 2021;22(5):2665. doi 10.3390/ijms22052665
- Seo S., Pokhrel A., Coleman J.J. The genome sequence of five genotypes of *Fusarium oxysporum* f. sp. *vasinfectum*: a resource for studies on *Fusarium* wilt of cotton. *Mol Plant Microbe Interact.* 2020;33(2):138-140. doi 10.1094/MPMI-07-19-0197-A
- Sridhar P.S., Trofimova D., Subramaniam R., González-Peña Fundora D., Foroud N.A., Allingham J.S., Loewen M.C. Ste2 receptor-mediated chemotropism of *Fusarium graminearum* contributes to its pathogenicity against wheat. *Sci Rep.* 2020;10(1):10770. doi 10.1038/s41598-020-67597-z
- Stegmann M., Monaghan J., Smakowska-Luzan E., Rovenich H., Lehner A., Holton N., Belkadir Y., Zipfel C. The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. *Science.* 2017;355(6322):287-289. doi 10.1126/science.aal2541
- Steinkellner S., Mammerler R., Vierheilig H. Microconidia germination of the tomato pathogen *Fusarium oxysporum* in the presence of root exudates. *J Plant Interact.* 2005;1(1):23-30. doi 10.1080/17429140500134334
- Sun X., Fang X., Wang D., Jones D.A., Ma L. Transcriptome analysis of *Fusarium*-tomato interaction based on an updated genome annotation of *Fusarium oxysporum* f. sp. *lycopersici* identifies novel effector candidates that suppress or induce cell death in *Nicotiana benthamiana*. *J Fungi (Basel).* 2022;8(7):672. doi 10.3390/jof8070672
- Tanaka S., Brefort T., Neidig N., Djamei A., Kahnt J., Vermerris W., Koenig S., Feussner K., Feussner I., Kahmann R. A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. *eLife.* 2014;3:e01355. doi 10.7554/eLife.01355
- Terrón-Camero L.C., Molina-Moya E., Peláez-Vico M.Á., Sandoval L.M., Romero-Puertas M.C. Nitric oxide and globin Gb1 regulate *Fusarium oxysporum* infection of *Arabidopsis thaliana*. *Antioxidants.* 2023;12(7):1321. doi 10.3390/antiox12071321
- Tintor N., Paauw M., Rep M., Takken F.L.W. The root-invading pathogen *Fusarium oxysporum* targets pattern-triggered immunity using both cytoplasmic and apoplastic effectors. *New Phytol.* 2020;227(5):1479-1492. doi 10.1111/nph.16618
- Torres D.E., Oggenfuss U., Croll D., Seidl M.F. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biol Rev.* 2020;34(3):136-143. doi 10.1016/j.fbr.2020.07.001
- Tsuda K., Katagiri F. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Curr Opin Plant Biol.* 2010;13(4):459-465. doi 10.1016/j.pbi.2010.04.006
- Turrà D., El Ghalid M., Rossi F., Di Pietro A. Fungal pathogen uses sex pheromone receptor for chemotropic sensing of host plant signals. *Nature.* 2015;527(7579):521-524. doi 10.1038/nature15516
- Validov S.Z., Kamilova F.D., Lugtenberg B.J.J. Monitoring of pathogenic and non-pathogenic *Fusarium oxysporum* strains during tomato plant infection. *Microb Biotechnol.* 2011;4(1):82-88. doi 10.1111/j.1751-7915.2010.00214.x
- van Dam P., Fokkens L., Ayukawa Y., van der Gragt M., Ter Horst A., Brankovics B., Houterman P.M., Arie T., Rep M. A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. *Sci Rep.* 2017;7(1):9042. doi 10.1038/s41598-017-07995-y
- van der Does H.C., Fokkens L., Yang A., Schmidt S.M., Langereis L., Lukasiewicz J.M., Hughes T.R., Rep M. Transcription factors encoded on core and accessory chromosomes of *Fusarium oxysporum* induce expression of effector genes. *PLoS Genet.* 2016;12(11):e1006401. doi 10.1371/journal.pgen.1006401
- van der Does H.C., Constantin M.E., Houterman P.M., Takken F.L.W., Cornelissen B.J.C., Haring M.A., van den Burg H.A., Rep M. *Fusarium oxysporum* colonizes the stem of resistant tomato plants, the extent varying with the R-gene present. *Eur J Plant Pathol.* 2019;154(1):55-65. doi 10.1007/s10658-018-1596-3
- Viljoen A., Mostert D., Chiconela T., Beukes I., Fraser C., Dwyer J., Murray H., ... Rose L.J., Beed F., Dusunceli F., Chao C.-P., Molina A. Occurrence and spread of the banana fungus *Fusarium oxysporum* f. sp. *cubense* TR4 in Mozambique. *S Afr J Sci.* 2020;116(11/12). doi 10.17159/sajs.2020/8608
- Vlaardingerbroek I., Beerens B., Rose L., Fokkens L., Cornelissen B.J.C., Rep M. Exchange of core chromosomes and horizontal transfer of lineage-specific chromosomes in *Fusarium oxysporum*. *Environ Microbiol.* 2016;18(11):3702-3713. doi 10.1111/1462-2920.13281
- Wang B., Yu H., Jia Y., Dong Q., Steinberg C., Alabouvette C., Edel-Hermann V., Kistler H.C., Ye K., Ma L.-J., Guo L. Chromosome-scale genome assembly of *Fusarium oxysporum* strain Fo47, a fungal endophyte and biocontrol agent. *Mol Plant Microbe Interact.* 2020;33(9):1108-1111. doi 10.1094/MPMI-05-20-0116-A
- Wang Y., Liu X., Yuan B., Chen X., Zhao H., Ali Q., Zheng M., ... Xu J., Shi J., Wu H., Gao X., Gu Q. *Fusarium graminearum* rapid alkalization factor peptide negatively regulates plant immunity and cell growth via the FERONIA receptor kinase. *Plant Biotechnol J.* 2024;22(7):1800-1811. doi 10.1111/pbi.14303
- Wefling R., Epple P., Altmann S., He Y., Yang L., Henz S.R., McDonald N., ... Weigel D., Schulze-Lefert P., Dangel J.L., Panstruga R., Braun P. Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe.* 2014;16(3):364-375. doi 10.1016/j.chom.2014.08.004
- Widinugraheni S., Niño-Sánchez J., van der Does H.C., van Dam P., García-Bastidas F.A., Subandiyah S., Meijer H.J.G., Kistler H.C., Kema G.H.J., Rep M. A *SIX1* homolog in *Fusarium oxysporum* f. sp. *cubense* tropical race 4 contributes to virulence towards Cavendish banana. *PLoS One.* 2018;13(10):e0205896. doi 10.1371/journal.pone.0205896
- Williams A.H., Sharma M., Thatcher L.F., Azam S., Hane J.K., Sperschneider J., Kidd B.N., ... Saxena R., Pande S., Ma L.-J., Varshney R.K., Singh K.B. Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting *Fusarium oxysporum* formae speciales facilitates identification of candidate effectors. *BMC Genomics.* 2016;17(1):191. doi 10.1186/s12864-016-2486-8
- Yadeta K.A., Thomma B.P.H.J. The xylem as battleground for plant hosts and vascular wilt pathogens. *Front Plant Sci.* 2013;4:97. doi 10.3389/fpls.2013.00097
- Ye X., Li Z., Luo X., Wang W., Li Y., Li R., Zhang B., ... Wang H., Huang Y., Cao H., Cui Z., Zhang R. A predatory myxobacterium controls cucumber *Fusarium* wilt by regulating the soil microbial community. *Microbiome.* 2020;8(1):49. doi 10.1186/s40168-020-00824-x
- Yuan J., Wen T., Zhang H., Zhao M., Penton C.R., Thomashow L.S., Shen Q. Predicting disease occurrence with high accuracy based on soil macroecological patterns of *Fusarium* wilt. *ISME J.* 2020;14(12):2936-2950. doi 10.1038/s41396-020-0720-5
- Yun S.-H., Arie T., Kaneko I., Yoder O.C., Turgeon B.G. Molecular organization of mating type loci in heterothallic, homothallic, and asexual *Gibberella/Fusarium* species. *Fungal Genet Biol.* 2000;31(1):7-20. doi 10.1006/fghi.2000.1226

- Zhang Y., Ma L.-J. Deciphering pathogenicity of *Fusarium oxysporum* from a phylogenomics perspective. In: Townsend J.P., Wang Z. (Eds) *Advances in Genetics*. Vol. 100. Academic Press, 2017;179-209. doi [10.1016/bs.adgen.2017.09.010](https://doi.org/10.1016/bs.adgen.2017.09.010)
- Zhang Y., Yang H., Turra D., Zhou S., Ayhan D.H., Delulio G.A., Guo L., ... Pearlman E., Schwartz D.C., Di Pietro A., Kistler H.C., Ma L.-J. The genome of opportunistic fungal pathogen *Fusarium oxysporum* carries a unique set of lineage-specific chromosomes. *Commun Biol*. 2020;3(1):50. doi [10.1038/s42003-020-0770-2](https://doi.org/10.1038/s42003-020-0770-2)
- Zhang Y., Liu S., Mostert D., Yu H., Zhuo M., Li G., Zuo C., ... Grigoriev I.V., Yi G., Viljoen A., Li C., Ma L.-J. Virulence of banana wilt-causing fungal pathogen *Fusarium oxysporum* tropical race 4 is mediated by nitric oxide biosynthesis and accessory genes. *Nat Microbiol*. 2024;9(9):2232-2243. doi [10.1038/s41564-024-01779-7](https://doi.org/10.1038/s41564-024-01779-7)
- Zhao C., Waalwijk C., de Wit P.J.G.M., van der Lee T., Tang D. EBR1, a novel Zn₂Cys₆ transcription factor, affects virulence and apical dominance of the hyphal tip in *Fusarium graminearum*. *Mol Plant Microbe Interact*. 2011;24(12):1407-1418. doi [10.1094/MPMI-06-11-0158](https://doi.org/10.1094/MPMI-06-11-0158)
- Zhao S., An B., Guo Y., Hou X., Luo H., He C., Wang Q. Label free proteomics and systematic analysis of secretome reveals effector candidates regulated by SGE1 and FTF1 in the plant pathogen *Fusarium oxysporum* f. sp. *cubense* tropical race 4. *BMC Genomics*. 2020;21(1):275. doi [10.1186/s12864-020-6695-9](https://doi.org/10.1186/s12864-020-6695-9)
- Zhou F., Emonet A., Dénervaud Tendon V., Marhavy P., Wu D., Lahaye T., Geldner N. Co-incidence of damage and microbial patterns controls localized immune responses in roots. *Cell*. 2020;180(3):440-453.e18. doi [10.1016/j.cell.2020.01.013](https://doi.org/10.1016/j.cell.2020.01.013)

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

Received April 11, 2025. Revised May 24, 2025. Accepted June 17, 2025.

doi 10.18699/vjgb-26-63

Using genomic screening methods to create high-quality rice breeding material

N.G. Tumanyan  , Zh.M. Mukhina 

Federal Rice Research Centre of the Ministry of Agriculture of the Russian Federation, Belozerny, Krasnodar, Russia

 tngerag@yandex.ru; agroplazma@gmail.com

Abstract. Rice, as a key model in the study of agroecosystem genomics, is the focus of research meant to address the challenges of producing sufficient food for the growing global population. In breeding programs developing new varieties, improving the physicochemical properties of the grain is crucial. Based on the analysis of national and international research, this article presents information on new molecular genetic methodologies and advances in the development of new valuable rice genotypes using genome sequencing data. Continuous enrichment of rice germplasm at global breeding centers is achieved through the use of highly effective approaches employing postgenomic and cellular technologies in combination with traditional phenotyping methods. This review examines the achievements of molecular genetic research in rice, focusing on valuable grain quality traits such as vitreousness (chalkiness) and shape (size). GWAS analysis is widely used in marker-assisted and genomic rice breeding programs. More recently, GBS analysis has been used to identify relationships between phenotype and genotype based on the analysis of bi-parental mapping populations and varietal accessions. The post-genomic research period, focused on the search for candidate genes for valuable quality traits, had started after the genomic reference sequences were obtained. As a result, hundreds of QTLs for the chalkiness trait were discovered across 12 chromosomes, few were accurately mapped or sequenced. By 2018, several major QTLs affecting grain size were sequenced and characterized. For example, the presence of the recessive *GS3* allele and the dominant *GW7^{TFA}* allele increases the grain length-to-width ratio. In 2023, it was shown that overexpression of *OsFIF3* inhibits the expression of *FLO2* and *SUT1*, thereby increasing chalkiness and reducing grain size. This breeding breakthrough is attributed, for example, to the use of non-digital markers for length, width, thickness, and the grain length-to-width ratio, *GS3RGS1* and *RM505*, as selection markers. All research is an ongoing process aimed at achieving the highest possible level of high-quality rice products.

Key words: rice; selection; genotyping; sequencing; marker; rice quality; vitreousness; grain shape

For citation: Tumanyan N.G., Mukhina Zh.M. Using genomic screening methods to create high-quality rice breeding material. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2026;30(4):625-635. doi 10.18699/vjgb-26-63

Funding. The research was carried out with the financial support of the Russian Science Foundation and Kuban Science Foundation grant No. 25-16-20103, <https://rscf.ru/project/25-16-20103/>. Federal State Budgetary Scientific Institution "Federal Rice Research Centre".

Использование геномных методов скрининга в создании селекционного материала риса с высоким качеством зерна

Н.Г. Туманьян  , Ж.М. Мухина 

Федеральный научный центр риса Министерства сельского хозяйства Российской Федерации, пос. Белозерный, Краснодар, Россия

 tngerag@yandex.ru; agroplazma@gmail.com

Аннотация. Рис как ключевая модель в изучении геномики агроэкосистем находится в центре внимания исследований по решению задач производства достаточного количества продуктов для растущего населения Земли. Качество зерна возделываемых сортов риса лежит в основе его конкурентоспособности и культуры потребления. На основе анализа отечественных и зарубежных исследований представлены сведения о молекулярно-генетических методах и достижениях в создании новых ценных генотипов риса при использовании данных секвенирования геномов. Непрерывное обогащение генплазмы риса в мировых селекционных центрах происходит благодаря высокоэффективным подходам с применением постгеномных и клеточных технологий в сочетании с традиционными методами фенотипирования. В обзоре рассматриваются достижения молекулярно-генетических исследований риса, относящихся к ценным признакам качества зерна – его стекловидности (мучнистости) и форме (размерам). В программах по маркер-ориентированной и геномной селекции риса широко используется анализ GWAS. В последние годы генотипирование методом GBS стало широко применяться

для выявления взаимосвязи между фенотипом и генотипом на основе анализа двуродительских картирующих популяций и выборок сортообразцов. Постгеномный период исследований, связанный с поиском генов-кандидатов ценных признаков качества, начался после получения референсных последовательностей геномов. В результате были обнаружены сотни QTL признака мучнистости по 12 хромосомам, немногие из них были картированы или секвенированы. К 2018 г. секвенировано и охарактеризовано несколько основных QTL, влияющих на размер зерна. Наличие рецессивного аллеля *GS3* и доминантного аллеля *GW7^{TFA}* увеличивает соотношение длины зерна к ширине. В 2023 г. было показано, что сверхэкспрессия *OsFIF3* ингибирует экспрессию *FLO2* и *SUT1*, тем самым увеличивает мучнистость зерна и снижает его размеры. Селекционный прорыв связывают с использованием неспецифических маркеров признаков длины, ширины, толщины, отношения длины зерновки к ширине, *GS3RG51* и *RM505* в качестве маркеров выбора. Все исследования представляют собой непрерывный процесс с целью выхода на максимально возможный уровень реализации высоких параметров качества продукции из риса.

Ключевые слова: рис; селекция; генотипирование; секвенирование; маркер; качество риса; стекловидность зерна; форма зерна

Introduction

Rice, a valuable cereal crop, has ranked first or second among cereals in Russia in recent years and is used in both traditional and dietary nutrition. Rice is a staple crop for half the world population, and it is a key model for studying the genomics of agroecosystems, making it a focus of research aimed at solving the problem of producing sufficient food for the growing world population (Wing et al., 2018). To date, the origin of rice has not been definitively established: centers of its cultivation have been discovered in Korea, India, and China.

Global rice consumption is approximately 500 million tons per year, ranging from 0.01–0.9 kg/person in Lebanon and Latvia to 230–280 kg/person in Vietnam and Myanmar, respectively. Approximately 90 % of rice is produced in traditional rice-consuming countries – East and Southeast Asia.

Rice consumption in Russia is approximately 670,000 tons per year. Russian rice production, being a significant contributor to national agricultural production (eight regions produce rice), ensures the country's food security in terms of rice.

The economic value of rice is determined by the profitability of producing the raw material (rice grain) and rice products. Among the factors determining this, the grain physicochemical properties, including those that contribute to its nutritional and culinary qualities (ECQ), are crucial. Therefore, the grain quality of cultivated rice varieties underlies consumer culture and its competitiveness.

Rice quality parameters required for production and consumption form the basis of breeding programs for developing new varieties. Rice quality is a complex, integrated indicator determined by numerous physical properties of the grain (Alyoshin E.P., Alyoshin N.E., 1997). The absolute requirements that promising rice varieties must meet are high quality performance and compliance with production and consumer requirements. The concept of rice quality includes a large number

of morphological, physicochemical and biochemical characteristics, which, as a rule, are the subject of fundamental, applied and interdisciplinary research (Han et al., 2004; Koutroubas et al., 2004; Addison et al., 2015; Odoo et al., 2021).

The set of physicochemical grain quality traits includes a wide variety of parameters assessed in scientific breeding programs and production processes: moisture content, grain calorific value, grain vitreousness, density, hardness, strength, grain cracking, total grain yield, whole kernel content, chaff, bulk density, etc. Morphological traits include size (length, width, thickness), shape (caryopsis length-to-width ratio), weight, awns, pubescence, caryopsis ribbing, and other attributes.

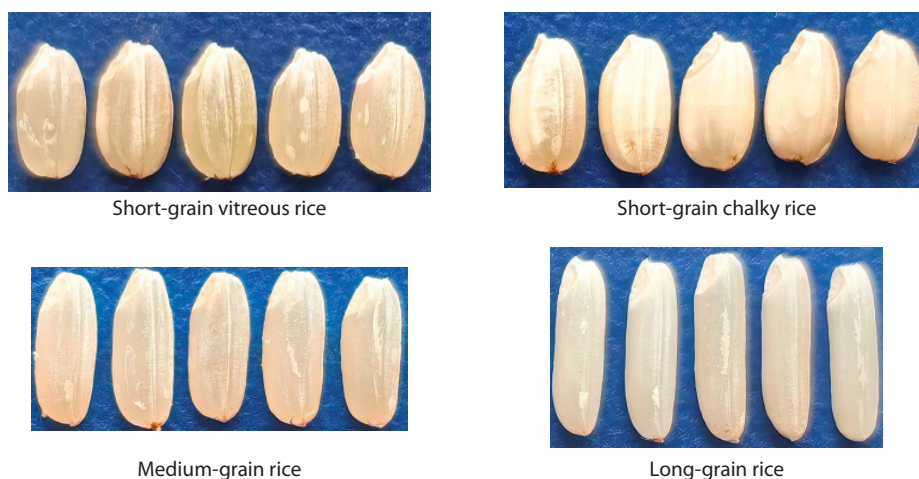
Achievements in international agricultural plant breeding are largely determined by the widespread use of modern biotechnologies in breeding programs, such as whole-genome sequencing, sequencing of plant genome regions within large-scale genotyping, targeted genotyping, haploid technologies, and others.

Rice grain quality.

Vitreousness, grain size and kernel shape

As living standards rise, improving rice quality becomes increasingly important. It is widely recognized that worldwide, improvements in quality lag behind yields. Rice quality is determined by numerous physical properties of the grain. The grain appearance and processing parameters are a priori determined by its basic physical properties (vitreousness/chalkiness, grain shape/size, etc.), which are the primary determinants of rice's market value.

Rice vitreousness is defined as the percentage of vitreous grains in the grain mass (see the Figure). Rice chalkiness is a highly undesirable grain quality that develops during grain filling and is one of the key indicators of grain appearance. It not only affects the grain appearance but also its physical and chemical properties, polishing quality, and cooking quality (Peng et al., 2014). Due to



Types of rice by grain vitreousness and grain shape.

the disordered structure of starch granules and air spaces, chalky grains easily break during hulling and polishing, reducing the yield of high-quality rice products. Rice grain chalkiness, including white heart, belly and back, refers to the opaque part of rice endosperm caused by loose starch granules, which may depend on both the cultivar potential and its response to natural and anthropogenic factors (Guo et al., 2011; Peng et al., 2014; Du et al., 2023; Gong et al., 2023).

Grain size assessment involves determining the length, width, and thickness of the grain in mm; and grain shape, as the ratio of grain length-to-width. Grain size, considered a morphological trait, is a key agronomic characteristic associated with yield and quality. Small-grain rice can have a grain length of up to 5 mm, while long-grain rice can have a grain length of over 6.5 mm. In Russia, rice breeding categorizes rice varieties as short-grain (with a grain length-to-width ratio, l/b , of up to 2.0), medium-grain (l/b up to 3.0), and long-grain ($l/b > 3.0$) ones.

Highly vitreous rice is valued more than rice with predominantly chalky grains. However, there are varieties used in traditional Italian and Spanish dishes, such as risotto and paella, that are characterized by a virtually complete absence of vitreous grains.

Therefore, this review aims to present global achievements in genomic rice breeding in terms of rice quality, which have made a significant contribution to the creation of new, sought-after genotypes of cultivated rice (see the Figure).

Biotechnological methods for enriching the germplasm of agricultural crops

Continuous enrichment of rice germplasm in the world breeding centers is based on the use of highly effective approaches using post-genomic and cellular technologies

in combination with traditional phenotyping methods, which ensures the acceleration of the breeding process, high information content of characteristics (traits of interest), and leads to the maximum positive result in the modern realities of the scientific breeding process (Addison et al., 2015; Garkusha et al., 2022).

The genomic approach is very popular abroad in practical agricultural plant breeding projects, and its high cost does not discourage its use due to its high efficiency. The French company Limagrain uses it to develop modern sunflower hybrids, whereas Euralis Semences uses it to develop breeding forms of sorghum crops (grain, sweet sorghum, sorghum-sudangrass hybrids), and Syngenta uses it to develop sunflower, corn, vegetables and other crops. The routine use of genomic screening methods and the creation of breeding material largely determines the breeding success of these companies.

The future of crop breeding is currently linked to the latest advances in crop genome sequencing and assembly technologies (Dmitriev et al., 2022). The development of molecular biology methods for restriction analysis, PCR DNA amplification, and subsequent nucleotide sequence determination by DNA sequencing has already led, at the initial stages, to the use of DNA regions (nucleotide sequences) linked to genes of interest (which determine certain quality traits of an agricultural crop) or located within their structure as a new group of molecular genetic markers.

The use of genome sequencing technologies has led to fundamental results: the identification of genes of interest with the assessment of polymorphism within the species and the development of molecular markers of economically valuable traits (including grain quality traits), on the basis of which marker-assisted selection (MAS) and genomic breeding programs for crops are implemented.

Genome-wide association studies (GWAS) analysis is widely used in marker-assisted and genomic selection programs. The more recent development of GBS (genotyping-by-sequencing) has been used to identify relationships between phenotype and genotype based on the analysis of bi-parental mapping populations and cultivar samples. This method is based on developed SNP chips (SNPs: single-nucleotide polymorphisms) and genotyping methods using NGS sequencing (NGS: next-generation sequencing, a short-read sequencing technology) of selected genome fractions. The resolution of GWAS can reach up to a single nucleotide. Candidate genes are identified, validated, and tagged (Baxla et al., 2025). The large-scale use of high-throughput sequencing (GBS) genotyping technology has been successfully applied to agricultural plants, as confirmed by numerous publications in the international scientific literature.

Genomic selection (GS) used in agricultural crop breeding offers advantages over marker-assisted selection for polygenic selection. While MAS is effective for monogenic trait control, GS involves analyzing a training population using phenotyping and genotyping data, identifying correlations between phenotype and genotype, and selecting genotypes from a set of selection candidates. The optimal number of training populations, markers, and their ratios are key to successful breeding. The introduction of accelerated genomic technologies – GS, NGS, SNP-mediated selection (using markers and GWAS), with accompanying phenotyping and bioinformatics – is already widely used in modern breeding, leading to the creation and use of new crop genotypes in agriculture.

The development of highly efficient sequencing technologies and the acquisition of reference sequences of plant genomes has led to a “post-genomic” period of research: within this period, an effective search for candidate genes of economically valuable traits, new markers, and the use of the obtained data in genomic selection are being carried out.

For the first time in Russia, scientists from Skoltech, together with colleagues from the University of Southern California, the Vavilov Federal Research Center, the Pustovoit All-Russian Research Institute of Oil Crops, and the Agroplasma breeding and seed company, conducted large-scale genotyping of a sunflower collection using the Illumina platform. The authors conducted the genetic analysis and phenotyping of 601 cultivated sunflower lines, which enabled bioinformatics analysis to identify genetic markers for monitoring the fatty acid composition of the oil (Goryunova et al., 2019). Within the framework of the grant from the Russian Science Foundation and the

Kuban Science Foundation (22-16-20015), phenotyping of samples of segregating rice populations with different amylose content and grain color was carried out in order to select SNP candidate markers for rice breeding based on amylose content, followed by sequencing of DNA libraries on an Illumina MiSeq device (Mukhina et al., 2024).

A development by scientists from Novosibirsk is of great importance for the use of biotechnology in breeding programs. Through the efforts of scientists from the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, the Kurchatov Genome Center of the Institute of Cytology and Genetics of the SB RAS, the Novosibirsk State Agrarian University, and Novosibirsk State University, the GBS-DP bioinformatics pipeline was developed based on the Snakemake process management system. The program enables fully automated GBS data analysis, including the installation of necessary software packages and the processing of large datasets (over 400 samples) (Pronozin et al., 2023).

At the Federal State Autonomous Educational Institution of Higher Education “Peter the Great St. Petersburg Polytechnic University”, St. Petersburg, an SNP panel was developed from a set of molecular markers based on single-nucleotide polymorphisms of the soybean genome with the help of which genotypes are identified. It makes it possible to assess the genomic value of soybean breeding material and conduct soybean breeding for a set of economically valuable traits: variety/line productivity, oil and protein content in seeds (Samsonova et al., 2021).

International research on genomic approaches in practical crop breeding projects, including rice, is of interest due to the importance of incorporating the latest global advances into the breeding process for developing national varieties.

Rice is a model plant for studying and understanding aspects of crop genomics and phenomics. The first crop for which the first reference genome sequence was published using NGS technologies (2002) was rice (*Oryza sativa*). Subsequent application of the latest sequencing technologies, which produce reads of 20,000 nucleotide pairs, resolved the problems of polyploidy and high repeat content. As early as 2010, information was published on a whole-genome analysis of 1,536 SNPs using Illumina technologies for genotyping 395 diverse accessions of rice (*O. sativa*).

The analysis was developed based on a subset of SNPs detected using Perlegen hybridization technology and served to validate the majority of the detected SNPs in a high-quality MBML-intersect dataset (Zhao K. et al., 2010). The technology was used on crops such as barley (Yao et al., 2018), wheat varieties and hybrids (Tyrka et

al., 2021), *Brassica napus* L., *Ralstonia solanacearum*, *Papaver bracteatum* (Rahmati et al., 2024), *Vigna radiata* L. (Kumari et al., 2023), *B. napus* L. (Kim et al., 2024), Logan (Thai fruit of the genus *Dimocarpus*) (Riangwong et al., 2023), *Eucommia ulmoides* (grown in China) (Liu C. et al., 2022), *Cucumis sativus* L. (Kishor et al., 2021) and many others in China, South Korea, USA, India, Iran and other countries.

Results from sequencing and resequencing crop genomes have enabled the identification of SNPs and the development of SNP chips used for genotyping multiple genotypes. RNA-seq and exome sequencing technologies also enable the identification of new SNPs.

In the field of genotyping, works on high-throughput sequencing of rice are known by scientific teams from China (Zhu et al., 2018; Eltahawy et al., 2020; Wang Y. et al., 2020; Yuyu et al., 2020; Pan et al., 2021; Yang W. et al., 2021b; Li J. et al., 2022; Du et al., 2023; Gong et al., 2023; Jin et al., 2023; Tang et al., 2023; Wang N. et al., 2023; Wu et al., 2023; Gao J. et al., 2024; Liu J. et al., 2024; Xu et al., 2024; Yang W. et al., 2024), India (Naidu et al., 2017; Ngangkham et al., 2018), the USA (Zhao K. et al., 2010; Kumar et al., 2023), the Philippines (Cuevas et al., 2016), Vietnam (Piao et al., 2023), Indonesia (Susiyanti et al., 2020), Israel (Heifetz, Soller, 2015); international teams from the Philippines, China and Germany (Sreenivasulu et al., 2022), China, the Philippines, France (Qiu et al., 2015), Uruguay, the USA, Argentina (Quero et al., 2018), India and the USA (Kumari et al., 2023), Egypt and China (Ali et al., 2020), etc.

The vast amount of rice genome sequencing and resequencing data has enabled the development of databases for a variety of purposes, including the Genome Variation Map (Song et al., 2018), RiceVarMap (Zhao H. et al., 2015), SNP-Seek database (Alexandrov et al., 2015), RPAN and MBK V1 (Institute of Genetics and Developmental Biology, CAS, 2018).

Based on the sequencing data of 3,010 rice genomes, tools for searching and visualizing rice pan-genomes were developed: Rice Pan-genome Browser (RPAN, <http://cgm.sjtu.edu.cn/3kricedb/>) and Rice Functional and Genomic Breeding (RFGB, <http://www.rmbreeding.cn/Index/>), containing genomic sequences, gene annotations, expression data, and PAV (presence/absence variations). More than 12,000 genes absent from the rice reference genome were included (Sun et al., 2017; Wang H. et al., 2020).

GBS technology was successfully used for genetic determination of the “days to flowering”, “plant height”, and “panicle productivity” traits under drought stress in two mapping rice populations obtained from crossing

the Swarna*2/Dular and IR11N121*2/Aus 196 varieties (Yadav et al., 2019). As a result of genotyping using a custom marker panel of the segregating population of F₂ rice of the Nipponbare variety with wild African rice *O. longistaminata*, 8,154 informative SNP markers were identified and quantitative loci (QTL) responsible for the “panicle number” trait were localized on chromosomes 1, 3, 4, and 8 (Furuta et al., 2017).

Genome editing technology (first used in plants in 2013) based on the CRISPR/Cas9 system (a multiplex genome editing system) allows for point mutations and the insertion and deletion of genes in specific regions of the genome. This breakthrough technology is being used extensively in rice. It has already been applied to multiple target genes in rice, improving the grain quality of breeding material, producing non-transgenic plants, and rapidly introducing genetic diversity into crop breeding.

In 2017, Chinese researchers reported the creation of a vector targeting eight rice agronomic genes using CRISPR/Cas9. Further genetic transformation and DNA sequencing revealed high mutation efficiency of the eight target genes, including in subsequent generations of various gene-editing-associated phenotypes in both homozygous and heterozygous states. These results demonstrate the potential of the CRISPR/Cas9 system for rapidly introducing genetic diversity into crop breeding (Shen et al., 2017).

Inactivation of the corresponding enzymes in this type of transformation made it possible to increase or decrease the amylose content, significantly reduce the content of cadmium ions in the grain; it also led to an increase in the content of stable oleic acid in the grain and imparting aroma (Lu, Zhu, 2017; Sun et al., 2017; Hua et al., 2019; Khlestkina et al., 2019).

Vitreousness/chalkiness traits in rice molecular breeding

Grain appearance is directly related to its properties: grain shape, size, and vitreousness/chalkiness (Guo et al., 2011). Over the past 20 years, international researchers have made significant progress in identifying QTL genes associated with key traits, including appearance, aroma, texture, and nutritional properties. Markers derived from these genetic discoveries have become effective breeding tools for improving rice quality.

Chalkiness trait of rice grain, caused by loose starch granules in the endosperm, reduces its quality and negatively affects the production of rice products. At numerous rice centers in China, the nature of chalky endosperm formation and grain shape have been a key focus of research over the past 10 years. Hundreds of QTLs for the chalky grain trait have been identified. Due

to the genetic complexity and instability of the rice chalky grain trait, only a few have been accurately mapped or sequenced to date. *Chalk5*, a positive regulatory factor particularly expressed in the endosperm and located on chromosome 5 encoding vacuolar pyrophosphatase translocating H⁺⁺, is the first major QTL sequenced for chalky grain. It influences the grain transport system by affecting pH, and also affects the subcellular ultrastructure of the endosperm and the formation of chalky spots (Li Y. et al., 2014).

In addition, several QTLs, such as *qPGWC-7* (Zhou et al., 2009), *qPGWC-8* (Guo et al., 2011), and *qACE-9* (Gao F. et al., 2016), have been fine-mapped, which is useful for gene sequencing and functional analysis. The QTL for the percentage of chalky grain (PCG), *qPCGI*, was localized to a 139 kb region on the long arm of chromosome 1 using segregating populations generated from residual heterozygotes of the Xieqingzao B (XB)/Zhonghui 9308 (9308) cross. The 9308 allele reduced PCG but had little effect on major agronomic traits, which can be used to improve grain appearance without reducing yield (Zhu et al., 2018).

A large number of QTLs for chalky spots in rice grain have been reported across all 12 chromosomes, but few of them have been fine-mapped or sequenced. In 2021, two QTLs for chalky spots were fine-mapped in two single-segment substitution lines (SSSL): 11-09 with a substituted segment from *O. sativa* and HP67-11 with a substituted segment from *O. glaberrima* (Yang W., 2021a).

For single nucleotide polymorphisms, 3,830 markers were mapped to 12 linkage groups with a total length of 2,456.4 cM with an average genetic distance of 0.82 cM. Seven grain quality parameters, including transparency (vitreousness), chalky grains content, and relative chalky spot area per grain, were examined in the F₂ population. A total of 15 QTLs with logarithm of odds (LOD) scores >4 were identified and mapped to chromosomes 6, 7, and 9. These loci include four QTLs for vitreousness and four for chalky content. Of these, only one matched previously published QTLs, while the rest were novel. By comparing the major QTL regions in the rice genome, several key candidate genes reported to play a crucial role in determining grain characteristics were identified. These results will accelerate the fine mapping of these QTLs and pyramiding of QTLs, which will contribute to the genetic improvement of rice quality (Jin et al., 2023).

At the Rice Research Institute, Anhui Academy of Agricultural Sciences and the College of Agronomy, Anhui Agricultural University (Hefei, China), a genome-wide association analysis of the degree of endosperm chalkiness (DEC) and percentage of grains with chalki-

ness (PGWC) was performed by combining 1.2 million SNPs with phenotypic data from 173 rice accessions and identifying thirteen QTLs for DEC and nine for PGWC (Xu et al., 2024).

Scientists from the Guangdong Provincial Laboratory of Molecular Plant Breeding, the State Key Laboratory of Subtropical Agricultural and Bioresource Conservation and Utilization of South China Agricultural University, and the Guangdong Laboratory of Modern Agriculture (Guangzhou) used six QTLs with low chalkiness to generate 17 pyramided lines with 2–4 QTLs. The results showed that chalkiness decreased with an increasing number of QTLs in the pyramided lines. It was concluded that pyramiding QTLs of genotypes with low chalkiness can be used to develop high-quality rice varieties with high grain vitreousness (low chalkiness) (Yang W., 2024).

The previously described QTL for the chalkiness trait *qWCR4* was confirmed in the segregating BC₃F₂ population and mapped to a 35.26 kb region (Shi et al., 2022).

Grain size and shape traits in rice molecular breeding

Grain shape is determined by three traits: grain length, width, and thickness, and simultaneously affects rice yield and quality. Although a large number of major genes regulating grain shape have been sequenced in recent years, many minor genes remain to be discovered.

Grain size is one of the key agronomic traits associated with yield and quality. The major quantitative trait loci *GS3* and *GL3.1* have been shown to play a predominant role in negatively regulating grain size (Yuyu et al., 2020). By 2018, more than 8,500 QTLs regulating various agronomically important rice traits, including grain size, had been mapped using different contrasting segregating populations derived from different parents, including several major QTLs affecting grain size and yield in rice that were sequenced and characterized (Ngangkham et al., 2018).

Association mapping based on 18,824 high-quality markers yielded 38 QTLs for 10 traits, five of which corresponded to known genes or precisely mapped QTLs (Qiu et al., 2015; Quero et al., 2018).

The results obtained by 2019 laid the foundation for studying grain length biodiversity and molecular breeding of different grain types. Analysis of the secondary F₂ population from a cross between the Nipponbare and Z741 rice varieties allowed the identification of 20 QTLs for important agronomic traits. The grain length of Z741 was shown to be controlled by a major QTL (*qKL3*) and a minor QTL (*qKL7*). Candidate gene sequencing revealed

that *qKL3* is a possible allele of *OsPPKL1*, encoding a protein phosphatase involved in brassinosteroid signaling, and *qKL7* is an unspecified QTL; eight QTLs (*qKL3*, *qKL7*, *qRLW3-1*, *qRLW7*, *qPH3-1*, *qKWT3*, *qKWT7*, and *qNPB6*) were confirmed using three selected single-segment substitution (SSSL) lines S1, S2, and S3; five QTLs (*qKL6*, *qKW3*, *qKW7*, *qKW6*, and *qRLW6*) were detected in S1, S2, and S3 that were not detected in the Nipponbare/Z741 F₂ population. Minor QTLs are likely influenced by environmental factors, since QTLs *qNPB3*, *qNPB7*, and *qPL3* were not validated by three SSSLs (Wang H. et al., 2020).

A study of the genetic effects of grain shape, amylose content, and other quality traits at the Laboratory of Crop Genetic Improvement and Biotechnology and the Key Laboratory of Rice Genetics and Breeding of the Guangxi Academy of Agricultural Sciences revealed that the *gs3*, *GW7*, *TFA*, *gw8*, and *chalk5* genes improve rice appearance. The presence of the *gs3*, *GW7*, and *TFA* alleles increases the rice grain length-to-width ratio, whereas the addition of the *gw8* allele can further reduce grain width.

Genome-wide association studies simultaneously determining size and vitreousness/chalkiness in rice

Much attention is paid to genome-wide association studies of grain appearance (size, shape, chalkiness, content of chalky grains in the total mass) (Zhao D. et al., 2022) of the world collection of rice germplasm and historical breeding populations.

Eight marker–trait associations were identified by screening 124 genotypes with primers specific for grain size traits. Specific markers for length, width, thickness, and grain length-to-width ratio, *GS3RGS1* and *RM505*, have the potential to be used as selection markers in rice breeding (Naidu et al., 2017).

A low correlation with chalkiness was demonstrated at grain widths <2.0 mm, with minimal differences between the *Chalk5* and *chalk5* alleles (the gene regulating chalkiness development in rice) (Gao J. et al., 2024). Scientists from the Guangxi Agricultural Academy (China) and others have shown that *OsMKK3* encodes a kinase of MAP-kinase that controls grain size and chalkiness by influencing cell proliferation in the spikelet coat; *OsSPL16*, *GS5*, and *GIF1* have a significant impact on the *OsMKK3*-regulated grain size pathway (Pan et al., 2021).

A set of recombinant inbred lines (RILs) derived from Gang46B (G46B) and K1075, a G46B introgression line with lower PGWC, was developed. Based on a linkage map containing 33 simple sequence repeat (SSR) markers, a total of 15 additive QTLs were detected, determining six quality traits, including chalkiness and grain shape (Gao F. et al., 2016). A new QTL, *qGL11*, was identified,

potentially encoding *cyclt1*, which plays a specific role in regulating grain length (Li J. et al., 2022).

Researchers at Hunan Agricultural University conducted DAP-seq of multiple rice accessions and showed that overexpression of *OsFIF3* inhibits the expression of *FLO2* and *SUT1*, thereby increasing grain chalkiness and reducing grain size (Tang et al., 2023).

In 2024, it was reported that using backcrosses between the Cypress and KY131 populations, a genetic linkage map was constructed with 151 polymorphic markers, and QTL mapping analysis identified 37 QTLs for grain shape, chalkiness, and amylose content. Among these, three QTLs for grain shape (*qGW3*, *qGL7.2*, and *qLWR7.1*) and four QTLs for grain chalkiness (*qWBR1.2*, *qBR2*, *qWCR3*, and *qWCR11*) demonstrated repeatable and detectable genetic effects across populations. Consequently, *LOC_Os03g45210* may be a candidate gene for *qWCR3*.

The genetic effects of three grain shape QTLs, *qGL7.2*, *qLWR7.2*, and *qGW3*, and three chalkiness QTLs, *qBR2*, *qWCR3*, and *qWCR11*, were then confirmed in a random population. Using the progeny analysis method, the *qWCR3* gene was precisely mapped to a 100-kb region on chromosome 3. The results of this study provide a basis for map sequencing of these QTLs (Liu J. et al., 2024).

Two closely linked quantitative trait loci (QTLs) controlling chalkiness and grain shape were identified on rice chromosome 8 using single-segment substitution lines (SSSLs) (Yang W. et al., 2021b).

In the *qWCR4* region, the *LOC_Os04g50060* and *LOC_Os04g50070* genes showed significant differences in endosperm expression levels between the two NILs, while the other four genes were not expressed. Chalkiness, grain number per plant, seed setting rate, grain width, and thousand-grain mass were increased by *qWCR4^{J23B}* (*qWCR4* allele in the *J23B* gene), contributing to higher yield per plant. The results of these studies provide a basis for *qWCR4* sequencing based on map data (Shi et al., 2022).

In 2023, the results of a study on high-resolution quantitative trait loci mapping using sequencing to determine rice grain quality characteristics, with a particular focus on chalkiness, were published. A genome-wide association study of rice grain shape and chalkiness was conducted using accessions from the Xi'an collection (China). A panel of 137 markers revealed significant genetic variation in six grain quality traits. Twenty-seven QTLs were identified using GWAS. The authors believe that the obtained results improve the understanding of the genetic basis of the physical traits of grain mass, shape, and chalkiness in rice and can be used to provide recommendations for future breeding using markers with

QTLs or gene pyramiding to stabilize and improve rice quality (Wang N. et al., 2023).

A total of 185 recombinant inbred F₁₂ lines (RILS), obtained in the United States by crossing Cypress (HNT-tolerant) and LaGrue (HNT-sensitive), were screened for grain quality traits: grain length, grain width, and relative chalky grain content under control and stress conditions. A total of 15 QTLs were identified on nine chromosomes, which were scanned for natural genetic variation in the Japanese Rice Diversity Panel (JDP) to identify candidate genes for these parameters. A total of 6,160 high-impact SNPs were detected (Kumar et al., 2023).

The mitochondria-associated protein PPR (demonstrated in the stable mutant WBG1) regulates grain chalkiness and size (Wu et al., 2023).

The *RGAI* gene, which encodes the α -subunit of the rice G protein, plays a key role in regulating rice structure, seed size, and response to abiotic stress. Research by Chinese scientists has shown that the *RGAI* gene is involved not only in controlling rice structure and grain size but also in regulating rice quality and seed germination (Yang X. et al., 2024).

Recently, 25 quantitative traits associated with the grain length, width and chalkiness under heat stress, or three sets of overlapping QTLs located on chromosomes 3, 5 and 11, were identified, which can be attributed to the priority candidate genes for QTLs associated with the quality of grain appearance (size and chalkiness under heat stress). Mutants *LOC_Os05g06920*, *LOC_Os05g06970* and *LOC_Os11g28104* had an increased grain length and grain length-to-width ratio, as well as a decreased grain width, including under heat stress. Mutants *LOC_Os05g06970* and *LOC_Os11g28104* significantly reduced grain vitreousness and transparency (Chen L. et al., 2025).

In summary, previous studies have shown that numerous grain quality-related genes affect both grain size and chalkiness in rice (Yang W. et al., 2021b; Zhao D. et al., 2022; Tang et al., 2023; Yang X. et al., 2024). For example, *OsMKK3*, *OsSPL16*, *GS5*, *GIF1* (Pan et al., 2021). Several genes responsible for grain shape, including *gw8*, *GW7*, *lgy3*, and *gs9*, can reduce starch content and significantly improve the grain appearance of rice (Liu Q. et al., 2018; Zhao D. et al., 2018).

Conclusion

The results obtained worldwide to date from the use of highly effective biotechnological methods deepen the understanding of the molecular mechanisms regulating the quality of rice grain. They provide theoretical support for molecular breeding to improve rice quality, in particular through the development of markers of economically

valuable traits and germplasm resources for the genetic improvement of crop plants, including rice, for traits associated with the most important physical properties – vitreousness/chalkiness, shape and size of rice grain.

Hundreds of QTLs for the undesirable grain quality trait – chalkiness – have been identified in rice production centers worldwide. Only a few have been accurately mapped or sequenced, including *Chalk5*, a positively regulated factor expressed in the endosperm located on chromosome 5 and the first major QTL sequenced for chalky grain.

Additionally, the *qPGWC-7*, *qPGWC-8*, and *qACE-9* QTLs were accurately mapped. Eight marker–trait associations were identified by screening genotypes with primers specific for rice grain size traits. Specific markers for length, width, thickness, grain length-to-width ratio, *GS3RGS1*, and *RM505*, were identified and can be used as markers of choice in breeding. Using DAP-seq of multiple rice accessions, it was shown that overexpression of *OsFIF3* inhibits the expression of *FLO2* and *SUT1*, thereby increasing grain chalkiness and reducing grain size. Three QTLs for grain shape (*qGW3*, *qGL7.2*, and *qLWR7.1*) and four QTLs for grain chalkiness (*qWBR1.2*, *qBR2*, *qWCR3*, and *qWCR11*) were identified (Liu J. et al., 2024). Numerous genes associated with grain quality were shown to influence both grain size and chalkiness in rice.

Thus, the level of understanding of the rice genome worldwide is high, and increased attention is being paid to the most important physical parameters of the rice grain at global rice research centers. Research is an ongoing process with a clearly defined goal of achieving the highest possible level of the best-quality rice products.

References

- Addison M., Sarfo-Mensah P., Edusah S.E. Assessing Ghana's initiative of increasing domestic rice production through the development of rice value chain. *Glob Sci Res J.* 2015;3(4):230-237
- Alexandrov N., Tai Sh., Wang W., Mansueto L., Palis K., Fuentes R. SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* 2015;63(2):D1023-D1027. doi 10.1093/nar/gku1039
- Ali N., Li D., Eltahawy M., Abdulmajid D., Bux L., Liu E., Dang X., Hong D. Mining of favorable alleles for seed reserve utilization efficiency in *Oryza sativa* by means of association mapping. *BMC Genetics.* 2020;21(1):4. doi 10.1186/s12863-020-0811-3
- Alyoshin E.P., Alyoshin N.E. Rice. Krasnodar, 1997 (in Russian)
- Baxla B., Thiagarajan K., Swaminathan M., Bellie A., Natarajan S., Govindan S.K. Advances in two-line hybrid rice breeding: leveraging thermosensitive genetic male sterility system in rice for improved global rice production. *Euphytica.* 2025;221:101. doi 10.1007/s10681-025-03550-3
- Chen L., Chen W., Li J., Wei Y., Qing D., Huang J., Yang X., ... Deng G., Dai G., Chen C., Liang T., Pan Y. Identifying heat adaptability QTLs and candidate genes for grain appearance quality at the flowering stage in rice. *Rice.* 2025;18(1):13. doi 10.1186/s12284-025-00770-y

- Cuevas R., Pede V., McKinley J., Velarde O., Demont M. Rice grain quality and consumer preferences: a case study of two rural towns in the Philippines. *PLoS One*. 2016;11(3):e0150345. doi 10.1371/journal.pone.0150345
- Dmitriev A.A., Pushkova E.N., Melnikova N.V. Plant genome sequencing: modern technologies and novel opportunities for breeding. *Mol Biol*. 2022;56(4):495-507. doi 10.1134/s0026893322040045
- Du Y., Long Ch., Deng X., Zhang Zh., Liu J., Xu Y., Xu Y., Liu D., Zeng Y. Physiological basis of high nighttime temperature-induced chalkiness formation during early grain-filling stage in rice (*Oryza sativa* L.). *Agronomy*. 2023;13(6):1475. doi 10.3390/agronomy13061475
- Eltahawy M.S., Ali N., Zaid I.U., Li D., Abdulmajid D., Bux L., Wang H., Hong D. Association analysis between constructed SNPLDBs and GCA effects of 9 quality-related traits in parents of hybrid rice (*Oryza sativa* L.). *BMC Genomics*. 2020;21(1):30. doi 10.1186/s12864-019-6428-0
- Furuta T., Ashikari M., Jena K.K., Doi K., Reuscher S. Adapting genotyping-by-sequencing for rice F₂ populations. *G3: Genes Genomes, Genetics*. 2017;7(3):881-893. doi 10.1534/g3.116.038190
- Gao F., Zeng L., Qiu L., Lu X., Ren J., Wu X., Su X., Gao Y., Ren G. QTL mapping of grain appearance quality traits and grain weight using a recombinant inbred populations in rice (*Oryza sativa* L.). *J Integr Agric*. 2016;15(8):1693-1702. doi 10.1016/S2095-3119(15)61259-X
- Gao J., Gao L., Chen W., Huang J., Qing D., Pan Y., Ma Ch., Wu H., Zhou W., Li J., Yang X., Dai G., Deng G. Genetic effects of grain quality enhancement in Indica hybrid rice: insights for molecular design breeding. *Rice*. 2024;17:39. doi 10.1186/s12284-024-00719-7
- Garkusha S.V., Tumanyan N.G., Mukhina Zh.M., Papulova E.Yu. Phenotyping of varieties of "Collections of genetic resources of rice, vegetable and melon crops" according to the technological characteristics of grain in connection with the development of breeding technology for the creation of rice varieties with high nutritional quality of grain. *Trudy Kubanskogo Gosudarstvennogo Agrarnogo Universiteta = Proceedings of the Kuban State Agrarian University*. 2022;56:59-63. doi 10.21515/1999-1703-98-59-63 (in Russian)
- Gong D., Zhang X., He F., Chen Y., Li R., Yao J., Zhang M., Zheng W., Yu G. Genetic improvements in rice grain quality: a review of elite genes and their applications in molecular breeding. *Agronomy*. 2023;13(5):1375. doi 10.3390/agronomy13051375
- Goryunova S.V., Goryunov D.V., Chernova A.I., Martynova E.U., Dmitriev A.E., Boldyrev S.V., Ayupova A.F., ... Gorlova L.A., Garkusha S.V., Mukhina Z.M., Savenko E.G., Demurin Y.N. Genetic and phenotypic diversity of the sunflower collection of the Pustovoit All-Russia Research Institute of Oil Crops (VNIIMK). *Helia*. 2019;42(70):45-60. doi 10.1515/helia-2018-0021
- Guo T., Liu X., Wan X., Weng J., Liu S., Liu X., Chen M., ... Guo X., Lei C., Wang J., Jiang L., Wan J. Identification of a stable quantitative trait locus for percentage grains with white chalkiness in rice (*Oryza sativa*). *J Integr Plant Biol*. 2011;53(8):598-607. doi 10.1111/j.1744-7909.2011.01041.x
- Han Y., Xu M., Liu X., Yan Ch., Korban S.S., Chen X., Gu M. Genes coding for starch branching enzymes are major contributors to starch viscosity characteristics in waxy rice (*Oryza sativa* L.). *Plant Sci*. 2004;166(2):357-364. doi 10.1016/j.plantsci.2003.09.023
- Heifetz E.M., Soller M. Targeted Recombinant Progeny: a design for ultra-high resolution mapping of Quantitative Trait Loci in crosses between inbred or pure lines. *BMC Genetics*. 2015;16:76. doi 10.1186/s12863-015-0206-z
- Hua K., Zhang J., Botella J.R., Ma Ch., Kong F., Liu B., Zhu J.-K. Perspectives on the application of genome-editing technologies in crop breeding. *Mol Plant*. 2019;12(8):1047-1059. doi 10.1016/j.molp.2019.06.009
- Jin S., Xu L., Yang Q., Zhang M., Wang Sh., Wang R., Tao T., Hong L., Guo Q., Song Sh., Leng Y., Cai X., Gao J. High-resolution quantitative trait locus mapping for rice grain quality traits using genotyping by sequencing. *Front Plant Sci*. 2023;13:860664. doi 10.3389/fpls.2022.860664
- Khlestkina E.K. Rice Genome editing using CRISPR/Cas system. *Biotehnologiya i Seleksiya Rastenij = Plant Biotechnology and Breeding*. 2019;2(1):49-54. doi 10.30901/2658-6266-2019-1-49-54 (in Russian)
- Kim W.J., Yang B., Kim D.G., Kim S.H., Lee Y.J., Kim J., Baek S.H., Kang S.Y., Ahn J.W., Choi Y.J., Bae C.H., Iwar K., Kim S.H., Ryu J. Genotyping-by-sequencing analysis reveals associations between agronomic and oil traits in gamma ray-derived mutant rapeseed (*Brassica napus* L.). *Plants*. 2024;13(11):1576. doi 10.3390/plants13111576
- Kishor D., Lee H.-Y., Hemasundar A., Kim J.-G., Lee S.-Y., Kang B.-Ch., Song K., You C.-R. Identification of an allelic variant of the *CsOr* gene controlling fruit endocarp color in cucumber (*Cucumis sativus* L.) using genotyping-by-sequencing (GBS) and whole-genome sequencing. *Front Plant Sci*. 2021;12:676884. doi 10.3389/fpls.2021.676884
- Koutroubas S., Mazzini F., Pons B., Ntanos D. Grain quality variation and relationships with morphophysiological traits in rice (*Oryza sativa* L.) genetic resources in Europe. *Field Crops Res*. 2004;86:115-130. doi 10.1016/S0378-4290(03)00117-5
- Kumar A., Thomas J., Gill N., Dwiningsih Yh., Ruiz Ch., Famoso A., Pereira A. Molecular mapping and characterization of QTLs for grain quality traits in a RIL population of US rice under high nighttime temperature stress. *Sci Rep*. 2023;13(1):4880. doi 10.1038/s41598-023-31399-w
- Kumari A., Sharma D., Sharma P., Sil S., Wang Ch., Verma V., Patil A., ... Chandel G., Grover A., Jagadish S.V.K., Katiyar-Agarwal S., Agarwal M. Meta-QTL and haplo-pheno analysis reveal superior haplotype combinations associated with low grain chalkiness under high temperature in rice. *Front Plant Sci*. 2023;14:1133115. doi 10.3389/fpls.2023.1133115
- Li J., Yang H., Xu G., Deng K., Yu J., Xiang S., Zhou K., Zhang Q., Li R., Li M., Ling Y., Yang Zh., He G., Zhao F. QTL analysis of Z414, a chromosome segment substitution line with short, wide grains, and substitution mapping of *qGL11* in rice. *Rice*. 2022;15(1):25. doi 10.1186/s12284-022-00571-7
- Li Y., Fan C., Xing Y., Yun P., Luo L., Yan B., Peng B., Xie W., Wang G., Li X., Xiao J., Xu C., He Y. *Chalk5* encodes a vacuolar H(+)-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat Genet*. 2014;46:398-404. doi 10.1038/ng.2923
- Liu C., Wang L., Lu W., Zhong J., Du H., Liu P., Du Q., Du L., Qing J. Construction of SNP-based high-density genetic map using genotyping by sequencing (GBS) and QTL analysis of growth traits in *Eucommia ulmoides* oliver. *Forests*. 2022;13(9):1479. doi 10.3390/f13091479
- Liu J., Zhang H., Wang Y., Liu E., Shi H., Gao G., Zhang Q., Zhang Q., Lou G., Jiang G., He Y. QTL analysis for rice quality-related traits and fine mapping of *qWCR3*. *Int J Mol Sci*. 2024;25(8):4389. doi 10.3390/ijms25084389
- Liu Q., Han R., Wu K., Zhang J., Ye Y., Wang S., Chen J., Pan Y., Li Q., Xu X., Zhou J., Tao D., Wu Y., Fu X. G-protein β subunits determine grain size through interaction with MADS-domain transcription factors in rice. *Nat Commun*. 2018;9:852. doi 10.1038/s41467-018-03047-9
- Lu Y., Zhu J.-K. Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol Plant*. 2017;10(3):523-525. doi 10.1016/j.molp.2016.11.013
- Mukhina Zh.M., Tumanyan N.G., Garkusha S.V., Papulova E.Yu., Chukhir N.P., Chukhir I.N., Gnenniy E.Yu., Esaulova L.V., Malyuchenko E.A., Vakhrusheva N.I. Improving coloured rice grain quality through accelerated breeding. *SABRAO J Breed Genet*. 2024;56(1):89-100. doi 10.54910/sabrao2024.56.1.8
- Naidu B.S.L., Reddy V.L.V., Naik B.J., Naik E.S. Allele mining and allelic diversity of genes governing grain size related traits in rice

- (*Oryza sativa* L.). *Can J Biotechnol.* 2017;1:200. doi 10.24870/cjb.2017-a186
- Ngangkham U., Samantaray S., Yadav M., Kumar A., Chidambaraman P., Katara J.L. Effect of multiple allelic combinations of genes on regulating grain size in rice. *PLoS One.* 2018;13(1):e0190684. doi 10.1371/journal.pone.0190684
- Odom D., Kugbe J., Dzomeku I., Berdjour A., Boateng D., Yaro R., Wireko P., Sam E., Ghanney Ph. Impact of production inputs and timing on crackness of rice in Northern Ghana. *Int J Agronomy.* 2021;4:1-14. doi 10.1155/2021/9982911
- Pan Y., Chen L., Zhao Y., Guo H., Li J., Rashid M.A.R., Lu Ch., ... Qing D., Gao L., Dai G., Li D., Deng G. Natural variation in *OsMKK3* contributes to grain size and chalkiness in rice. *Front Plant Sci.* 2021;12:784037. doi 10.3389/fpls.2021.784037
- Peng B., Wang L., Fan C., Jiang G., Luo L., Li Y., He Y. Comparative mapping of chalkiness components in rice using five populations across two environments. *BMC Genet.* 2014;15:49. doi 10.1186/1471-2156-15-49
- Piao R.H., Chen M.-J., Meng F.-M., Qi C.-Y., Koh H.-J., Gao M.-M., Song A.-Q., Jin Y.-M., Yan Y.-F. Identification and characterization of the chalkiness endosperm gene *CHALK-H* in rice (*Oryza sativa* L.). *J Integr Agric.* 2023;22(10):2921-2933. doi 10.1016/j.jia.2023.04.020
- Pronozin A.Y., Salina E.A., Afonnikov D.A. GBS-DP: a bioinformatics pipeline for processing data coming from genotyping by sequencing. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed.* 2023;27(7):737-745. doi 10.18699/VJGB-23-86
- Qiu X., Pang Y., Yuan Z., Xing D., Xu J., Dingkuhn M., Li Z., Ye G. Genome-wide association study of grain appearance and milling quality in a worldwide collection of indica rice germplasm. *PLoS One.* 2015;10(12):e0145577. doi 10.1371/journal.pone.0145577
- Quero G., Gutiérrez L., Monteverde E., Blanco P., Pérez de Vida F., Rosas J., Fernández S., Garaycochea S., McCouch S., Berberian N., Simondi S., Bonnacarrère V. Genome-wide association study using historical breeding populations discovers genomic regions involved in high-quality. *Plant Genome.* 2018;11(3):170076. doi 10.3835/plantgenome2017.08.0076
- Rahmati R., Nemati Z., Naghavi M., Pfanzt S. Phylogeography and genetic structure of *Papaver bracteatum* populations in Iran based on genotyping-by-sequencing (GBS). *Sci Rep.* 2024;14(1):16309. doi 10.1038/s41598-024-67190-8
- Riangwong K., Saensuk Ch., Pitaloka M.K., Dumhai R., Ruanjai-chon V., Toojinda Th., Wanchana S., Arikrit S. Genetic diversity and population structure of a longan germplasm in Thailand revealed by genotyping-by-sequencing (GBS). *Horticulturae.* 2023;9(6):726. doi 10.3390/horticulturae9060726
- Samsonova M.G., Nuzhdin S.V., Samsonova A.A., Kanapin A.A., Ayupova A.F., Khajtovich F.E., Boldyrev S.V. Markers for Marker Selection of Soya According to Utility Signs. *Invention patent RU 2740798 C1. Publ. 21.01.2021* (in Russian)
- Shen L., Hua Y., Fu Y., Li J., Liu Q., Jiao X., Xin G., Wang J., Wang X., Yan C., Wang K. Rapid generation of genetic diversity by multiplex CRISPR/Cas9 genome editing in rice. *Sci China Life Sci.* 2017;60:506-515. doi 10.1007/s11427-017-9008-8
- Shi H., Zhu Y., Yun P., Lou G., Wang L., Wang Y., Gao G., Zhang Q., Li X., He Y. Fine mapping of *qWCR4*, a rice chalkiness QTL affecting yield and quality. *Agronomy.* 2022;12(3):706. doi 10.3390/agronomy12030706
- Song Sh., Tian D., Li C., Tang B., Dong L., Xiao J., Bao Y., Zhao W., He H., Zhang Zh. Genome Variation Map: a data repository of genome variations in BIG Data Center. *Nucleic Acids Res.* 2018;46(46)(D1):944-949. doi 10.1093/nar/gkx986
- Sreenivasulu N., Zhang C., Tiozon R.N. Jr., Liu Q. Post-genomics revolution in the design of premium quality rice in a high-yielding background to meet consumer demands in the 21st century. *Plant Commun.* 2022;3(3):100271. doi 10.1016/j.xplc.2021.100271
- Sun C., Hu Z., Zheng T., Lu K., Zhao Y., Wang W., Shi J., Wang C., Lu J., Zhang D., Li Z., Wei C. RPAN: rice pan-genome browser for ~3000 rice genomes. *Nucleic Acids Res.* 2017;45(2):597-605. doi 10.1093/nar/gkw958
- Susiyaniti S., Rusmana R., Maryani Y., Sjaifuddin S., Krisdianto N., Syabana M. The physicochemical properties of several Indonesian rice varieties. *Biotropia.* 2020;27(1):41050. doi 10.11598/btb.2020.27.1.1030
- Tang X., Zhong W., Wang K., Gong X., Xia Y., Nong J., Xiao L., Xia S. Regulation of grain chalkiness and starch metabolism by FLO2 interaction factor 3, a bHLH transcription factor in *Oryza sativa*. *Int J Mol Sci.* 2023;24(16):12778. doi 10.3390/ijms241612778
- Tyrka M., Mokrzycka M., Bakera B., Tyrka D., Szeliga M., Stojalowski S., Matysik P., Rokicki M., Rakoczy-Trojanowska M., Krajewski P. Evaluation of genetic structure in European wheat cultivars and advanced breeding lines using high-density genotyping-by-sequencing approach. *BMC Genomics.* 2021;22(1):81. doi 10.1186/s12864-020-07351-x
- Wang H., Zhang J., Farkhanda N., Li J., Sun S., He G., Zhang T., Ling Y., Zhao F. Identification of rice QTL for important agronomic traits with long-kernel CSSL-Z741 and three SSSLs. *Rice Sci.* 2020;27(5):414-423. doi 10.1016/j.rsci.2020.04.008
- Wang N., Chen H., Qian Y., Liang Z., Zheng G., Xiang J., Feng T., Li M., Zeng W., Bao Y., Liu E., Zhang C., Xu J., Shi Y. Genome-wide association study of rice grain shape and chalkiness in a worldwide collection of Xian accessions. *Plants.* 2023;12(3):419. doi 10.3390/plants12030419
- Wang Y., Wang J., Zhai L., Liang Ch., Chen K., Xu J. Identify QTLs and candidate genes underlying source-, sink-, and grain yield-related traits in rice by integrated analysis of bi-parental and natural populations. *PloS One.* 2020;15:e0237774. doi 10.1371/journal.pone.0237774
- Wing R., Purugganan M., Zhang Q. The rice genome revolution: from an ancient grain to Green Super Rice. *Nat Rev Genet.* 2018;19:505-517. doi 10.1038/s41576-018-0024-z
- Wu M., Cai M., Zhai R., Ye J., Zhu G., Yu F., Ye Sh., Zhang X. A mitochondrion-associated PPR protein, WBG1, regulates grain chalkiness in rice. *Front Plant Sci.* 2023;14:1136849. doi 10.3389/fpls.2023.1136849
- Xu Q., Jiang J., Jing C., Hu C., Zhang M., Li X., Shen J., Hai M., Hai M., Zhang Y., Zhang Y., Wang D., Dang X. Genome-wide association mapping of quantitative trait loci for chalkiness-related traits in rice (*Oryza sativa* L.). *Front Genet.* 2024;15:1423648. doi 10.3389/fgene.2024.1423648
- Yadav S., Sandhu N., Singh V., Catolos M., Kumar A. Genotyping-by-sequencing based QTL mapping for rice grain yield under reproductive stage drought stress tolerance. *Sci Rep.* 2019;9(1):14326. doi 10.1038/s41598-019-50880-z
- Yang W., Liang J., Hao Q., Luan X., Tan Q., Lin S., Zhu H., Liu G., Liu Z., Bu S., Wang S., Zhang G. Fine mapping of two grain chalkiness QTL sensitive to high temperature in rice. *Rice.* 2021a;14:33. doi 10.1186/s12284-021-00476-x
- Yang W., Xiong L., Liang J., Hao Q., Luan X., Tan Q., Lin S., Zhu H., Liu G., Liu Z., Bu S., Wang S., Zhang G. Substitution mapping of two closely linked QTLs on chromosome 8 controlling grain chalkiness in rice. *Rice.* 2021b;14:85. doi 10.1186/s12284-021-00526-4
- Yang W., Chen S., Hao Q., Zhu H., Tan Q., Lin S., Chen G., Li Z., Bu S., Liu Z., Liu G., Wang S., Zhang G. Pyramiding of low chalkiness QTLs is an effective way to reduce rice chalkiness. *Rice.* 2024;17(1):4. doi 10.1038/s41598-019-50880-z
- Yang X., Lu J., Shi W., Chen Y., Yu J., Chen S., Zhao D., Huang L., Fan X., Zhang C., Zhang L., Liu Q., Li Q. *RGAL1* regulates grain size, rice quality and seed germination in the small and round grain mutant *srg5*. *BMC Plant Biol.* 2024;24(1):167. doi 10.1186/s12870-024-04864-5

- Yao X., Wu K., Yao Y., Bai Y., Ye J., Chi D. Construction of a high-density genetic map: genotyping by sequencing (GBS) to map purple seed coat color (*Psc*) in hulless barley. *Hereditas*. 2018;155:37. doi 10.1186/s41065-018-0072-6
- Yuyu C., Aike Z., Pao X., Xiaoxia W., Yongrun C., Beifang W., Yue Z., Liaqat S., Shihua C., Liyong C., Yingxin Zh. Effects of *GS3* and *GL3.1* for grain size editing by CRISPR/Cas9 in rice. *Rice Sci*. 2020; 27(5):405-413. doi 10.1016/j.rsci.2019.12.010
- Zhao D., Li Q., Zhang C., Zhang C., Yang Q., Pan L., Ren X., Lu J., Gu M., Liu Q. *GS9* acts as a transcriptional activator to regulate rice grain shape and appearance quality. *Nat Commun*. 2018;9(1):1240. doi 10.1038/s41467-018-03616-y
- Zhao D., Zhang C., Li Q., Liu Q. Genetic control of grain appearance quality in rice. *Biotechnol Adv*. 2022;60:108014. doi 10.1016/j.biotechadv.2022.108014
- Zhao H., Yao W., Ouyang Y., Yang W., Wang G., Lian X., Xing Y., Chen L., Xie W. RiceVarMap: a comprehensive database of rice genomic variations. *Nucleic Acids Res*. 2015;43(D1):D1018-D1022. doi 10.1093/nar/gku894
- Zhao K., Wright M., Kimball J., Eizenga G., McClung A., Kovach M., Tyagi W., Ali M.L., Tung C.W., Reynolds A., Bustamante C.D., McCouch S.R. Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLoS One*. 2010;5(5):e107780. doi 10.1371/journal.pone.0010780
- Zhou L., Chen L., Jiang L., Zhang W., Liu L., Liu X., Zhao Z., Liu S., Zhang L., Wang J., Wan J. Fine mapping of the grain chalkiness QTL *qPGWC-7* in rice (*Oryza sativa* L.). *Theor Appl Genet*. 2009; 118(3):581-590. doi 10.1007/s00122-008-0922-0
- Zhu A., Zhang Y., Zhang Z., Wang B., Xue P., Cao Y., Chen Y., Li Z., Liu Q., Cheng S., Cao L. Genetic dissection of *qPCG1* for a quantitative trait locus for percentage of chalky grain in rice (*Oryza sativa* L.). *Front Plant Sci*. 2018;9:1173. doi 10.3389/fpls.2018.01173

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

Received October 10, 2025. Revised January 14, 2026. Accepted February 16, 2026.

doi 10.18699/vjgb-26-64

Development of purple-grain triticale

N.V. Petrash ^{1, 2}, P.I. Stepochkin ^{1, 2}¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia² 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 pnv11@bionet.nsc.ru

Abstract. Purple-colored grains of cereal crops are characterized by high antioxidant activity. Anthocyanins, polyphenolic compounds found in the pericarp of their grains, have beneficial effects on human health. However, triticale has not yet developed forms with anthocyanin-rich purple grain color. The aim of this work was to obtain new forms of wheat-rye amphiploids with purple grain color using marker-assisted breeding and to compare their anthocyanin content and productivity indicators. Molecular DNA markers were used to determine the genotype of hybrids produced by triticale Sadko (*× Triticosecale* Wittmack) and a purple-colored emmer wheat line 27-3/17 (*T. dicoccum* (Schrank) Schuebl.). Purple-colored F₃ hybrids carried two complementary dominant genes *Pp3* and *Pp-B1* in a homozygous state responsible for the high content of anthocyanins in the grain. In subsequent generations, the wheat-rye amphiploids had a purple grain color. The total anthocyanin content in the whole grain flour of the hybrids ranged from 36 to 529.3 µg/g. The high content was recorded in sample 2-1-6-6 (529.3 µg/g). Samples 2-1-1-4e, 2-1-5-10a and 2-1-6-4b were at the control level (emmer wheat – 382.6 µg/g). The F₅₋₆ hybrid plants had a typical hexaploid triticale phenotype. The spike length and the number of spikelets exceeded those of emmer wheat. The number of spike grains in the hybrids was less than that in the Sadko triticale averaging at 28.0 and 34.4 in 2024 and 2025, respectively. 1,000 grain weight of purple-grained triticale families in 2025 was comparable to the Sadko maternal form and averaged 47.6 g. The yield per unit area of hybrid families (470 g/m²) in 2025 was higher than that of emmer wheat (306 g/m²), but lower than that of Sadko (584 g/m²). Thus, the breeding material of purple-grained triticale forms was obtained, which in a number of ways is similar to the triticale Sadko maternal form, but differs from the paternal form of the purple grain donor emmer wheat.

Key words: triticale; purple grain color; marker-assisted selection; anthocyanin

For citation: Petrash N.V., Stepochkin P.I. Development of purple-grain triticale. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2026;30(4):636-644. doi 10.18699/vjgb-26-64

Funding. This work was supported by grant No. 25-26-20153 from the Russian Science Foundation and the Ministry of Education, Science and Innovation Policy of the Novosibirsk Region (No. 30-2025-000980 dated May 21, 2025).

Acknowledgements. Hybrid propagation was carried out at the Collective Use Center for Plant Reproduction of the Institute of Cytology and Genetics SB RAS and supported by the state budget-funded project No. FWNR-2026-0033.

Создание тритикале с фиолетовой окраской зерна

Н.В. Петраш ^{1, 2}, П.И. Стёпочкин ^{1, 2}¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия² 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 pnv11@bionet.nsc.ru

Аннотация. Зерна злаковых культур с фиолетовой окраской отличаются высокой антиоксидантной активностью благодаря содержанию антоцианов в оболочке и представляют интерес для производства полезных продуктов питания. Однако у тритикале пока не выведены формы с богатой антоцианами фиолетовой окраской зерна. Цель настоящей работы – получение с использованием маркер-вспомогательной селекции новых форм пшенично-ржаных амфиплоидов с фиолетовой окраской зерна и сравнительная оценка их по содержанию антоцианов и показателям продуктивности. Определение генотипа гибридов тритикале Садко (*× Triticosecale* Wittmack) с фиолетовозерной линией полбы 27-3/17 (*T. dicoccum* (Schrank) Schuebl.) проводили с применением молекулярных ДНК-маркеров. У всех гибридов F₃, имеющих фиолетовую окраску зерна, установлено наличие двух комплементарно действующих доминантных генов, *Pp3* и *Pp-B1*, отвечающих за повышенное содержание антоцианов в зерне. В последующих поколениях растения пшенично-ржаных амфиплоидов имели фиолетовую окраску зерна. Общее содержание антоцианов в цельнозерновой муке гибридов варьировало от 36 до 529.3 мкг/г. Высокое содержание определено у образца 2-1-6-6 (529.3 мкг/г). На уровне контроля (полба – 382.6 мкг/г) были образцы 2-1-1-4e, 2-1-5-10a и 2-1-6-4b. Растения гибридов F₅₋₆ обладали типичным фенотипом гексаплоидных тритикале. По длине колоса и количеству колосков превышали показатели полбы. Число зерен колоса у гибридов было меньше, чем у контроля тритикале Садко, и достигло в среднем 28.0 и 34.4 шт. в 2024 и 2025 гг. соответственно. Масса 1000 зерен

фиолетовозерных семей тритикале в 2025 г. была на уровне материнской формы Садко и составила в среднем 47.6 г. Средняя урожайность с единицы площади у гибридов (470 г/м²) в 2025 г. была выше, чем у полбы (306 г/м²), но ниже, чем у Садко (584 г/м²). Таким образом, получен селекционный материал фиолетовозерных форм тритикале, по ряду признаков схожий с материнской формой тритикале Садко, но отличающийся от отцовской формы полбы – донора фиолетовой окраски зерна.

Ключевые слова: тритикале; фиолетовая окраска зерна; маркер-вспомогательная селекция; антоцианы

Introduction

The artificial allopolyploid triticale ($\times$ *Triticosecale* Wittmack) was developed by combining the genomes of wheat (*Triticum* sp.) and rye (*Secale* sp.). This species has existed for less than 150 years. The first fertile octoploid ($2n = 8x = 56$) plants of wheat-rye amphiploid were made in 1888 by the German researcher W. Rimpau (Muntzing, 1974). They had three doubled common wheat genomes (BBA^uA^uDD) and one doubled rye genome (RR). Due to their cytogenetic instability, meiotic disturbances are often observed in primary octoploid triticales resulting in the appearance of aneuploid plants with a reduced chromosome number in the progeny (Silkova et al., 2021). Such plants can reach up to 70 % of the population (Stepochkin, Vladimirov, 1978). Divergent forms of hexaploid (6x) triticales often arise in the progenies of aneuploid plants (Kalinka, Achrem, 2018). They are more productive and cytogenetically stable than octoploid ones, which explains the fact that triticale cultivars worldwide are predominantly hexaploid.

Along with high yield potential and good grain quality, triticale exhibits resistance to many diseases and adverse environmental conditions, making this crop promising for agricultural production. Over recent decades, global triticale production has gradually increased, exceeding 10 million tons (<https://www.fao.org/faostat/en/#data/QCL>). The largest producers are Poland, Belarus, Germany, Russia, and France (Hamid et al., 2024).

Triticale is widely used for nutritional and forage purposes, as well as in the production of biofuels, alcohol, and in brewing (Cantale et al., 2016; Mirontseva et al., 2018; Zhu, 2018; Gaviley et al., 2024; Latini et al., 2024). Triticale grain has a higher biological value than wheat and rye. Due to its high lysine concentration, improved protein digestibility, and balanced mineral content, it is used as health-promoting nutrition in the diet of people and animals (Kamanova et al., 2023). Furthermore, triticale is suitable for producing high-quality flour used in bread baking and confectionery industries, including for making dietary bread and breakfast cereals (Leonova et al., 2022).

Despite its many advantages, triticale is still underutilized and consumed in much smaller quantities than other major cereals. This is primarily due to the wrinkled grain, flour color, and specific taste characteristics. Breeders put continuous efforts aiming to improve triticale

grain features and successfully create smooth-grained varieties with flour quality not inferior to that of wheat (Camerlengo, Kiszonas, 2023). It is worth noting that the low genetic diversity of triticale confirmed by the studies based on molecular data (Niedziela et al., 2016; Losert et al., 2017) is determined by the gene pool of the parental wheat and rye species. Therefore, it is important to expand the gene diversity of triticale by introducing donors of new traits.

Nowadays, consumers are increasingly striving for healthier diets, making a demand for alternative solutions, when it comes to cereal crops used in human and animal nutrition. Cereals with blue, purple, and black colored grains as well as increased anthocyanin content show enhanced functional and nutritional properties and attract the increasing attention of researchers (Sharma S. et al., 2018; Loskutov, Khlestkina, 2021; Yudina et al., 2021). Anthocyanins are a group of water-soluble polyphenolic compounds that have a positive impact on human health due to their anti-inflammatory, anti-diabetic, anti-cancer, and antibacterial properties (Zhu, 2018; Francavilla, Joye, 2020; Gard et al., 2022).

It is known that purple grain color is determined by the synthesis and accumulation of anthocyanins in the pericarp of the grain. This process is initiated as a result of complementary interaction between two key genes, *Pp3* and *Pp-1* (Khlestkina, 2013). The *Pp3* gene was transferred from *T. aestivum* Jakubz to chromosome 2A of wheat, and the *Pp-1* gene is localized on the short arms of homeologous group 7 chromosomes and controls the coloration of the coleoptile and other plant organs (Khlestkina, 2013; Shoeva et al., 2014; Khlestkina et al., 2015).

At present, major efforts are underway in Russia to develop initial breeding material with colored grain in wheat (Khlestkina et al., 2017; Vasilova et al., 2021; Rubets et al., 2022; Shamanin et al., 2022; Gordeeva et al., 2024; Chumanova et al., 2025), barley (Kukoeva et al., 2024), and triticale (Petraш, Stepochkin, 2023). Several purple-grained common wheat cultivars, namely Nadira, Pamyati Konovalova, and Ef 22, have been registered in the State Register of Breeding Achievements (<https://gossortrf.ru>).

Creating triticale with increased anthocyanin content in the grain will redefine the development potential of this crop as a product for healthy, balanced, and func-

tional human and animal nutrition. Currently, there are no registered triticales cultivars with purple grain color. Purple-grained forms and lines of wheat and emmer wheat can act as donors of this trait. However, when hybridizing purple-grained soft wheat (genomic formula BBAADD) with hexaploid triticales (genomic formula BBAARR), diploidization of a hybrid may cause the risk of complete or partial loss of R-genome chromosomes in the progeny, leading to the formation of 6x forms with an incomplete set of rye chromosomes, which is observed in hybrids of triticales with spring common wheat over a number of generations (Badaev et al., 1985).

Crossing purple-grained common wheat with rye can produce haploid sterile hybrids that require colchicine treatment for chromosome doubling to obtain purple-grained octoploid wheat-rye amphiploids. These, in turn, are to be crossed with hexaploid triticales forms to subsequently select 6x amphiploids with purple grain color in the progeny. The example of successful development of new hexaploid-level triticales forms in crosses of 28-chromosome emmer wheat (*T. dicoccum* (Schrank) Schuebl.) with hexaploid triticales (Silkova et al., 2023) gave an impetus to developing purple-grained triticales as a result of crosses of 6x triticales with the previously obtained emmer wheat line 27-3/17 with a characteristically high anthocyanin content in the grain (Stepochkin et al., 2023).

The goal of the present paper was to obtain new forms of wheat-rye amphiploids with purple grain color using marker-assisted breeding and to comparatively evaluate their anthocyanin content and productivity parameters.

Materials and methods

Plant material. F₂ hybrids (*Triticosecale* × *T. dicoccum*) were used as plant material. The hybrids created by crossing the hexaploid triticales cultivar Sadko from the global collection of the Vavilov Institute of Plant Genetic Resources (VIR k-3927) and the purple-grained emmer wheat line 27-3/17 (Stepochkin et al., 2023) were grown using the *in vitro* embryo culture method (Petrash, Steepochkin, 2023). The plant material production scheme is shown in Figure 1.

DNA isolation and molecular analysis. Total DNA was isolated from young leaves of individual plants according to the method described by J. Plaschke et al. (1995). The PCR reaction mixture contained 100 ng of template DNA, reaction buffer (100 mM Tris-HCl, pH 8.5 (at 25 °C), 100 mM KCl, 0.4 mM of each deoxynucleoside triphosphate, 4 mM MgCl₂, 0.06 units/μL Taq DNA polymerase, 0.2 % Tween 20, stabilizers for HS-Taq DNA polymerase) (“Biolabmix”, Russia), 0.3 μM forward and reverse primers (“Biosset”, Russia). Plant

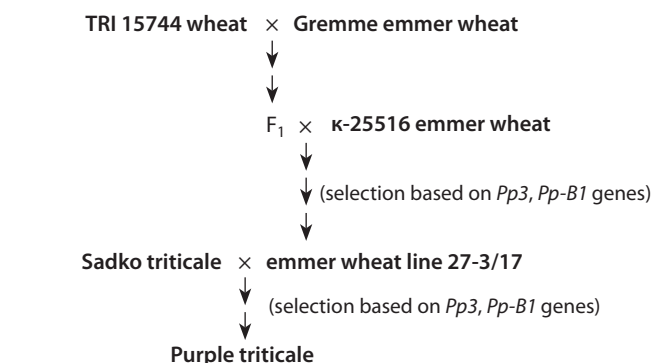


Fig. 1. The purple-grain triticales production scheme.

screening was performed using the intragenic marker Pp3-diagnostic (Shoeva et al., 2021) developed for the *Pp3* gene on chromosome 2A and the microsatellite marker Xgwm0046 flanking the *Pp-B1* gene located on chromosome 7B. PCR procedures with the markers and separation of PCR products in agarose gel were carried out according to the described protocols (Shoeva et al., 2021; Gordeeva et al., 2023).

Determination of total anthocyanin content. Anthocyanin content was determined with modifications according to the principles and methods specified in the publication (Sharma A. et al., 2023). Mature grains of the studied samples were ground using an EM-3A laboratory mill (Russia). Anthocyanins were extracted from whole grain flour in 1 % HCl in methanol for 12 hours at +4 °C. After centrifugation for 25 minutes at 12,000 rpm and +4 °C, the optical density of the extract was measured using a SmartSpec™ Plus spectrophotometer (Bio-Rad Laboratories, Inc., USA) at wavelengths of 530 and 700 nm. The analysis was performed in triplicate for each sample. Anthocyanin content was calculated in terms of the equivalent amount of the most common grain anthocyanidin, namely cyanidin-3-glucoside, using the formula as follows:

$$C = (A/e) \times (V/1,000) \times MW \times (1/Sample\ Wt) \times 10^6,$$

where *C* – the total anthocyanin content, μg cyanidin-3-glucoside/g flour; *A* – the calculated absorption value (*A*₅₃₀–*A*₇₀₀); *e* – the molar extinction coefficient of cyanidin-3-glucoside (25.965 cm^{–1}·M^{–1}); *V* – the total extract volume, mL; *MW* – the molecular weight of cyanidin-3-glucoside (449); *Sample Wt* – the sample weight, mg.

Field evaluation of plant productivity. Study of productivity traits of the developed purple-grained triticales samples and their parents was conducted in 2024 and 2025 on the experimental field of the Siberian Research Institute of Plant Production and Breeding – Branch of the Institute of Cytology and Genetics, Siberian Branch

of Russian Academy of Sciences (SibRIPP&B – Branch of IC&G SB RAS). Plants were grown on plots 100 cm wide with 20 cm between rows.

Weather conditions during the years of field experiments were characterized by excessive moisture and high temperatures. By the beginning of September 2024, the sum of effective temperatures was 1,590.8 °C, and in 2025, it was 1,601 °C, compared to the long-term average of 1,428 °C. The moisture supply level of the 2024 growing season was excessive (hydrothermal coefficient (HTC) of 1.91 in May–August), with August receiving 2.3 times the normal precipitation, which led to the development of fungal infections and pre-harvest sprouting. The 2025 growing season was generally characterized by sufficient moisture supply (HTC of 1.13 in May–August), with the precipitation being below the long-term average in June and July and above that in August.

Statistical analysis. Statistical analysis was performed using the Statistica 12 and Microsoft Excel 2016 software. Values of plant trait indicators were compared using Student's *t*-test.

Results and discussion

High spike sterility and low germination were noted in early generations of triticale and purple-grained emmer wheat hybrids. Dwarf sterile plants with spikes resembling those of emmer wheat were observed. Completely sterile plants were no longer observed starting from F₄. All spikes of plants in the next generation were typical for hexaploid triticale.

Analysis of plants using molecular markers was conducted in F₄ hybrids (Fig. 2). All grains in this generation had purple coloration, and the coleoptile had purple color upon germination. According to PCR results with the Pp3-diagnostic marker, it was established that

triticale and emmer wheat produced different amplification fragments. As a result of the PCR, all hybrids synthesized a fragment approximately 186 bp in length, which coincided with the fragment from emmer wheat. Therefore, the hybrids carried the dominant *Pp3* gene on chromosome 2A.

Analysis with the microsatellite marker Xgwm0046, closely linked to the *Pp-B1* gene (responsible for colored coleoptile) showed that the hybrids amplified a fragment of approximately 186 bp in length similarly to the parental emmer wheat, whereas a longer fragment was synthesized with triticale DNA. Based on molecular analysis, it was established that all studied F₄ hybrids carried the dominant *Pp3* and *Pp-B1* genes in a homozygous state.

Four best plants (2-1-1; 2-1-3, 2-1-5, and 2-1-6) were selected from the F₄ generation based on spike fertility and were subsequently studied as separate families. All plants turned out to be purple-grained in subsequent growing seasons, yet they were morphologically different. In Figure 3, images of the spike (in two projections), as well as grains of some promising F₆ hybrids and their parents are presented. It can be seen from the images that the obtained samples had colored grain and an awned spike, morphologically more similar to that of Sadko triticale. It is worth noting that the hybrids, like the maternal cultivar Sadko, had hairy peduncles. This trait is controlled by the dominant *Hp* gene located on chromosome 5R (Korzun et al., 1996), therefore, the selected triticale forms carried chromosome 5 of rye or its fragment.

The F₅₋₆ progeny was tested for productivity traits under field conditions in 2024–2025 along with the parental forms (see the Table). Structural analysis was performed in 10 typical plants from each family with Sadko triticale used as control. Field testing showed that,

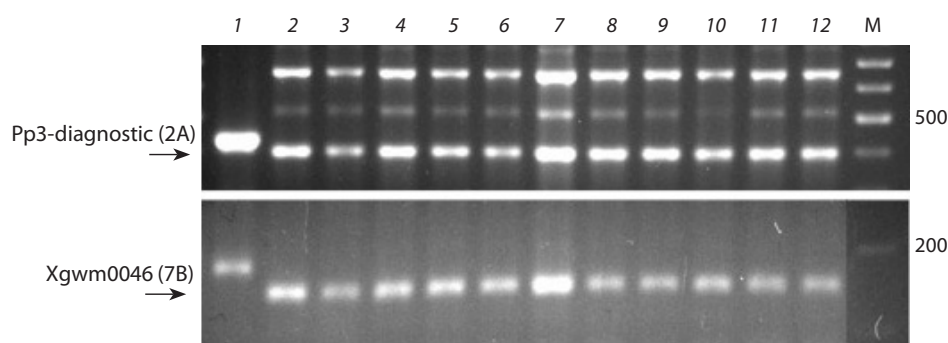


Fig. 2. Gel electrophoresis images of PCR products from F₄ hybrids as follows: *Triticosecale* × *T. dicoccum* and their parental forms with the Pp3-diagnostic and Xgwm0046 DNA markers.

1 – Sadko triticale; 2 – emmer wheat line 27-3/17; 3–12 – F₄ hybrids: Sadko × 27-3/17; M – Step 100 DNA marker. Arrows indicate the DNA fragments corresponding to the dominant allele of the *Pp3* gene on chromosome 2A (about 398 bp) and the Xgwm0046 marker linked to the dominant allele of the *Pp-B1* gene on chromosome 7B (about 186 bp).



Fig. 3. The spikes and grains of purple-grain F₆ hybrids Sadko × 27- 3/17 (*Triticosecale* × *T. dicoccum*) and their parental forms.

Productivity indicators of F₅₋₆ hybrids of purple-grained triticale forms and their parents in 2024–2025

Specimen	Main spike length, cm		Spikelets per main spike, pcs		Grains per main spike, pcs		Grain weight per main spike, g		1,000 grain weight, g		Yield, g/m ²	
	2024 г.	2025 г.	2024 г.	2025 г.	2024 г.	2025 г.	2024 г.	2025 г.	2024 г.	2025 г.	2024 г.	2025 г.
2-1-1	9.67 ± 0.49	9.71 ± 0.67	24.4 ± 1.1	24.6 ± 1.4	27.7* ± 4.9	38.6 ± 6.7	0.85* ± 0.22	1.98 ± 0.43	30.1 ± 3.4	47.5 ± 5.9	216* ± 52	493 ± 49
2-1-3	9.56 ± 0.68	9.42 ± 0.78	23.6 ± 1.4	23.0 ± 1.9	21.6* ± 8.2	29.6* ± 4.6	0.62* ± 0.35	1.39 ± 0.42	29.2 ± 5.7	46.0 ± 0.8	163** ± 30	384* ± 86
2-1-5	9.66 ± 0.76	9.28 ± 0.42	23.2 ± 1.1	23.6 ± 1.4	32.5* ± 4.0	29.5* ± 3.8	0.99 ± 0.22	1.38* ± 0.25	29.5 ± 3.4	46.6 ± 6.5	244** ± 32	404* ± 35
2-1-6	9.29 ± 0.46	9.56 ± 0.34	24.2 ± 1.1	23.8 ± 0.9	30.2* ± 5.7	39.8 ± 3.5	0.95 ± 0.24	2.00 ± 0.23	30.5 ± 2.6	50.1 ± 3.0	281* ± 79	486 ± 94
Mean value	9.55 ± 0.17	9.49 ± 0.18	23.9 ± 0.6	23.8 ± 0.7	28.0* ± 4.7	34.4 ± 5.6	0.85* ± 0.16	1.69 ± 0.34	29.8 ± 0.6	47.6 ± 1.8	275** ± 44	470 ± 61
Sadko triticale	10.19 ± 0.36	10.10 ± 0.34	24.6 ± 0.6	24.8 ± 0.9	44.8 ± 2.9	45.7 ± 2.9	1.53 ± 0.20	2.28 ± 0.12	33.8 ± 2.7	50.2 ± 3.3	473 ± 29	584 ± 40
Emmer wheat line 27-3/17	6.79** ± 0.46	7.73** ± 0.39	17.0** ± 0.9	17.8** ± 0.9	29.9* ± 2.3	35.1* ± 2.0	1.00 ± 0.08	1.38** ± 0.05	33.7 ± 2.6	39.5* ± 2.6	235** ± 36	306** ± 29

Note. Mean value is the mean trait value across triticale hybrid forms. Significant differences from Sadko triticale: **p* < 0.05; ** *p* < 0.01.

similarly to Sadko triticale, the hybrids were resistant to lodging whereas emmer wheat had a weak straw and lodged completely. Unlike emmer wheat, hybrid triticale forms and the Sadko cultivar were also free-threshing cereals with sturdy rachis.

In the years of study, the hybrid lines exhibited the spike length and the number of spikelets per spike comparable to those of the Sadko control (over 9 cm and 23 pcs, respectively), whereas emmer wheat spike length did not exceed 8 cm, the number of its spikelets being 17–18 pcs. In terms of the number of grains in the main spike, the hybrids were inferior to the control; significantly lower values were observed for families 2-1-3 and 2-1-5 in 2025 (29.6 and 29.5 pcs, compared to 45.7 pcs in Sadko). Meanwhile, emmer wheat exhibited a significantly lower number of grains than Sadko in both years due to the smaller number of spikelets per spike. The low number of set grains in hybrid triticale forms is associated with partial sterility. The grain weight per spike in hybrid families in the years of study was comparable to the control level but did not exceed it, averaging 0.85 g in 2024 and 1.69 g in 2025; the exception being sample 2-1-3 in 2024 with a significantly lower grain weight per spike at 0.62 g.

The 1,000 grain weight acts as an indicator of grain size; in the years of the study, the obtained purple-grained triticale forms showed 1,000 grain weight values comparable to those of the control cultivar, specifically 29.8 g in 2024 and 47.6 g in 2025 compared respectively to 33.8 g and 50.2 g in Sadko. In terms of yield per square meter, the hybrids were inferior to the control triticale cultivar, especially in 2024, when they showed the average yield of 275 g/m², compared to 473 g/m² in Sadko. In 2025, more favorable weather conditions prevailed during the growing season, and the hybrid yield increased to 470 g/m². The Sadko cultivar also showed a higher result yielding 584 g/m².

Based on the evaluation of productivity traits, it can be concluded that the developed purple-grained triticale forms exhibited trait values similar to those of Sadko hexaploid triticale. However, most hybrids still showed partial spike sterility possibly related to meiotic disturbances. As shown by previous cytological studies in F₆ hybrids of emmer wheat with triticale, meiotic anomalies persisted over several generations (Silkova et al., 2023). It is also known that, when it comes to triticale hybrids, genome stabilization and meiotic normalization can occur even in later generations (Kalinka, Achrem, 2018).

To determine anthocyanin content in the grains, nine plants were taken from four families in the F₆ generation based on the set grains per spike and productivity values. All grains had anthocyanin coloration ranging from

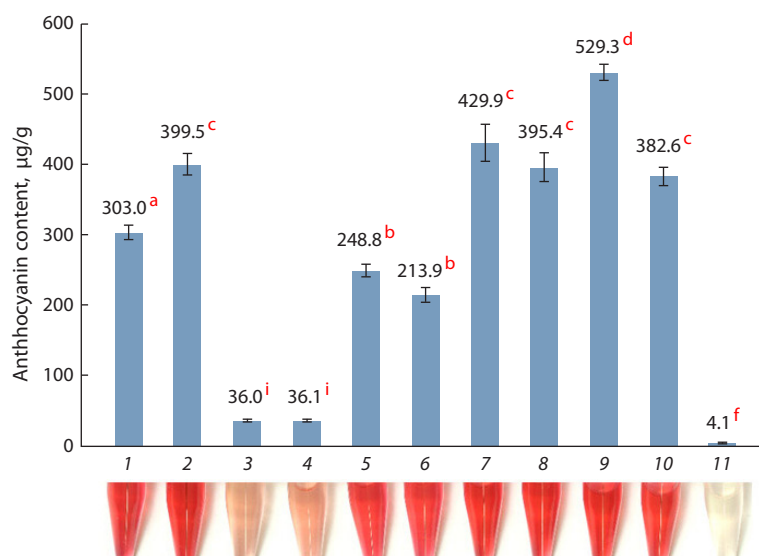


Fig. 4. Total anthocyanin content in terms of the equivalent amount of Cy-3-Glu in methanol extracts of whole-grain flour from nine hybrids of purple-grain triticale and their parental forms.

1 – F₆ 2-1-1-4d; 2 – F₆ 2-1-1-4e; 3 – F₆ 2-1-3-3a; 4 – F₆ 2-1-3-3b; 5 – F₆ 2-1-3-5a; 6 – F₆ 2-1-5-7; 7 – F₆ 2-1-5-10a; 8 – F₆ 2-1-6-4b; 9 – F₆ 2-1-6-6; 10 – emmer wheat line 27-3/17; 11 – Sadko triticales. Letters a–f indicate statistically significant differences between the compared groups at $p < 0.05$.

moderate to intense. Total anthocyanin content (TAC) was determined in whole grain flour obtained from the 2025 field material. The results of the tested plants were compared to those of emmer wheat line 27-3/17, i. e. the donor of purple coloration (Fig. 4). The obtained data showed that TAC values in hybrids ranged from 36 µg/g to 529.3 µg/g. The control (emmer wheat) value was 382.6 µg/g. A significantly higher value of 529.3 µg/g was observed in specimen 2-1-6-6. The specimens 2-1-1-4e at 399.5 µg/g, 2-1-5-10a at 429.9 µg/g, and 2-1-6-4b at 395.4 µg/g showed the results comparable to that of emmer wheat. High TAC values (of over 200 µg/g) were also recorded in samples: 2-1-1-4d, 2-1-3-5a, and 2-1-5-7. The lowest TAC values (36 µg/g) were found in the whole grain flour of the samples 3a and 3b from the family 2-1-3. The white-grained maternal triticales cultivar Sadko showed a total anthocyanin content of 4.1 µg/g flour.

According to estimates by various authors, the total anthocyanin content in purple-grained wheat grain ranges from 16 to 477 µg/g (Gordeeva et al., 2024). In this paper, triticales forms with high anthocyanin content of 529 µg/g were obtained exceeding the results of the donor purple-grained emmer wheat line. It is known that anthocyanin biosynthesis in the grain pericarp requires a complementary interaction of two genes, *Pp-1* and *Pp3*. The latter is a unique gene inherited from an Ethiopian wheat (Khlestkina, 2014). At the same time, the *Pp-1* gene controlling the coloration of the coleoptile

and other plant organs, is widespread among rare wheat species (Shoeva et al., 2024) and common wheat cultivars (Gordeeva et al., 2024).

Pigmentation of plant organs is widespread among triticales cultivars; for example, the spring triticales cultivar Amore has strong anthocyanin coleoptile coloration (Skatova et al., 2018), cultivars Venets, Ullubiy, and Boguslav exhibit moderate coloration (<https://gossortrf.ru>). The cultivar description for Sadko triticales discussed in this paper cites the anthocyanin coloration of the flag leaf auricles (Grib et al., 2011). It is possible that Sadko carries one of the *Pp-1* gene alleles, which, in combination with the *Pp3* and *Pp-B1* genes from emmer wheat, determines greater expression of the grain coloration trait in some hybrids.

Obtaining constant forms or lines of wheat-rye amphiploids using hybridization of triticales with emmer wheat is a lengthy process. Its complexity is due to ongoing shaping associated both with the recombination of genes from the two species and with genomic rearrangements. To accelerate the solution of this problem at the next breeding stage, it is important to apply the doubled haploid method based on the obtained purple-grained triticales forms.

Conclusion

In the present paper, a successful attempt has been made to develop triticales forms with purple grain based on crossing the triticales cultivar Sadko and emmer wheat

line 27-3/17 acting as the donor of the *Pp3* and *Pp-B1* genes responsible for grain color. The obtained hybrid triticale forms have a purple grain color, a high total anthocyanin content, and approach the maternal triticale form in terms of productivity indicators and spike shape. Further research should be aimed at obtaining constant, genetically stable lines of purple-grained triticale, which will serve as initial breeding material and become donors for high anthocyanin content in grain for other triticale cultivars. It is assumed that purple-colored triticale grains will be used as an additional source of antioxidants in functional nutrition for humans and animals.

References

- Badaev N.S., Badaeva E.D., Bolsheva N.L., Maximov N.G., Zelenin A.V. Cytogenetic analysis of forms produced by crossing hexaploid triticale with common wheat. *Theor Appl Genet.* 1985;70: 536-541. doi 10.1007/BF00305987
- Camerlengo F., Kiszonas A.M. Genetic factors influencing triticale quality for food. *J Cereal Sci.* 2023;113:103744. doi 10.1016/j.jcs.2023.103744
- Cantale C., Petrazzuolo F., Correnti A., Farneti A. Triticale for bio-energy production. *Agric Agric Sci Procedia.* 2016;8:609-616. doi 10.1016/j.aaspro.2016.02.083
- Chumanova E.V., Efremova T.T., Sobolev K.V., Kosyaeva E.A. Development and characterization of wheat hybrids *Triticum petropavlovskyi* Udacz. et Migusch. × *Triticum aestivum* L. with large colored grain and increased antioxidant content. *Pisma v Vavilovskii Zhurnal Genetiki i Seleksii = Lett Vavilov J Genet Breed.* 2025;11(2):80-89. doi 10.18699/letvjgb-2025-11-12 (in Russian)
- Francavilla A., Joye I.J. Anthocyanins in whole grain cereals and their potential effect on health. *Nutrients.* 2020;12(10):2922. doi 10.3390/nu12102922
- Gard M., Kaur S., Sharma A., Kumari A., Tiwari V., Sharma S., Kapoor P., Sheoran B., Goyal A., Krishania M. Rising demand for healthy foods-anthocyanin biofortified colored wheat is a new research trend. *Front Nutr.* 2022;9:878221. doi 10.3389/fnut.2022.878221
- Gaviley O.V., Katerynych O.O., Ionov I.A., Dekhtiarova O.O., Griffin D.K., Romanov M.N. Triticale: a general overview of its use in poultry production. *Encyclopedia.* 2024;4(1):395-414. doi 10.3390/encyclopedia4010027
- Gordeeva E.I., Shoeva O.Y., Shamanin V.P., Khlestkina E.K. The molecular markers applying in breeding of spring bread wheat (*Triticum aestivum* L.) lines with different anthocyanin coloration of the grains. *Pisma v Vavilovskii Zhurnal Genetiki i Seleksii = Letters to Vavilov J Genet Breed.* 2023;9(2):86-99. doi 10.18699/LettersVJ-2023-9-11 (in Russian)
- Gordeeva E.I., Shamanin V.P., Khlestkina E.K., Shoeva O.Yu. On peculiarities of breeding purple-grained wheat based on varieties with anthocyanin pigmentation of coleoptiles and stems. *Sel'skokhozyaistvennaya Biologiya = Agricultural Biology.* 2024;59(3):507-524. doi 10.15389/agrobiologiya.2024.3.507rus (in Russian)
- Grib S.I., Bushtevich V.N., Polyakova E.L., Krylova T.M., Pavlova L.D., Bondarchuk V.A., Filatova T.F., Pilipenko D.V. Triticale variety "Sadko". Patent for a Selection Achievement No. 307. Republic of Belarus, 15.06.2011 (in Russian)
- Hamid, Kathuria D., Gautam S., Suri S., Jaiswal A.K. Triticale. In: Singh J., Kaur S., Rasane P., Singh J. (Eds) *Cereals and Nutraceuticals.* Singapore: Springer, 2024;163-189. doi 10.1007/978-981-97-2542-7_8
- Kalinka A., Achrem M. Reorganization of wheat and rye genomes in octoploid triticale (× *Triticosecale*). *Planta.* 2018;247(4):807-829. doi 10.1007/s00425-017-2827-0
- Kamanova S., Yermekov Y., Shah K., Mulati A., Liu X., Bulashev B., Toimbayeva D., Ospankulova G. Review on nutritional benefits of triticale. *Czech J Food Sci.* 2023;41(4):248-262. doi 10.17221/67/2023-CJFS
- Khlestkina E.K. Genes determining the coloration of different organs in wheat. *Russ J Genet Appl Res.* 2013;3(1):54-65. doi 10.1134/S2079059713010085
- Khlestkina E.K., Shoeva O.Y., Gordeeva E.I. Flavonoid biosynthesis genes in wheat. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2014;18(4/1):784-796 (in Russian)
- Khlestkina E.K., Shoeva O.Y., Gordeeva E.I. Flavonoid biosynthesis genes in wheat. *Russ J Genet Appl Res.* 2015;5(3):268-278
- Khlestkina E.K., Usenko N.I., Gordeeva E.I., Stabrovskaya O.I., Sharfunova I.B., Otmakhova Y.S. Evaluation of wheat products with high flavonoid content: justification of importance of marker-assisted development and production of flavonoid-rich wheat cultivars. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2017;21(5):545-553. doi 10.18699/VJ17.25-o (in Russian)
- Korzun V., Melz G., Börner A. RFLP mapping of the dwarfing (*Ddw1*) and hairy peduncle (*Hp*) genes on chromosome 5 of rye (*Secale cereale* L.). *Theor Appl Genet.* 1996;92:1073-1077. doi 10.1007/BF00224051
- Kukoeva T.V., Molobekova C.A., Totsky I.V., Vasiliev G.V., Pronozin A.Y., Afonnikov D.A., Khlestkina E.K., Shoeva O.Y. Enrichment of grain anthocyanin content through marker-assisted breeding for *Ant1*, *Ant2* or *HvMyc2* genes in barley (*Hordeum vulgare* L.). *Agronomy.* 2024;14(6):1231. doi 10.3390/agronomy14061231
- Latini A., Cantale C., Gazza L., Nocente F., Migliore G., Maccioni O., Marconi O., Floridi S., De Francesco G., Ammar K., Galeffi P. Exploring the potential of triticale lines for bioethanol production. *Explor Foods Foodomics.* 2024;2:613-625. doi 10.37349/eff.2024.00054
- Leonova S., Badamshina E., Koshchina E., Kalugina O., Gareeva I., Leshchenko N. Triticale flour in bakery and rusk products. *Food Sci Technol Int.* 2022;28(6):524-534. doi 10.1177/10820132211023273
- Losert D., Maurer H.P., Marulanda J.J., Würschum T. Phenotypic and genotypic analyses of diversity and breeding progress in European triticale (× *Triticosecale* Wittmack). *Plant Breed.* 2017;136(1): 18-27. doi 10.1111/pbr.12433
- Loskutov I.G., Khlestkina E.K. Wheat, barley, and oat breeding for health benefit components in grain. *Plants.* 2021;10(1):86. doi 10.3390/plants10010086
- Mirontseva A., Tsed E., Volkova S. Justification of bioactivated grain triticale use in alcohol production. *Tekhnika i Tekhnologiya Pishchevykh Proizvodstv = Food Processing: Techniques and Technology.* 2018;48(1):57-65. doi 10.21603/2074-9414-2018-1-57-65 (in Russian)
- Muntzing A. Historical review of the development of triticale. In: *Triticale: Proceedings of an International Symposium, El Batán, Mexico.* 1974:13-30
- Niedziela A., Orłowska R., Machczyńska J., Bednarek P.T. The genetic diversity of triticale genotypes involved in Polish breeding programs. *Springerplus.* 2016;5(1):355. doi 10.1186/s40064-016-1997-8
- Petrash N.V., Stepochkin P.I. Development of purple-grain hybrids in distant crosses of triticale, bread wheat and emmer using the embryo culture. *Pisma v Vavilovskii Zhurnal Genetiki i Seleksii = Lett Vavilov J Genet Breed.* 2023;9(4):183-188. doi 10.18699/LettersVJ-2023-9-25 (in Russian)
- Plaschke J., Ganai M.W., Roeder 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
- Rubets V.S., Voronchikhina I.N., Igonin V.N., Sidorenko V.S., Voronchikhin V.V. Characteristics of violet-green variety of spring soft wheat in the conditions of the central region of the Non-Chernozem zone of Russia. *Mezhdunarodnyi Sel'skokhozyaistvennyi Zhurnal.* 2022;5:525-529. doi 10.55186/25876740_2022_65_5_525 (in Russian)

- Skatova S.E., Vasiliev V.V., Tyslenko A.M., Zuev D.V. Spring triticale 'Amore' is a new variety for adaptive agriculture in the Non Black soil zone. *Vladimirskij Zemledelets*. 2018;2(84):10-15 (in Russian)
- Shamanin V.P., TekinCakmak Z.H., Gordeeva E.I., Karasu S., Pototskaya I., Chursin A.S., Pozherukova V.E., Ozulku G., Morgounov A.I., Sagdic O., Koksel H. Antioxidant capacity and profiles of phenolic acids in various genotypes of purple wheat. *Foods*. 2022; 11(16):2515. doi 10.3390/foods11162515
- Sharma A., Yadav M., Tiwari A., Ali U., Krishania M., Bala M., Sharma P., Goudar G., Roy J.K., Navik U., Garg M. A comparative study of colored wheat lines across laboratories for validation of their phytochemicals and antioxidant activity. *J Cereal Sci*. 2023;112: 103719. doi 10.1016/j.jcs.2023.103719
- Sharma S., Chunduri V., Kumar A., Kumar R., Khare P., Kondepudi K.K., Bishnoi M., Garg M. Anthocyanin bio-fortified colored wheat: nutritional and functional characterization. *PLoS One*. 2018; 13(4):0194367. doi 10.1371/journal.pone.0194367
- Shoeva O.Y., Gordeeva E.I., Khlestkina E.K. The regulation of anthocyanin synthesis in the wheat pericarp. *Molecules*. 2014;19(12): 20266-20279. doi 10.3390/molecules191220266
- Shoeva O.Yu., Gordeeva E.I., Khlestkina E.K. Intragenic DNA Marker for Selecting Wheat with High Anthocyanin Content in the Pericarp of the Grain. Russian patent RU 2774444C1, 29.11.2021. <https://patentimages.storage.googleapis.com/7f/6b/b4/734418d3890f53/RU2774444C1.pdf> (in Russian)
- Shoeva O.Yu., Gordeeva E.I., Khlestkina E.K., Gashimov M.E., Kurkiev K.U. Study of rare species of wheat as donors for breeding for functional nutrition. *Sel'skokhozyaistvennaya Biologiya = Agricultural Biology*. 2024;59(5):955-972. doi 10.15389/agrobiol.2024.5.955rus (in Russian)
- Silkova O.G., Ivanova Y.N., Loginova D.B., Solovey L.A., Sycheva E.A., Dubovets N.I. Karyotype reorganization in wheat-rye hybrids obtained via unreduced gametes: is there a limit to the chromosome number in triticale? *Plants*. 2021;10(10):2052. doi 10.3390/plants10102052
- Silkova O.G., Ivanova Y.N., Stepochkin P.I. Creation and study of emmer (*Triticum dicoccum*) × triticale hybrids. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2023;27(4):323-332. doi 10.18699/VJGB-23-39
- Stepochkin P.I., Vladimirov N.S. Chromosome number, seed set and winter hardiness characteristics of C₁ winter lines of homogenomic 8x triticale. *Genetika = Genetics (Moscow)*. 1978;14(9):1597-1603 (in Russian)
- Stepochkin P.I., Gordeeva E.I., Khlestkina E.K. Marker-assisted breeding of hybrid lines of *Triticum dicoccon* (Schränk) Schuebl. × *Triticum aethiopicum* Jakubz. with purple grain. *Proceedings on Applied Botany, Genetics and Breeding*. 2023;184(2):139-148. doi 10.30901/2227-8834-2023-2-139-148 (in Russian)
- Vasilova N.Z., Askhadullin D.F., Askhadullin D., Bagavieva E.Z., Tazutdinova M.R., Khusainova I.I. Violet-green variety of spring soft wheat Nadira. *Zernobobovye i Krupnyanye Kul'tury = Legumes Groat Crops*. 2021;4(40):66-75. doi 10.24412/2309-348X-2021-4-66-75 (in Russian)
- Yudina R.S., Gordeeva E.I., Shoeva O.Yu., Tikhonova M.A., Khlestkina E.K. Anthocyanins as functional food components. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2021; 25(2):178-189. doi 10.18699/VJ21.022 (in Russian)
- Zhu F. Triticale: nutritional composition and food uses. *Food Chem*. 2018;241:468-479. doi 10.1016/j.foodchem.2017.09.009

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

Received November 10, 2025. Revised January 14, 2026. Accepted January 15, 2026.

doi 10.18699/vjgb-26-65

Development of naked *Triticum durum* × *Triticum dicoccum* hybrids with anthocyanin color of the grain pericarp

E.I. Gordeeva ¹, A.Yu. Novoselskaya-Dragovich ², O.Yu. Shoeva ¹, E.K. Khlestkina ^{1, 3}¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia² N.I. Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, Russia³ Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia elgordeeva@bionet.nsc.ru

Abstract. The production of emmer hybrids with a high content of anthocyanins in the grains for the production of functional foods is a promising breeding direction. Phenotyping and preliminary assessment of the inheritance of gliadin-coding genes were performed for the most promising purple-grained emmer hybrids obtained previously after a complex three-stage crossing of purple-grained durum wheat (*T. durum* Desf.) with two different forms of spring emmers (*T. dicoccum* Schrank): the hybrid naked-grained variety Gremme and the red-grained awnless mutant line k25516. Genotyping hybrids for the storage protein genes in wheat grain, gliadins (*Gli*), enabled the selection of a purple-grained line that fully inherited gliadin-coding genes from emmer wheat k-25516, and a line inheriting these genes from durum wheat and emmer wheat k-25516. To improve the breeding material, backcrossing of three phenotypically and qualitatively different purple-grained hybrid lines with the parental variety Gremme, which demonstrated the highest yield, was conducted. During the *Pp* (Purple pericarp) genes selection of the plants in F₂₋₃ progenies, the use of microsatellite markers located close to *Pp* genes did not demonstrate reliable linkage to the target genes. The intragenic polymorphic PCR markers made it possible to accurately select plants carrying dominant alleles of two complementarily interacting genes, *Pp-B1* and *Pp3* in F₂₋₄. Based on the ease of grain threshing, the plants were selected in F₄. Thus, over two years, using small areas of the greenhouse and marker-controlled selection, a collection consisting of 25 naked and semi-naked spring purple-grained lines of wheat-emmer hybrids, constant in anthocyanin coloration and differing in gliadin-coding genes and other quality traits, was obtained.

Key words: naked emmer; purple grain; anthocyanins; gliadins; marker assisted breeding; PCR markers

For citation: Gordeeva E.I., Novoselskaya-Dragovich A.Yu., Shoeva O.Yu., Khlestkina E.K. Development of naked *Triticum durum* × *Triticum dicoccum* hybrids with anthocyanin color of the grain pericarp. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2026;30(4):645-657. doi 10.18699/vjgb-26-65

Funding. This work was supported by the Russian Science Foundation grant No. 25-26-20157 (<https://grant.rscf.ru/site/user/bids?role=master>) and the Ministry of Education, Science and Innovation Policy of the Novosibirsk Region (agreement dated April 21, 2025, No. 30-2025-001058).

Acknowledgements. The plants were grown at the Plant Reproduction Center of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, under budget project FWNr-2026-0029. The authors are grateful to Galina Vladimirovna Generalova and Olga Viktorovna Zakharova for their technical support.

Получение голозерных гибридов *Triticum durum* с *Triticum dicoccum* с антоциановой окраской перикарпа зерновки

Е.И. Гордеева ¹, А.Ю. Новосельская-Драгович ², О.Ю. Шоева ¹, Е.К. Хлесткина ^{1, 3}¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия² Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия³ Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР), Санкт-Петербург, Россия elgordeeva@bionet.nsc.ru

Аннотация. Одно из перспективных направлений селекции – получение гибридов полбы с повышенным содержанием антоцианов в зерновке для производства функциональных продуктов питания. Проведено фенотипирование и дана предварительная оценка по наследованию глиадинкодирующих генов наиболее перспективных фиолетовозерных полбяных гибридов, полученных нами ранее после сложного трехступенчатого скрещивания фиолетовозерной твердой пшеницы (*T. durum* Desf.) с двумя разными формами яровой полбы

(*T. dicoccum* Schrank): гибридным голозерным сортом Греммэ и безостой краснозерной мутантной линией к-25516 (коллекции ВИР). Генотипирование гибридов по генам (*Gli*) запасных белков зерновки пшеницы – глиадинам – позволило отобрать фиолетовозерную линию, целиком наследующую глиадинкодирующие гены от полбы двузернянки к-25516, и линию с наследованием этих генов от твердой пшеницы и полбы двузернянки к-25516. Для улучшения селекционного материала было проведено возвратное скрещивание фиолетовозерных пшенично-полбяных гибридов, содержащих доминантные аллели *Pp-B1* и *Pp3*, с показавшим наилучшую урожайность белозерным голозерным сортом Греммэ. Произведен маркер-контролируемый отбор растений поколения F_{2-4} по генам *Pp* (*Purple pericarp*), регулирующим биосинтез антоцианов в перикарпе зерновки. При отборе фиолетовозерных растений в потомствах F_{2-3} использование микросателлитных маркеров, близко расположенных к генам антоциановой окраски, не показало надежного сцепления с целевыми генами, однако позволило проследить у гибридов сложное наследование генетического материала от разных родителей. Разработанные нами внутривидовые полиморфные ПЦР-маркеры позволили точно отобрать образцы растений, несущие в ДНК одновременно доминантные аллели двух комплементарно взаимодействующих генов, *Pp-B1* и *Pp3*. Растения отбирались по признаку легкой обмолачиваемости зерен в F_4 . Таким образом, за два года при использовании малых площадей тепличного комплекса при помощи маркер-контролируемой селекции в поколении F_4 получена коллекция, состоящая из 25 голозерных и полуголозерных яровых фиолетовозерных линий пшенично-полбяных гибридов, константных по признаку антоциановой окраски перикарпа зерновок и различающихся по глиадинкодирующим генам.

Ключевые слова: голозерная полба; фиолетовая окраска зерна; антоцианы; глиадины; маркер-контролируемая селекция; ПЦР-маркеры

Introduction

The purple color of cereal grains is determined by anthocyanins, a class of flavonoid compounds. Numerous studies have shown that anthocyanins had positive effects on human health, possessing antioxidant, anti-inflammatory, hypoglycemic, and antitumorigenic properties (Liu et al., 2021; Dwivedi et al., 2022; Mohammadi, 2024). Wheat grain anthocyanins had a preventive effect against tumors and neurodegenerative diseases (Tikhonova et al., 2020, 2024; Geyik et al., 2023).

Breeding programs to develop cereals with a purple grain color as a source for the production of fortified, high-quality cereal products and pasta are attractive. One promising candidate for such nutrition is emmer (emmer wheat, *Triticum dicoccum* Schrank., genome BBAA, $2n = 28$) – an ancient hulled wheat species, the grain of which was traditionally used in porridge. Emmer was displaced by the more productive naked-grained durum wheat due to difficulty in threshing and low yields, and as a cultivated crop, it remained in limited regions of the Volga region, Siberia, and the North Caucasus (Badaeva, 2015).

Currently, increasing interest in emmer is explained by its superior nutritional properties, genetic variability, and broad ecological plasticity (Gilev et al., 2018; Biradar et al., 2022). It is easily digestible and surpasses modern cultivated durum and bread wheat in terms of vegetable protein, unsaturated fatty acids, fiber, iron, zinc, and B vitamins, accumulating more antioxidants in the grain, making it an important component of a healthy diet (Zakharova, Tolstova, 2019; Şahin, Karakas, 2022; Cabas-Lühmann et al., 2023; Temirbekova et al., 2024). It can be used as a source for breeding wheat

with enhanced nutritional qualities. A number of emmer samples are resistant to pathogens, making it possible to grow plants without the use of any chemicals or synthetic fertilizers hazardous to human health (Zverev et al., 2016; Kuznetsova et al., 2020; Biradar et al., 2022).

One of the main obstacles to the practical use of emmer wheat is the difficulty of separating the grain from its hull. In recent years, breeders have focused their efforts on developing varieties in which the grain is easily separated from the spikelet and flower hulls, making it easier to thresh. It has been proposed to classify these new emmer wheat varieties as the subspecies *Triticum dicoccon* (Schrank) Schübl. subsp. *nudicoccon* Kobyl. et Smekal. (Smekalova, Kobylansky, 2019).

Purple-grained wheats were previously obtained through classical hybridization and selection from the tetraploid species *T. aethiopicum* Jakubz. from Ethiopia (BBAA genome, $2n = 28$) (Zeven, 1991). The purple anthocyanin coloration trait of wheat pericarp has been well studied and is controlled by complementarily interacting anthocyanin biosynthesis regulatory genes *Pp-B1* on chromosome 7B and *Pp3* on chromosome 2A in tetraploid species, and by *Pp-D1* on chromosome 7D and *Pp3* on chromosome 2A in hexaploid species (Khlestkina et al., 2010; Tereshchenko et al., 2012). By hybridizing hexaploid wheat *T. aestivum* L. (BBAADD genome, $2n = 42$) and performing subsequent hybrid selection using molecular markers closely linked to these genes, it is possible to easily control DNA regions carrying dominant alleles of the purple coloration genes of the grains (Gordeeva et al., 2020). In our laboratory, we have developed diagnostic intragenic DNA markers for key regulatory genes involved in anthocyanin biosynthesis in grains: *Pp1-diagnostic* and *Pp3-diagnostic* (Shoeva

Table 1. Pedigree of samples used for hybridization and their characteristics

No.	Sample/ pedigree	Feature characteristics	Source (References)
1	<i>T. durum</i> sample Tri15744	Purple-grained, early-ripening, short-stemmed, awned spikes, naked durum wheat	IPK Gatersleben collection
2	Gremme (<i>T. dicoccum</i> Belka v.*3/ <i>T. durum</i> Svetlana v.)	White grained, early-ripening, long-stemmed (medium-sized), awned spikes, naked emmer	Chistopol State Variety Testing Station, Tatarstan
3	k-25516 <i>T. dicoccum</i> Schrank. Schuebl. convvar. <i>serbicum</i> Zakharova	Red grained emmer wheat, long-stemmed, mid-season, red stem and spike, hulled grains, awnless spikes mutant, selected from the field population	VIR collection (Volga-Balkan emmer)
4	27-15/8 F ₉ (Tri15744/Gremme//k-25516)	Purple-grained, early-ripening, long-stemmed, awnless spikes, semi-hulled hybrid	ICG SB RAS (Stepochkin et al., 2023)
5	27-3 F ₈ (Tri15744/Gremme//*2 k-25516)	Purple-grained, mid-season, long-stemmed, awnless spikes, semi-hulled hybrid	
6	31-16/6 F ₈ (Tri15744/Gremme//*2 k-25516)	Purple-grained, early-ripening, short- stemmed, awnless spikes, nearly naked hybrid	

et al., 2022; Gordeeva et al., 2023). These markers have been successfully used to create new purple-grained wheat varieties with enhanced functional properties (Gordeeva et al., 2023).

Despite advances in hexaploid wheat breeding (Fisenko et al., 2020; Vasilova et al., 2021; Rubets et al. 2022; Shamanin et al., 2024), there are no tetraploid durum wheat varieties with high anthocyanin content in the country. Using multi-stage hybridization of two different forms of spring emmer based on tetraploid durum wheat, which accumulates anthocyanins in the pericarp of the grain, and marker-assisted selection, we created purple-grained semi-naked hybrids (Stepochkin et al., 2023). However, the main disadvantages of the new hybrids include incomplete threshing of the grain from the flower and spikelet scales, fragility of the rachis, and low yield. The emmer grains with dense scales are difficult to thresh without damaging the outer shells, which contain the main anthocyanin reserves. Moreover, this barrier dramatically reduces the breeding value of such samples. Therefore, the development of emmer lines that are naked and easier to thresh without damaging the integrity of the embryo and outer grain shell, allowing for the preservation of the beneficial properties of emmer with anthocyanin pericarp, is an urgent task.

In 2012, the Gremme spring emmer variety, which is characterized by its nakedness and was bred in Tatarstan, was included in the State Register (<http://reestr.gossort.com/reestr/sort/9052467>). The variety was obtained from an interspecific cross of emmer *T. dicoccum* Schrank. Belka variety with durum wheat *T. durum* Desf. Svetlana

variety, followed by backcrosses on emmer (Temirbekova et al., 2020), and can serve as the starting material for the creation of naked emmer hybrids.

The aim of this study was to develop new, higher-yielding, naked wheat-emmer hybrids with increased anthocyanin content in the grain pericarp. For this purpose the following was performed: 1) a selective assessment of the gluten protein content of previously obtained wheat-emmer hybrids and their parents; 2) crossing of selected promising samples with the naked variety Gremme, followed by marker-assisted selection of purple-grained plants.

Materials and methods

Plant material and crossing scheme. New lines of wheat-emmer hybrids were used in this study. The lines were obtained from complex saturating crosses of purple-grained tetraploid wheat *T. durum* Tri15744 with naked emmer Gremme and then with red-grained emmer *T. dicoccum* k-25516 at the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (Stepochkin et al., 2023). The wheat samples used in hybridization and their origin sources are presented in Table 1. The full crossing scheme is presented in Supplementary Figure S1¹.

To obtain completely naked, more productive plants, direct and reverse backcrosses were carried out on three contrasting samples of the hybrids we had previously obtained (No. 4, 5, and 6; Table 1), which had a rich content of anthocyanins in the grains, with naked emmer

¹ Supplementary Figures S1–S5 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Gord_Engl_30_4.pdf

Gremme, which showed the highest productivity when grown in field conditions (Stepochkin et al., 2023). No. 4 and 6 were selected from early-ripening purple-grained samples after genotyping the genes encoding storage proteins (gliadins). Mid-season No. 5 plants showed the highest total anthocyanin content in their grains after field cultivation in 2019 (Stepochkin et al., 2023).

After backcrossing in the F₂ and F₃ generations, a selection of purple-grained hybrids homozygous for key genes of anthocyanin biosynthesis in the grain pericarp was carried out using phenotypic and molecular markers.

Genotyping of storage protein genes. Genotyping of hybrids during the breeding process was performed using genes encoding wheat grain storage proteins – gliadins (*Gli*). Gliadin-coding genes (GCGs) are characterized by multiple allelism, which allows for genotype recognition, and are localized in two of the seven chromosomes of the A genome and B genome (the *Gli-A1*, *Gli-A2*, *Gli-B1*, and *Gli-B2* loci on chromosomes 1A, 6A, 1B, and 6B, respectively). To identify alleles of gliadin-coding loci, we used polyacrylamide gel electrophoresis (PAGE) of the gliadin from individual grains in an acidic aluminum-lactate buffer (pH 3.2) (Laboratory Analysis..., 2013).

According to the catalog (Melnikova et al., 2012), the electrophoretic spectrum of each grain contains a group of polypeptides controlled by different alleles of gliadin-coding loci. Three to five grains per spike were analyzed. For each electrophoresis, the homo- or heterozygous state of the alleles for each locus is provided, as well as from which parents these alleles were inherited.

Cultivation, phenotyping and genotyping of hybrid material. Plant cultivation and hybridization were conducted in a hydroponic greenhouse at the shared-use center of the artificial plant cultivation laboratory and in the experimental fields of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (55°02'N, 82°56'E). The selection was performed in the F₂ and F₃ generations. The phenotypic markers such as anthocyanin coloration of vegetative and reproductive parts of plants, and DNA PCR markers for plants genotyping were used for selection.

The dark red coleoptiles coloration as a visual phenotypic marker of the dominant functional gene *Pp-B1* was used (Khlestkina et al., 2008). Coleoptile coloration in 4–5-day-old seedlings soaked in Petri dishes was assessed before planting. The purple coloration of the grains as a phenotypic marker for the complementary interaction of two key dominant genes, *Pp-B1* and *Pp3*, was used (Tereshchenko et al., 2012). Visual assessment of pericarp coloration in mature grains of the studied plants was performed after harvest.

DNA genotyping and molecular markers. DNA was extracted from young plant leaves using the method described by J. Plaschke et al. (1995). To analyze the isolated DNA of the hybrids and their parental forms, polymerase chain reaction (PCR) with intragenic markers developed for the anthocyanin biosynthesis genes *Pp-1* and *Pp3*, as well as microsatellite markers linked to these genes, were used (Table 2; Fig. S2).

Table 2. Molecular markers used in the work

Marker	Localization in chromosome	Annealing t°	Structure of forward and reverse primers (References)	Sizes of PCR products, bp
<i>Pp3-diagnostic</i>	2A	Tdn 65–56°	5'-TAGTGCCGTCTAACTGGTGA-3' 5'-ACGACGCCTAAGGAAACAC-3' (Shoeva et al. 2022)	429/398
<i>Xgwm0312</i>	2A	60°	5'-ATCGCATGATGCACGTAGAG-3' 5'-ACATGCATGCCTACCTAATGG-3' (Röder et al., 1998)	~200/230/300
<i>Pp1-diagnostic</i>	7B	Tdn 65–56°	5'-ATGGGGAGGAGGGCGT-3' 5'-TGCCGAGCGTGCTGTT-3' (Gordeeva et al., 2023)	478/434
<i>Xgwm0046</i>	7B	60°	5'-GCACGTGAATGGATTGGAC-3' 5'-TGACCCAATAGTGGTGGTCA-3' (Röder et al., 1998)	~180/190
<i>Xgwm0400</i>	7B	60°	5'-GTGCTGCCACCACTTGC-3' 5'-TGTAGGCACTGCTTGGGAG-3' (Röder et al., 1998)	~150/160/180
<i>Xgwm0537</i>	7B	60°	5'-ACATAATGCTTCCTGTGCACC-3' 5'-GCCACTTTTGTGTCGTTCT-3' (Röder et al., 1998)	~225/230

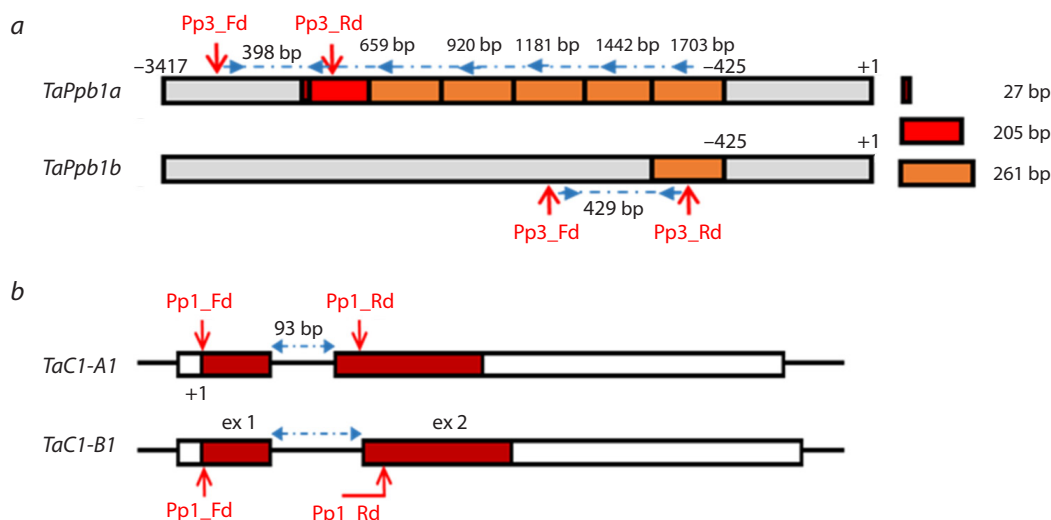


Fig. 1. Design of intragenic PCR-markers for DNA of the *Pp3* and *Pp1* genes.

a – scheme of differences in the structural organization of nucleotide sequences in the promoter region of functional *TaPpb1a* (NCBI accession number MG066455) and non-functional *TaPpb1b* (NCBI accession number MG066456: <https://www.ncbi.nlm.nih.gov/>) alleles of the *Pp3* gene (Purple pericarp) based on homology to the nucleotide sequence of the *Pp3* gene (NCBI accession number KJ747954). 261-nucleotide repeats of its part are shown in orange and red; *b* – schematic representation of differences in the structural organization of nucleotide sequences between the nonfunctional and functional alleles of *Pp-A1* and *Pp-B1* (syn. *TaC1-A1* and *TaC1-B1*; (Himi, Taketa, 2015)) on chromosomes 7A and 7B, respectively. The functional allele of *Pp-B1* has a longer region of a single intron between two exons, 28 base pairs longer than the corresponding region of the nonfunctional allele of *Pp-A1* (blue arrows). Red arrows indicate the binding sites of the PCR primers of the *Pp3*-diagnostic and *Pp1*-diagnostic markers.

To determine the substitution of chromosome regions and intervarietal inheritance of chromosomal rearrangements in hybrids, polymorphic SSR markers (simple sequence repeats) were used. These markers were selected from the group of microsatellite markers linked to the *Pp3* gene on chromosome 2A – marker *Xgwm0312*, and to the *Pp-B1* gene on chromosome 7B – markers *Xgwm0046*, *Xgwm0400*, *Xgwm0537* (Gatersleben Wheat Marker) (Röder, 1998; Khlestkina et al., 2008; Gordееva et al., 2023). The PCR conditions specified in the work of M.S. Röder et al. (1998) were followed. PCR products were separated on a 5 % HR (High resolution) agarose gel “HyAgarose™ HR Agarose” (ACTGene, Inc., Piscataway, NJ, USA).

To select wheat plants carrying dominant alleles, allele-specific diagnostic markers amplifying PCR products of different lengths together with microsatellites linked to key *Pp* genes were used in the study.

The intragenic allele-specific PCR marker *Pp3*-diagnostic, which we previously designed, allows us to clearly identify dominant and recessive alleles of the *Pp3* gene in tetra- and hexaploid wheat varieties, which makes it possible to select dominant alleles of this gene in a homozygous state (Shoeva et al., 2022). Previously, six tandem repeats of 261 nucleotides in the promoter of the dominant functional allele of the *Pp3* gene were identified, while only one such repeat in the recessive non-functional allele in white and red wheat varieties was found (Jiang et al., 2018).

Using the PrimerQuest™ Tool software, specific oligonucleotide primers were designed such that the forward primer *Pp3_Fd* anneals to the promoter region up to the start of the repeated regions, which differ in length and quantity between the dominant and recessive alleles, and the reverse primer *Pp3_Rd* anneals to the 261-nucleotide repeat. Due to the presence of truncated fragments of 27 and 205 nucleotides, the PCR products obtained using the developed primers differ in length between the dominant and recessive alleles, amounting to 398 and 429 nucleotide pairs, respectively (Fig. 1a).

The primers of the intragenic PCR marker *Pp1*-diagnostic flank the single intron of the *Pp-B1* gene, which varies in DNA length, on both sides. This allows for the simultaneous detection of both the dominant allele of the *Pp-B1* gene and the recessive non-functional alleles *pp-A1* and *pp-B1* (Fig. 1b). This marker was created using known nucleotide sequences of the *Rc* gene alleles (red coleoptile, a synonym of the *Pp* gene) isolated from the Chinese Spring wheat cultivar and the purple-grained isogenic line of the Novosibirskaya 67 variety (Himi, Taketa, 2015).

Diagnostic primers for the *Pp1*-diagnostic PCR marker were designed so that the forward primer *Pp1_Fd* annealed at the beginning of exon 1 for both the functional and non-functional alleles, while the reverse primer *Pp1_Rd* annealed 161 nucleotides from the beginning of exon 2. Due to differences in the length of the intron regions, the PCR products obtained using our primers

Table 3. Characteristics of the emmer hybrid plants based on the inheritance of gliadins from parental forms

Plant	<i>Gli-A1</i> (Chromosome A1S)	<i>Gli-B1</i> (Chromosome B1S)	<i>Gli-A2</i> (Chromosome A6S)	<i>Gli-B2</i> (Chromosome B6S)
Hybrid 27-15/3	k-25516	k-25516	Gremme	k-25516
Hybrid 27-15/4				
Hybrid 27-15/8*			k-25516	
Hybrid 27-15/10			Gremme	
Hybrid 31-16/6	Tri15744		k-25516	Tri15744
Hybrid 31-19/3	k-25516		Gremme	k-25516
Hybrid 31-19/4				
Hybrid 31-19/5				

* Inheritance of alleles of all four gliadin-coding loci from emmer *T. dicoccum* k-25516.

differed in length between the non-functional and functional alleles, amounting to 472 and 434 nucleotides, respectively.

The recessive *pp-B1* alleles on chromosome 7B did not differ in the length of their PCR marker products from the non-functional *pp-A1* alleles. In this case, standard DNA polymerase was used to perform PCR analysis; 2 % agarose gel (LE Agarose, Lonza Rockland, Inc., Rockland, ME, USA) prepared in TAE buffer (40 mM Tris-HCl pH 8.0, 20 mM sodium acetate, 1 mM EDTA) with the addition of ethidium bromide was used to separate PCR products.

The 100 bp DNA fragment length marker (MEDIGEN Laboratory LLC) served as the standard. Electrophoresis was performed in a horizontal chamber for 1–5 hours at 7 volts/cm. UV visualization and image analysis were performed using the Molecular Imager® Gel DocTM XR+ System (Bio-Rad Laboratories, Inc., Hercules, CA, USA).

Results

Selection of hybrid parent plants based on gliadin-coding genes

Initially, the parental lines were studied: *T. durum* Tri15744; Gremme variety, as well as *T. dicoccum* k-25516; the GCGs alleles were determined for each of the loci. The spectra of the parental samples differed from each other, i. e. they were controlled by different allelic variants of gliadin-coding loci, and were homogeneous, i. e. each sample had a single spectrum. During the electrophoretic analysis of the hybrid material, these samples (*T. durum* Tri1574, Gremme v., *T. dicoccum* k-25516) were used as standards necessary for the precise determination of the belonging of a block of components (allelic variant of a particular locus) to one of the three parental samples.

For plant genotyping using the GCGs method, 10 plants from each of three early-ripening purple-grained hybrid lines were selected: line 27-12, semi-hulled, with all normal, awnless spikes; line 31-16, nearly naked, the spikes with rudimentary awns at the top; line 31-19, semi-hulled, the spikes with an awned top (Stepochkin et al., 2023). Three to five grains from each plant were analyzed. The results of the study are presented in Table 3.

We haven’t provided genetic formulas for gliadins, as durum wheat and emmer have different nomenclatures – different catalogs of component blocks, in which the same letters denote different blocks. These catalogs haven’t yet been compiled into a unified system, thus, to ensure complete accuracy, the parent from which a particular component block/allelic variant of the gliadin-coding locus was inherited is named for each locus.

First, plants inheriting gliadin-coding genes from emmer *T. dicoccum* k-25516 were selected, and then those from naked emmer Gremme variety were selected. Only one hybrid plant, 27-15/8 (highlighted with * in Table 3, No. 4 in this paper), inherited all four gliadin-coding loci from emmer k-25516. Hybrid plant 31-16/6 (No. 6 in this paper) inherited the *Gli-B1* and *Gli-A2* loci from emmer k-25516, and the *Gli-A1* and *Gli-B2* loci from *T. durum* Tri1574. Plants of line 31-19 inherited 3 loci (*Gli-A1*, *Gli-B1* and *Gli-B2*) from emmer k-25516 and one (*Gli-A2*) from naked emmer Gremme variety.

Selection by phenotypic traits

In addition to the purple coloration of the grain’s pericarp, plants of the hybrid lines showed intense coloration of the coleoptiles and stems (Fig. 2).

Phenotypic traits of line 27-15/8 plants (No. 4 in Figure 2) resembled emmer k-25556 (No. 3), with long, red stems at maturity and awnless red spikes. The flat



Fig. 2. Photograph of stems, spikelets and grains of parental lines and early-ripening hybrids selected for backcrossing with the Gremme variety.

1 – purple-grained *T. durum* Tri15744; 2 – white-grained Gremme variety; 3 – red-grained emmer *T. dicoccum* k-25516; 4 – hybrid No. 27-15/8 (pedigree: Tri15744/Gremme// $\times$ 2 k-25516); 5 – hybrid No. 31-16/6 (Tri15744/Gremme//k-25516).

skeletons contained two dark purple, semi-hulled, difficult-to-thresh grains.

Phenotypic traits of the dwarf line 31-16/3 plants (No. 6 in Figure 2) included rudimentary awns at the top of their compact, club-shaped, brittle ears and nearly naked grains.

Based on phenotypic traits, tall, thin-stemmed plants of line 31-19 had rudimentary awns at the top of compact, club-shaped, brittle spikes and were nearly naked. A negative trait – lodging tendency and a lower anthocyanin content in the grains (39.7 $\mu\text{g/g}$) compared to the donor *T. durum* Tri15744 (68.4 $\mu\text{g/g}$) (Stepochkin et al., 2023) – prevented us from further study of plants in this line.

Mid-season white-spiked plants of line 27-3 demonstrated the highest anthocyanin content in the grains (82.5 $\mu\text{g/g}$) under field conditions in 2019 (Stepochkin et al., 2023) and were also selected for backcrossing with more higher yielding, naked-grained variety Gremme.

Selection by coleoptile color

Preliminary phenotypic (visual) assessment of coleoptile coloration in the parental plants showed that seedlings of awnless emmer k-25516, like the tetraploid donor line *T. durum* Tri15744, had pronounced anthocyanin coloration of coleoptiles, which corresponded to the

presence of dominant alleles of the *Pp-B1* gene (Fig. S3). The white-grained Gremme variety had weakly colored coleoptiles, similar to that of the hexaploid wheat variety Saratovskaya 29, associated with the presence of dominant alleles of the *Pp-A1* gene, which have a weak effect on anthocyanin biosynthesis in the grains (Gordeeva et al., 2015).

After direct and reverse backcrosses of the wheat-emmer hybrids with the awned, naked-grain Gremme variety, the F_1 plants were self-pollinated. Sixty grains from F_1 plants from each cross were soaked for germination in Petri dishes. In the F_2 generation, after crossing with the Gremme variety, which carries a recessive allele of the *Pp-B1* gene and, as a result, does not exhibit bright anthocyanin coloration of the coleoptiles, a visual assessment of seedling coloration was initially performed. After the F_2 hybrid coleoptiles had germinated for 4–5 days, ungerminated grains and seedlings with colorless coleoptiles were discarded for planting in the ground. Only seedlings with anthocyanin-colored coleoptiles were selected for sowing.

Molecular marker-assisted selection (MAS)

In the tetraploid wheat *T. durum* Tri15744, anthocyanin biosynthesis in the pericarp is controlled by complementarily interacting genes *Pp3* on chromosome 2A and *Pp-A1/B1* on chromosome 7A or 7B.

Initially, SSR markers on DNA extracted from the leaves of parental samples of purple-grained wheat-emmer hybrids were tested. Microsatellite markers *Xgwm0046* and *Xgwm0537*, located close to the target gene *Pp-B1* on chromosome 7B, showed no differences between the PCR products of emmer samples k-25516 with colored coleoptiles and the Gremme variety with uncolored coleoptiles, and, consequently, no differences between the dominant and recessive alleles of *Pp-B1* (Fig. 3a, b). Based on the results of PCR analysis with the *Xgwm0400* polymorphic marker, all three hybrid lines (No. 4, 5, 6 in Table 1) selected for crossing with the Gremme variety inherited dominant *Pp-B1* alleles only from the emmer k-25516 (Fig. 3c). For DNA genotyping of hybrids in the F_2 generation, the *Xgwm0400* polymorphic marker was selected. Further, for a more precise selection of purple-grained hybrids in the F_3 generation, the *Pp1-diagnostic* intragenic PCR marker was used (Fig. 3f).

Testing of the microsatellite marker *Xgwm0312*, linked to the *Pp3* gene on chromosome 2A of the tetraploid wheat genome, revealed significant polymorphism and discrepancy between the PCR products of the purple-grained parental samples and the donor line of the functional *Pp3* gene, *T. durum* Tri15744 (Fig. 3d). Therefore,

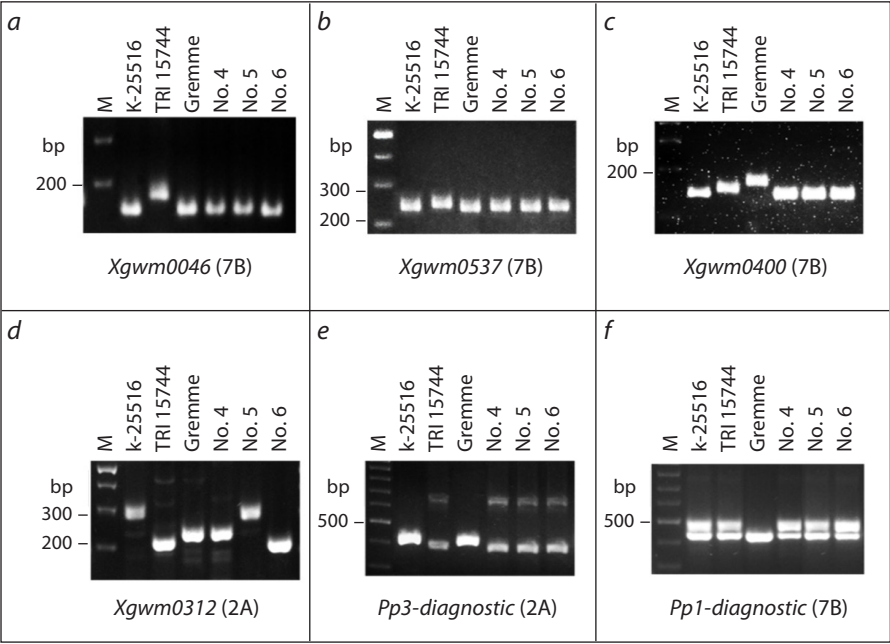


Fig. 3. Electropherograms of PCR products obtained using microsatellite markers linked to the *Pp-B1* and *Pp-3* genes on chromosomes 7B and 2A, respectively, on 5 % agarose gel (a–d). Electropherograms of PCR products obtained using intragenic PCR markers to the *Pp-3* and *Pp-B1* genes on chromosomes 2A and 7B, respectively, on 2 % agarose gel (e–f). No. 4, 5, 6 – the DNA products of the hybrids selected for crossing with the Gremme variety; M – the length marker.

selection of homozygous purple-grained hybrid samples in the F₂ and F₃ generations was carried out using the polymorphic intragenic PCR marker *Pp3-diagnostic*, which we previously developed (Fig. 3e).

After evaluating the coleoptile and grain color of F₂ plants in three pairs of direct and backcross combinations, we selected five purple-grained samples from 113 plants analyzed using molecular analysis. These samples contained both dominant alleles of the *Pp-B1* and *Pp3* genes (Table 4). The grains of one plant were nonviable.

Among the 24 mature F₂ plants from the cross between hybrid No. 4 and the white-grained Gremme variety, not a single dark-purple-grained plant carrying homozygous markers for both functional genes, *Pp-B1* and *Pp3*, was found, with the PCR products corresponding only to No. 4. Among the 14 F₂ plants from the backcross between the Gremme variety and emmer hybrid No. 4, two plants with dark-purple grains were found carrying homozygous markers for both functional genes, *Pp-B1* and *Pp3*. These plants (7 and 14) (Fig. S4) had loose spikes containing five and two grains, respectively, which were sown for further generation.

Thus, using molecular analysis, we were able to select only 4 viable plants with dark purple grains, homozygous for both dominant genes, *Pp-B1* and *Pp3* in the F₂ generation. Two plants were selected from a cross between

the white-grained Gremme variety and the emmer hybrid No. 4, and two plants were from a cross between emmer hybrid No. 6 and the Gremme variety (Table 4).

For F₃ planting and further analysis, dark purple grains were chosen from plants representing three groups:

- 1) with homozygous dominant alleles of *Pp3* and heterozygous for the SSR marker *Xgwm400*, linked to the *Pp-B1* gene;
- 2) with dominant alleles of *Pp3*, which are in a heterozygous state and homozygous for the *Xgwm400* marker;
- 3) highly productive plants heterozygous for both the *Pp3* alleles and the *Xgwm400* marker.

For the control and further study, seeds of a white-grained plant with recessive *Pp3* alleles and a homozygous dominant SSR marker *Xgwm400* linked to the *Pp-B1* gene were selected and planted, as well as seeds of a white-grained plant with homozygous dominant *Pp3* alleles and a homozygous recessive SSR marker *Xgwm400*. Phenotypic and molecular analysis of the plants continued in the F₃ generation (Table 5).

In the first cross combination of hybrid No. 4 with the Gremme variety among 42 hybrids, six dark purple grained plants, homozygous for the *Pp3* gene, carrying the marker of the dominant functional allele *Pp-B1* were selected by using *Pp3-diagnostic*, *Pp1-diagnostic* and SSR marker *Xgwm0400* (PCR products corresponded

Table 4. DNA analysis of F₂ plants from crossing hybrids with the Gremme variety

No.	Forward cross		Backcross	
	Samples analyzed	Homozygous samples (<i>Pp-B1</i> * + <i>Pp3</i>)	Samples analyzed	Homozygous samples (<i>Pp-B1</i> * + <i>Pp3</i>)
4	24	0	14	2
5	18	0	14	0
6	13	2	18	1 (sterile)
Total		2		3

* *SSR Xgwm400* served as a marker for identifying homozygous alleles of *Pp-B1*.

Table 5. DNA analysis of F₃ plants from crossing hybrids with Gremme variety

No.	Forward cross		Backcross	
	Samples analyzed	Homozygous samples (<i>Pp-B1</i> * + <i>Pp3</i>)	Samples analyzed	Homozygous samples (<i>Pp-B1</i> * + <i>Pp3</i>)
4	42	6	47	5
5	53	5	28	5
6	63	11	20	3
Total		22		13

* *Pp1-diagnostic* served as a marker for identifying homozygous alleles of *Pp-B1*.

to No. 4) (Fig. S5). It was noted that one plant with dark purple grains had recessive PCR products of the SSR marker *Xgwm400* linked to the *Pp-B1* gene from the Gremme variety, and dominant PCR products of the intragenic marker *Pp1-diagnostic* (Fig. S5). This fact indicates incomplete linkage of the microsatellite marker with the target gene and the existence of a hot recombination point located between the *Pp-B1* gene and the SSR marker *Xgwm400*.

In a backcross combination of the Gremme variety with line No. 4 in F₃, based on the results of phenotypic and molecular analysis, five dark-purple-grained plants homozygous for the *Pp3* gene and carrying the marker of the dominant allele *Pp-B1* (PCR products corresponded to No. 4) were selected from 47 hybrid samples. In this cross combination, one white-grained plant was noted with homozygous dominant alleles *Pp3* and homozygous for the SSR marker *Xgwm400*, linked to the *Pp-B1* gene, but lacking the PCR product of the marker *Pp1-diagnostic*, corresponding to the dominant allele *Pp-B1* (PCR products as in No. 4). This fact indicated that recombination occurred on chromosome 7B between the *Pp-B1* gene and the SSR marker *Xgwm400*.

Thus, using molecular analysis, 35 dark purple grain plants homozygous for the *Pp3* gene and containing the dominant *Pp-B1* allele were selected in the F₃ generation (Table 5). Grains from 29 productive plants were planted in the field, giving rise to new purple grain lines. In the

F₄ generation, four of the 29 lines selected exhibited a segregation for purple/white grain color (three lines inherited from family No. 5, and one line inherited from family No. 6). This indicated a heterozygous state of the dominant allele of the *Pp-B1* gene in the parental samples in the F₃ generation. The remaining 25 wheat-emmer hybrid lines were constant for the purple color of grains, of which seven lines had hard glumes and were difficult to thresh. The purple grains for the F₄ generation giving rise to new hybrid lines were sown in the field for the selection of economic value. That is beyond the scope of this paper.

The time spent crossing the plants and selecting the appropriate hybrids was two years: three greenhouse growing seasons and two field growing seasons.

Discussion

Enriching the genetic material of tetraploid wheat with new, unique characteristics associated with the accumulation of anthocyanins in the grains and long-forgotten emmer varieties with valuable economic properties allows for expanding the industry and providing a better range of healthy food products.

Gliadin-encoding genes

This study examined wheat-emmer hybrids obtained previously through saturation crosses to increase the emmer genetic material, along with the naked trait.

Genotyping of the hybrids during the breeding process was performed using genes encoding grain storage proteins – gliadins (*Gli*). Gliadins are controlled by four major unlinked loci on chromosomes 1 and 6 of the homeologous group (loci *Gli-A1*, *Gli-A2*, *Gli-B1*, *Gli-B2* on chromosomes 1A, 6A, 1B, 6B, respectively) (Metakovsky et al., 2019; Vinichenko, Salina, 2021). Each locus is a cluster of tightly linked genes. Genes within a single locus encode polypeptide components of different molecular weights and charges. Components controlled by a single allele are inherited in a linked manner, as a single Mendelian trait, forming a polypeptide block. The block structure is stable and does not depend on external or internal factors.

Multiple alleles have been identified for gliadin-coding loci: each locus controls over 20 alleles, allowing the theoretical identification of over one million genotypes. In fact, each variety and each line has its own unique electrophoretic spectrum, controlled by different alleles of four gliadin-coding loci (Kudryavtsev et al., 2014; Novoselskaya-Dragovich, 2015). This allows us to identify genotypes and trace the inheritance of parental alleles across generations of hybrids, determining their affiliation with specific samples at each locus.

Genotyping for gliadin-coding genes, thanks to the uniqueness of parental genotypes, makes it possible to select hybrid offspring that inherit the genotype of a specific parent, with a characteristic set of alleles at gliadin-coding loci. Analysis of the allelic composition of gliadin-coding loci in hybrid plants allowed us to determine from which parent the genes for each gliadin-coding locus were inherited, i. e. to determine the degree of hybridity in lines during their creation.

To obtain a wide range of naked, purple-grained hybrids for crossing with the naked emmer Gremme variety, contrasting hybrid lines differing in their gliadin-coding gene content were selected. One of the selected lines, awnless line No. 4, inherited all four *Gli* loci from the awnless emmer *T. dicoccum* k-25516. The nearly naked, low-yielding line No. 6 inherited two loci (*Gli-B1* and *Gli-A2*) from the awnless emmer k-25516, and two (*Gli-A1* and *Gli-B2*) from the tetraploid wheat *T. durum* Tri1574.

Crossbreeding and phenotyping

Pyramiding multiple genes within one plant genotype is a complex process. The primary focus at this stage of breeding was on producing plants with purple-colored grains.

Anthocyanin pigmentation in wheat is regulated by two transcription factors. The R2R3-Myb-like factor, which activates anthocyanin biosynthesis in both vegetative organs (coleoptile, stem) and pericarp, is encoded by a functional allele of the *Pp-B1* gene (also known as

Rc-1 or *Red coleoptile*) (Himi, Taketa, 2015). Coleoptile coloration occurs if at least one allele of the *Pp-B1* gene is dominant (Khlestkina et al., 2008). Therefore, to select plants inheriting dominant *Pp-B1* alleles, visual analysis of seedlings was used before sowing. Only samples with colored coleoptiles were selected from these seedlings. The functional allele of the *Pp3* gene, encoding the second regulatory factor Myc, has an endemic origin from the tetraploid wheat *T. aestivum* (Zeven, 1991). Anthocyanin biosynthesis in grains occurs only in the presence of complementarily interacting genes *Pp-B1* (or *Pp-D1*) and *Pp3* in one genotype (Tereschenko et al., 2012). While phenotyping is undoubtedly beneficial for plant breeding, the difficulty is that *Pp* regulatory genes are dominant and, therefore, capable of expressing coloration in seedling coleoptiles and grain pericarps in both homozygous and heterozygous states.

Molecular markers

Molecular PCR markers were used to select plants homozygous for both dominant alleles of the *Pp3* and *Pp-B1* genes, which exhibit stable purple coloration of the grains.

Testing of SSR markers on the DNA of parental samples of purple-grained wheat-emmer hybrids did not demonstrate 100 % correspondence of inheritance of the association of PCR products of microsatellites with the *Pp* genes. Thus, the polymorphic microsatellite marker *Xgwm0312*, linked to the *Pp3* gene on chromosome 2A, showed a discrepancy between the PCR products of purple-grained parental accessions and the donor line of the functional *Pp3* gene *T. durum* Tri15744 (Fig. 3d). Therefore, the polymorphic intragenic PCR marker *Pp3-diagnostic*, which we previously developed for hexaploid bread wheat, was successfully suited for molecular analysis during the selection of homozygous purple-grained samples of hybrids (Fig. 3e, and Fig. S4–S5). This made it possible to accurately select plant samples homozygous for the dominant alleles of the *Pp3* genes in the F₂₋₃ generations.

Testing of dark purple hybrid lines (No. 4, 5, and 6) used for backcrosses using the polymorphic SSR marker *Xgwm0400* revealed the inheritance of a region of chromosome 7B, and consequently, functional alleles of the *Pp-B1* gene, from the DNA of red-stemmed emmer k-25516 (Fig. 3c). This marker was used for genotyping the DNA of hybrids in the F₂ generation. However, despite the proximity of this marker to the *Pp-B1* gene under study, cases of incomplete linkage associated with the presence of a hotspot located between them must be considered.

In contrast to the SSR markers *Xgwm0044* and *Xgwm0111* on chromosome 7D, which demonstrated reliable linkage to the equivalent *Pp-D1* gene locus

during breeding of hexaploid purple-grained wheat (Gordeeva et al., 2020), the probability of linkage of the *Xgwm0400* marker to the *Pp-B1* gene locus was extremely low. This fact is consistent with the work of E. Khlestkina et al. (2008), where a short region of chromosome 7B with the target *Rc-B1* (= *Pp-B1*) locus, which did not contain any of the nearby mapped SSR markers, was identified in inbred recombinant lines.

A hotspot breakpoint between the microsatellite marker *Xgwm0312* and the *Pp3* gene locus on chromosome 2A in hybrid lines inheriting introgression from the wild tetraploid species *T. timopheevii* Zhuk. was demonstrated by O.Y. Tereshchenko et al. (2012). This is apparently explained by a wide spectrum of genetic diversity in tetraploid wheats compared to their domesticated hexaploid relatives. Therefore, to select homozygous purple-grained hybrid samples in the F_3 generation, we used the polymorphic intragenic marker *Pp1-diagnostic* (Fig. 3f, and Fig. S5), which we had previously designed for hexaploid bread wheat.

Since the recessive *pp-B1* alleles on chromosome 7B did not differ in the length of their PCR marker products from the non-functional *pp-A1* alleles, only after verifying the phenotyping of F_4 generation grains were the final plant families selected that would yield new purple-grained lines homozygous simultaneously for two functional alleles of the *Pp3* and *Pp-B1* genes with stable, purple-colored anthocyanin-stained grains.

According to the literature, purple-grained tetraploid wheats *T. durum* Desf. were bred in Italy as a niche product by crossing the commercial durum wheat variety Primadur (Blondur//2587-8-6/Leeds) and *T. durum* Ethiopian variety T1303 (PI352395) having purple pericarp (Ficco et al., 2014, 2018). Italian scientists have also recently developed molecular allele-specific markers of the *Pp-B1* and *Pp-A3* (*Pp3*) gene loci for marker-associated breeding of purple-grained durum wheat (Esposito et al., 2024). The markers were constructed based on a comparative analysis of the primary DNA of the genotypes of purple-grained wheat T1303 and local yellow-grained durum wheat varieties Primadur and Svevo.

Similar to the results obtained in bread wheat (Jiang et al., 2018), six tandem repeats 261 bp in length were detected in the promoter region of the functional dominant allele *TdPpb1* (*Pp3*) on chromosome 2A in purple-grained T1303, while Svevo had only one repeat. The size of the PCR products on DNA obtained from purple-grained wheat was 3,446 bp long, compared to 2,141 bp in yellow-grained varieties.

The nonfunctional *TdPpm1b* (*Pp-B1*) allele in the Primadur and Svevo varieties revealed a large insertion region of approximately 1.6 kbp, 56 bp from ATG,

which disrupted the first exon. S. Esposito et al. (2024) selected primers that spanned this insertion. The PCR products obtained from the functional alleles (*TdPpm1a*) in purple-grained genotypes were approximately 500 bp in length, while those from the nonfunctional alleles (*TdPpm1b*) in local yellow-grained varieties and lines were 1,800 bp.

The intragenic allele-specific markers that we designed, *Pp1-diagnostic* and *Pp3-diagnostic*, have broader applicability and are suitable for diagnosing both hexaploid and tetraploid wheat varieties. The primers proposed in the articles by W. Jiang et al. (2018) and S. Esposito et al. (2024) amplify long PCR products that require a specialized, highly efficient polymerase for synthesis. Standard DNA polymerase is suitable for our markers.

The primers proposed in this study for the intragenic PCR marker *Pp1-diagnostic*, starting from the beginning of the first exon ATG, span a longer, variable-length region of the *Pp-1* gene DNA. However, we did not detect long amplicons of the *Pp1-diagnostic* marker associated with the presence of a long insertion in the white-grained variety Gremme. It would be interesting to test the presence of this insert in cultivated varieties of durum wheat from Siberian selection.

Using a greenhouse complex during the off-season allowed us to accelerate selection and produce F_4 hybrids in two years (three greenhouse and two field growing seasons). A disadvantage of greenhouse cultivation was the limited space available for sowing seeds. This resulted in a double molecular analysis and selection of the hybrid plants with the qualities we need in the F_{2-4} generations.

Conclusion

In this study, after genotyping previously developed purple-grained, difficult-to-thresh, and low-yielding emmer hybrids for their grains storage protein genes, backcrossing with the naked-grained Gremme variety and MAS selection, a set of three families of spring, naked-grained and semi-naked, purple-grained wheat-emmer plants was obtained, differing in gliadin-coding genes and other qualitative traits.

Twenty-five wheat-emmer hybrid lines were finally obtained in the F_4 generation. These lines differed in morphological traits and were consistent in the purple anthocyanin coloration of grain pericarp. This coloration is expressed by the presence of homozygous functional alleles of the *Pp-B1* and *Pp3* genes in the genotype. The presence of only one functional allele or heterozygosity leads to segregation by grains color, followed by loss of anthocyanin pigment.

The intragenic PCR markers for the *Pp-B1* and *Pp3* genes that we previously designed based on the bread wheat genome demonstrated high efficiency. In small cultivation areas, the use of *Pp3-diagnostic* allowed for the immediate and accurate identification of homozygous genotypes of tetraploid wheat plants. The use of *Pp1-diagnostic* required subsequent verification of the grain color in the offspring to select samples with consistent color. These markers are a useful tool for future breeding programs.

References

- Badaeva E.D., Keilwagen J., Knüpfer H., Waßermann L., Dedkova O.S., Mitrofanova O.P., Kovaleva O.N., Liapunova O.A., Pukhalskiy V.A., Özkan H., Graner A., Willcox G., Kilian B. Chromosomal passports provide new insights into diffusion of emmer wheat. *PLoS One*. 2015;10(5):e0128556. doi 10.1371/journal.pone.0128556
- Biradar S.S., Yashavanthakumar K.J., Navathe S., Reddy U.G., Bavisar V.S., Gopalareddy K., Lamani K., Desai S.A. Dicoccum wheat: current status and future perspectives. In: New Horizons in Wheat and Barley Research. Singapore: Springer, 2022;531-563. doi 10.1007/978-981-16-4449-8_21
- Cabas-Lühmann P., Arriagada O., Matus I., Marcotuli I., Gadaleta A., Schwember A.R. Comparison of durum with ancient tetraploid wheats from an agronomical, chemical, nutritional, and genetic standpoints: a review. *Euphytica*. 2023;219(6):61. doi 10.1007/s10681-023-03188-z
- Dwivedi S.L., Mattoo A.K., Garg M., Dutt S., Singh B., Ortiz R. Developing germplasm and promoting consumption of anthocyanin-rich grains for health benefits. *Front Sustain Food Syst*. 2022;6:867897. doi 10.3389/fsufs.2022.867897
- Esposito S., Palombieri S., Vitale P., Angione G., D'Attilia C., Taranto F., Sestili F., De Vita P. Identification and development of functional markers for purple grain genes in durum wheat (*Triticum durum* Desf.). *Theor Appl Genet*. 2024;137:256. doi 10.1007/s00122-024-04755-1
- Ficco D.B., Mastrangelo A.M., Trono D., Borrelli G.M., De Vita P., Fares C., Beleggia R., Platani C., Papa R. The colours of durum wheat: a review. *Crop Pasture Sci*. 2014;65(1):1-15. doi 10.1071/CP13293
- Ficco D.B.M., Borrelli G.M., Giovanniello V., Platani C., De Vita P. Production of anthocyanin-enriched flours of durum and soft pigmented wheats by air-classification, as a potential ingredient for functional bread. *J Cereal Sci*. 2018;79:118-126. doi 10.1016/j.jcs.2017.09.007
- Fisenko A.V., Kalmykova L.P., Kuznetsova N.L., Kuzmina N.P., Ermolenko O.I., Upelnik V.P. Selection of purple-grain common wheat and its technological properties. *Agrarnaya Rossiya*. 2020; 10:43-48. doi 10.30906/1999-5636-2020-10-43-48 (in Russian)
- Geyik Ö.G., Tekin-Cakmak Z.H., Shamanin V.P., Karasu S., Pototskaya I.V., Shepelev S.S., Chursin A.S., Morgounov A.I., Yaman M., Sagdic O., Koksels H. Effects of phenolic compounds of colored wheats on colorectal cancer cell lines. *Qual Assur Saf Crops Food*. 2023;15(4):21-31. doi 10.15586/qas.v15i4.1354
- Gilev S.D., Tsybalenko I.N., Kopylov A.N., Filippova E.A., Kozlova T.A. Emmer wheat is a promising grain crop for organic agriculture. *Zernovoye Khozyajstvo Rossii = Grain Economy of Russia*. 2018;4:6-11. doi 10.31367/2079-8725-2018-58-4-6-11 (in Russian)
- Gordeeva E.I., Shoeva O.Y., Khlestkina E.K. Marker-assisted development of bread wheat near-isogenic lines carrying various combinations of purple pericarp (*Pp*) alleles. *Euphytica*. 2015;203(2):469-476. doi 10.1007/s10681-014-1317-8
- Gordeeva E., Shamanin V., Shoeva O., Kukoeva T., Morgounov A., Khlestkina E. The strategy for marker-assisted breeding of anthocyanin-rich spring bread wheat (*Triticum aestivum* L.) cultivars in Western Siberia. *Agronomy*. 2020;10(10):1603. doi 10.3390/agronomy10101603
- Gordeeva E.I., Shoeva O.Y., Shamanin V.P., Khlestkina E.K. The molecular markers applying in breeding of spring bread wheat (*Triticum aestivum* L.) lines with different anthocyanin coloration of the grains. *Pisma v Vavilovskii Zhurnal Genetiki i Selekcii = Lett Vavilov J Genet Breed*. 2023;9(2):86-99. doi 10.18699/LettersVJ-2023-9-11 (in Russian)
- Himi E., Taketa S. Isolation of candidate genes for the barley *Ant1* and wheat *Rc* genes controlling anthocyanin pigmentation in different vegetative tissues. *Mol Genet Genomics*. 2015;290(4):1287-1298. doi 10.1007/s00438-015-0991-0
- Jiang W., Liu T., Nan W., Jeewani D.C., Niu Y., Li C., Wang Y., Shi X., Wang C., Wang J., Li Y., Gao X., Wang Z. Two transcription factors *TaPpm1* and *TaPpb1* co-regulate anthocyanin biosynthesis in purple pericarps of wheat. *J Exp Bot*. 2018;69(10):2555-2567. doi 10.1093/jxb/ery101
- Khlestkina E.K., Röder M.S., Salina E.A. Relationship between homoeologous regulatory and structural genes in allopolyploid genome – a case study in bread wheat. *BMC Plant Biol*. 2008;8(1):88. doi 10.1186/1471-2229-8-88
- Khlestkina E.K., Röder M.S., Börner A. Mapping genes controlling anthocyanin pigmentation on the glume and pericarp in tetraploid wheat (*Triticum durum* L.). *Euphytica*. 2010;171(1):65-69. doi 10.1007/s10681-009-9994-4
- Kudryavtsev A.M., Dedova L.V., Melnik V.A., Shishkina A.A., Upelnik V.P., Novoselskaya-Dragovich A.Yu. Genetic diversity of modern Russian durum wheat cultivars at the gliadin-coding loci. *Russ J Genet*. 2014;50(5):483-488. doi 10.1134/S1022795414050093
- Kuznetsova E.A., Uchasov D.S., Shayapova L.V., Kuznetsova E.A. Study of the composition and properties of *Triticum dicoccum* (Schränk) grain. *Vsyo o Myase*. 2020;(S5):188-190. doi 10.21323/2071-2499-2020-S5-188-190 (in Russian)
- Laboratory Analysis of Wheat Seed Proteins: instruction manual. Moscow: Vash Format Publ., 2013 (in Russian)
- Liu J., Zhou H., Song L., Yang Z., Qiu M., Wang J., Shi S. Anthocyanins: promising natural products with diverse pharmacological activities. *Molecules*. 2021;26(13):3807. doi 10.3390/molecules26133807
- Melnikova N.V., Kudryavtseva A.V., Kudryavtsev A.M. Catalogue of alleles of gliadin-coding loci in durum wheat (*Triticum durum* Desf.). *Biochimie*. 2012;94(2):551-557. doi 10.1016/j.biochi.2011.09.004
- Metakovsky E., Melnik V.A., Pascual L., Wrigley C.W. Gliadin genotypes worldwide for spring wheats (*Triticum aestivum* L.) 2. Strong differentiation of polymorphism between countries and regions of origin. *J Cereal Sci*. 2019;87(6):311-317. doi 10.1016/j.jcs.2019.04.015
- Mohammadi N., Farrell M., O'Sullivan L., Langan A., Franchin M., Azevedo L., Granato D. Effectiveness of anthocyanin-containing foods and nutraceuticals in mitigating oxidative stress, inflammation, and cardiovascular health-related biomarkers: a systematic review of animal and human interventions. *Food Funct*. 2024;15(7):3274-3299. doi 10.1039/D3FO04579J
- Novoselskaya-Dragovich A.Y. Genetics and genomics of wheat: storage proteins, ecological plasticity, and immunity. *Russ J Genet*. 2015;51(5):476-490. doi 10.1134/S102279541505004X
- Plaschke J., Ganai M.W., Röder M.S. Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor Appl Genet*. 1995;91:1001-1007. doi 10.1007/BF00223912
- Röder M.S., Korzun V., Wendehake K., Plaschke J., Tixier M.H., Leroy P., Ganai M.W. A microsatellite map of wheat. *Genetics*. 1998; 149:2007-2023. doi 10.1093/genetics/149.4.2007
- Rubets V.S., Voronchikhina I.N., Igonin V.N., Sidorenko V.S., Voronchikhin V.V. Characteristics of violet-green variety of spring soft wheat in the conditions of the central region of the Non-Chernozem

- Zone of Russia. *Mezhdunarodnyi Sel'skokhozyaistvennyi Zhurnal*. 2022;5:525-529. doi 10.55186/25876740_2022_65_5_525 (in Russian)
- Şahin Y., Karakas F.P. Chemical composition of einkorn (*Triticum monococcum* ssp. *monococcum*), emmer (*Triticum dicoccum*), and spelt (*Triticum spelta*). In: Ancient Wheats. Cham: Springer, 2022; 119-145. doi 10.1007/978-3-031-07285-7_6
- Shamanin V.P., Pototskaya I.V., Chursin A.S., Shepelev S.S., Nardin D.S., Pozherukova V.E., Köksel H., Morgounov A.I. Breeding spring bread wheat (*Triticum aestivum* L.) varieties with functional properties of grain for environmentally friendly growing in Western Siberia. *Sel'skokhozyaistvennaya Biologiya = Agricultural Biology*. 2024;59(3):492-506. doi 10.15389/agrobiology.2024.3.492eng
- Shoeva O.Yu., Gordeeva E.I., Khlestkina E.K. Intragenic DNA marker for the selection of wheat with a high content of anthocyanins in the grain pericarp. Patent RU2774444C1. 2022 (in Russian)
- Smekalova T.N., Kobylansky V.D. A new subspecies of wheat: *Triticum dicoccon* (Schränk) Schuebl. subsp. *nudicoccon* Kobyl. et Smekal. *Trudy po Prikladnoi Botanike, Genetike i Seleksii = Proc Appl Bot Genet Breed*. 2019;180(4):148-151. doi 10.30901/2227-8834-2019-4-148-151 (in Russian)
- Stepochkin P.I., Gordeeva E.I., Khlestkina E.K. Marker-assisted breeding of hybrid lines of *Triticum dicoccon* (Schränk) Schuebl. × *Triticum aethiopicum* Jakubz. with purple grain. *Trudy po Prikladnoi Botanike, Genetike i Seleksii = Proc Appl Bot Genet Breed*. 2023;184(2):139-148. doi 10.30901/2227-8834-2023-2-1-10 (in Russian)
- Temirbekova S.K., Begeulov M.Sh., Afanas'eva Yu.V., Kulikov I.M., Ionova N.E. Adaptive capacity of einkorn huskless in the second, third and seventh Russian Federation regions. *Vestnik Rossijskoi Sel'skokhozyajstvennoi Nauki = Vestn Russ Agric Sci*. 2020; 1:34-38. doi 10.30850/vrsn/2020/1/34-38 (in Russian)
- Temirbekova S.K., Begeulov M.Sh., Baida I.D., Zverev S.V., Politukha O.V. Technological properties of grain of new varieties of ancient wheat species. *Agrarnaya Nauka = Agrarian Science*. 2024; 11:157-163. doi 10.32634/0869-8155-2024-388-11-157-163 (in Russian)
- Tereshchenko O.Y., Gordeeva E.I., Arbuzova V.S., Börner A., Khlestkina E. The D genome carries a gene determining purple grain colour in wheat. *Cereal Res Commun*. 2012;40(3):334-341. doi 10.1556/CRC.40.2012.3.2
- Tikhonova M., Shoeva O., Tenditnik M., Ovsyukova M., Akopyan A., Dubrovina N., Amstislavskaya T.G., Khlestkina E. Evaluating the effects of grain of isogenic wheat lines differing in the content of anthocyanins in mouse models of neurodegenerative disorders. *Nutrients*. 2020;12:3877. doi 10.3390/nu12123877
- Tikhonova M.A., Shoeva O.Y., Tenditnik M.V., Akopyan A.A., Litvinova E.A., Popova N.A., Amstislavskaya T.G., Khlestkina E.K. Antitumor effects of an anthocyanin-rich grain diet in a mouse model of lewis lung carcinoma. *Int J Mol Sci*. 2024;25:5727. doi 10.3390/ijms25115727
- Vasilova N.Z., Askhadullin D.F., Askhadullin D., Bagavieva E.Z., Tazutdinova M.R., Khusainova I.I. Violet-green variety of spring soft wheat Nadira. *Zernobobovye i Krupyanye Kultury = Legumes and Groat Crops*. 2021;4(40):66-75. doi 10.24412/2309-348X-2021-4-66-75 (in Russian)
- Vinichenko N.A., Salina E.A. Gliadins: structure and toxic epitopes. *Pisma v Vavilovskii Zhurnal Genetiki i Seleksii = Lett Vavilov J Genet Breed*. 2021;7(4):163-171. doi 10.18699/LettersVJ2021-7-21 (in Russian)
- Zakharova I.I., Tolstova E.G. Features of quality control in the production of spelt dishes. *Kontsept = The Concept*. 2019;4:238-246. doi 10.24411/2304-120X-2019-14028 (in Russian)
- Zeven A.C. Wheats with purple and blue grains: a review. *Euphytica*. 1991;56(3):243-258. doi 10.1007/bf00042371
- Zverev S.V., Pankratieva I.A., Politukha O.V., Chirkova L.V., Vitol I.S., Starichenkov A.A. Study of emmer properties. *Khleboprodukty = Grain Products*. 2016;1:66-67 (in Russian)

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

Received October 20, 2025. Revised February 20, 2026. Accepted February 20, 2026.

doi 10.18699/vjgb-26-66

Genetic characterization and meta-analysis of the population of the Northern Black Sea region in the 1st millennium CE based on ancient DNA data

E.D. Aituganova , A.S. Kon'kov ¹, T.V. Andreeva ^{1, 2, 3}, E.V. Rozhdestvenskikh ¹,
A.D. Manakhov ^{1, 2}, E.I. Rogaev ^{1, 4} 

¹ Sirius University of Science and Technology, Sirius Federal Territory, Krasnodar region, Russia

² Vavilov Institute of General Genetics of the Russian Academy of Sciences, Moscow, Russia

³ Centre of Genetics and Genetic Technologies, Department of Genetics, Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia

⁴ Department of Psychiatry, UMass Chan Medical School, Shrewsbury, MA, USA

 ajtuganova.ed@talantiuspeh.ru; evivrecc@gmail.com

Abstract. Over the past two decades, the introduction of whole-genome sequencing analysis of ancient DNA has led to a breakthrough in archaeogenetic research, significantly expanding our understanding of human genetic history. In this context the Northern Black Sea region during first millennium CE (1–1,000 CE) is of particular relevance, as it remained a hub of intense cultural exchange and migration. Despite its historical importance, ancient genomic data from this period remains scarce, and a comprehensive synthesis of existing findings is lacking. This study presents a systematic review and meta-analysis of published whole-genome sequencing data from 48 ancient samples associated with key archaeological cultures of the region: Late Scythian, Sarmatian, Alan, Bulgar, Saltovo-Mayaki and Chernyakhov. Through the systematization of data, we trace genetic continuity at Late Scythian and Alanian sites relative to preceding populations. Episodes of large-scale migration and population replacement have been documented, most clearly evident in the Sarmatian expansion of the 1st–4th centuries CE, with genetic traces extending from the Urals to the Carpathians. Based on limited evidence, genetic continuity has been identified between representatives of the Chernyakhov culture and early Slavic groups. Through our meta-analysis, we further detect intercultural connections between the Alans, Bulgars, and bearers of the Saltovo-Mayaki culture, whose genetic structure reveals the influence of Caucasian and East Eurasian components. Collectively, these findings underscore the complex genetic landscape of the region, shaped by successive migration waves and multifaceted intercultural contacts. We conclude by outlining key unresolved questions and future directions for archaeogenetic research in the Northern Black Sea region during the first millennium CE.

Key words: archaeogenetics; ancient DNA; genome; Northern Black Sea region; Sarmatians; Chernyakhov culture; Alans; Bulgars; Saltovo-Mayaki culture

For citation: Aituganova E.D., Kon'kov A.S., Andreeva T.V., Rozhdestvenskikh E.V., Manakhov A.D., Rogaev E.I. Genetic characterization and meta-analysis of the population of the Northern Black Sea region in the 1st millennium CE based on ancient DNA data. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2026;30(4):658-668. doi 10.18699/vjgb-26-66

Funding. This work was supported by the grant of the state program of the Sirius Federal Territory "Scientific and technological development of the 'Sirius' Federal Territory", Agreement No. 18-03 dated September 10, 2024, project code GEN-BFT-2407.

Генетическая характеристика и метаанализ населения Северного Причерноморья I тысячелетия н. э. по данным древней ДНК

Э.Д. Айтуганова , А.С. Коньков ¹, Т.В. Андреева ^{1, 2, 3}, Е.В. Рождественских ¹,
А.Д. Манахов ^{1, 2}, Е.И. Рогаев ^{1, 4} 

¹ Научно-технологический университет «Сириус», федеральная территория Сириус, Краснодарский край, Россия

² Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия

³ Центр генетики и генетических технологий, кафедра генетики биологического факультета МГУ им. М.В. Ломоносова, Москва, Россия

⁴ Медицинская школа Чан Массачусетского университета, департамент психиатрии, Шрусбери, США

 ajtuganova.ed@talantiuspeh.ru; evivrecc@gmail.com

Аннотация. За последние два десятилетия внедрение полногеномного анализа древней ДНК привело к прорыву в археогенетических исследованиях, существенно расширив наше понимание генетической истории человечества. На этом фоне особую актуальность приобретает изучение Северного Причерноморья I тысячелетия нашей эры (1–1000 гг.), региона, который в течение всего периода оставался зоной интенсивного культурного обмена и миграций. Несмотря на его историческую значимость, объем доступных данных древней ДНК по этому периоду пока остается ограниченным. В настоящей работе систематизированы и проанализированы опубликованные результаты полногеномного секвенирования 48 древних образцов, относящихся к различным археологическим культурам региона: позднескифской, сарматской, аланской, болгарской, салтово-маяцкой и черняховской. Систематизация данных и применение методов популяционно-генетического анализа позволили проследить генетическую преемственность на памятниках позднескифской, а также отдельно – аланской культуры по отношению к предшествующему населению. Удалось обнаружить межкультурные связи алан, болгар и носителей салтово-маяцкой культуры, в генетической структуре которых отмечается влияние кавказского и восточноевропейского компонентов. С другой стороны, зафиксированы эпизоды масштабных миграций и полной смены населения, наиболее ярко выраженные в сарматской экспансии I–IV вв. н.э., следы которой генетически прослеживаются от Урала до Карпат. По единичным образцам выявлена генетическая преемственность между представителями черняховской культуры и ранними славянскими группами. Полученные результаты демонстрируют мозаичность генетического ландшафта региона, формировавшегося под воздействием многоволновых миграций и межкультурных контактов. Статья освещает современное состояние исследований, показывает основные достижения и намечает перспективы дальнейшего изучения в области археогенетики Северного Причерноморья в I тысячелетии н.э.

Ключевые слова: археогенетика; древняя ДНК; геном; Северное Причерноморье; сарматы; черняховская культура; аланы; болгары; салтово-маяцкая культура

Introduction

For centuries, the Northern Black Sea region, as part of the Pontic-Caspian steppe, has been characterised by high population mobility, which has shaped its complex demographic and ethno-cultural history.

During this period, the population of the Northern Black Sea coast consisted of Hellenised groups who, in the 7th–5th centuries BC, founded numerous poleis: Olbia, Chersonesos, Panticapaeum, Phanagoria and others (Koshelenko et al., 1984). In the first centuries CE, these centers maintained a stable antique cultural zone, sustaining close contacts both with the metropolitan centers and with the local population – Scythians, Taurians, Maeotians, Sindi, and Getae (Alekseeva, 1991). Throughout the first millennium they fell within the sphere of influence of Rome, Byzantium, and later the Khazar Khaganate (Pletneva, 2003; Afanasyev et al., 2015a).

At the same time, the region's steppe lands became a corridor for successive waves of migration by nomadic tribal confederations. By the beginning of the first millennium CE, groups of the Sarmatian culture began to succeed the European Scythians, establishing dominance in the region and extending their influence from the Don to the Danube (Mordvintseva, 2013; Koryakova, 2018; Kovács, 2023). Later, in the 1st–3rd centuries, the Alans came to dominate the Azov region and the Ciscaucasia (Perevalov, 2014). From the late 2nd to the early 3rd century, Germanic tribes expanded into the region from northwestern Europe. These groups left a distinct imprint on the cultural and political structure of the Northern Black Sea region's population, contributing to the formation of the Chernyakhov culture (3rd–5th centuries) (Kazanski, 2011; Matveev, 2017; Onishchuk, 2018).

A major turning point in the demographic history of the region came with the Hunnic invasion in the 4th–5th centuries, which marked the beginning of the Migration Period and was accompanied by large-scale population movements (Shu-

shunova, Yartsev, 2024). In the 5th–7th centuries, the Huns were succeeded by nomadic confederations of the Onogurs, Kutrigurs, Savirs, and later the Avars (Bubenok, 2014). In the mid-7th century, Great Bulgaria emerged in the region, soon to be replaced by the Khazar Khaganate (Pletneva, 2003; Afanasyev et al., 2015a; Kazanski, 2020). From the late 7th to the 8th centuries, the Khazar Khaganate became the dominant force in the region, controlling the steppes from the Don to the Caucasus, as well as the eastern Crimea and the Taman Peninsula. Within these territories, under Khazar control, the Saltovo-Mayaki culture also developed (Pletneva, 2003; Berezina et al., 2012; Reshetova, 2012; Afanasyev et al., 2015a). Concurrently, in the more northern forest-steppe areas of the region, the settlement of Slavic tribes began (Sedov, 1979).

The complexity and multi-layered nature of the historical and cultural processes in the Northern Black Sea region have long attracted the attention of researchers. In recent years, the reconstruction of the region's population history has increasingly been based on the analysis of ancient DNA using a range of population genetics methods. In autosomal analysis, the principal component method is used to visualise the structure of genetic diversity (Patterson et al., 2006; McVean, 2009), while the ADMIXTURE method is employed to identify likely ancestral components (Alexander et al., 2009). Population origin models based on *f*-statistic methods are used, which measure the covariance of allele frequencies, thereby allowing the estimation of genetic distances, the identification of mixed ancestry, and the direction of gene flow (Patterson et al., 2012).

Based on these principles, the *qpAdm* and *qpWave* tools have been developed, which are widely used to test demographic hypotheses and quantify admixture proportions (Haak et al., 2015; Harney et al., 2021). IBD analysis identifies regions of the genome inherited from common ancestors, enabling

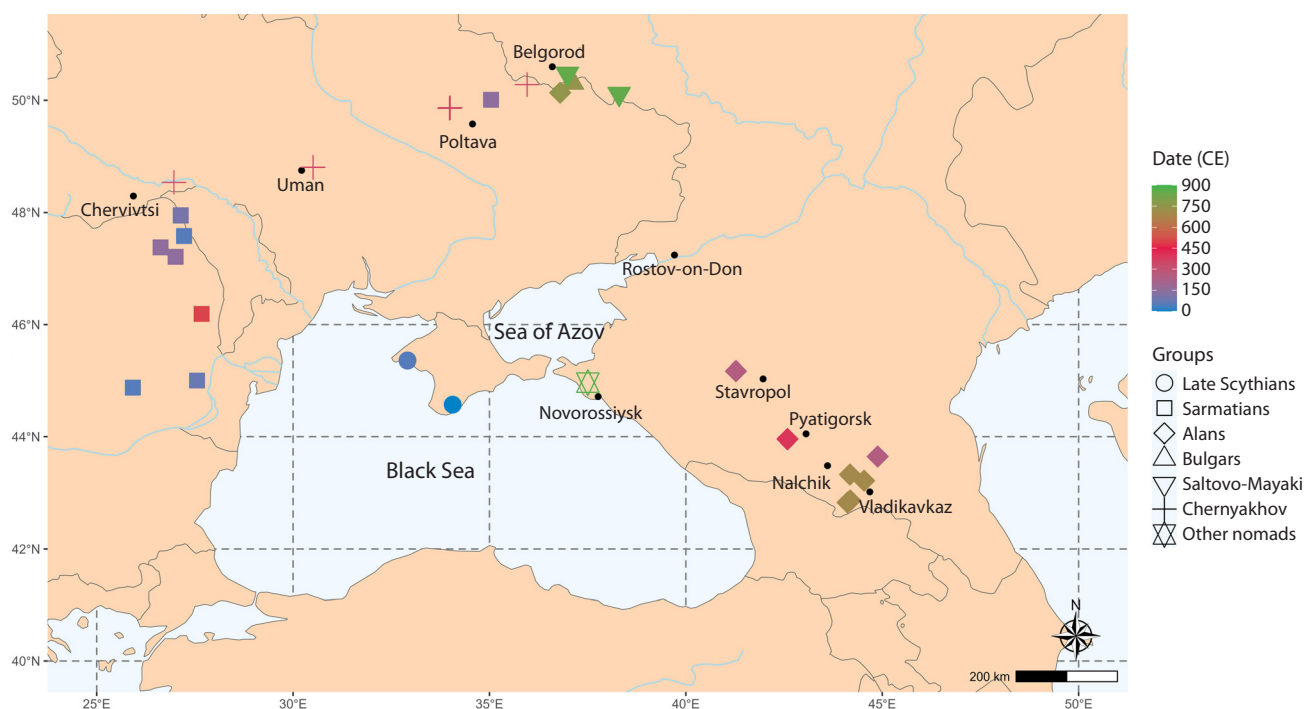


Fig. 1. Geographic location of the archaeological sites from which 48 ancient individuals used in this study.

Samples from archaeological sites are indicated by symbols corresponding to their respective archaeological culture. The color scale indicates sample dating, based on archaeological chronology or radiocarbon analysis (150-year intervals, spanning 0 to 900 CE). Scale bar: 200 km.

the assessment of genealogical relationships between individuals (Ringbauer et al., 2024). Analysis of mitochondrial DNA (mtDNA) and Y-chromosome haplogroups allows for a detailed reconstruction and investigation of maternal and paternal lineages.

Despite the rapid growth in archaeogenetic research, data from the Northern Black Sea region remain fragmentary and unevenly represented both chronologically and culturally. Consequently, the systematization, comparison, and meta-analysis of published data become particularly critical. This approach allows for the identification of patterns in genetic dynamics, assessment of the robustness of proposed models of origin and migration, and outlining the limitations of the current state of research as well as future perspectives.

In this study, we conducted a systematic review and meta-analysis of published whole-genome ancient DNA data from the Northern Black Sea region dating to the 1st millennium CE (Fig. 1).

Materials and methods

Sources were selected based on their peer-reviewed status and the availability of reliable data regarding the archaeological context, geographic provenance, and dating of the samples. We used archaeogenetic descriptions and genomic data from articles and/or AADR v62.0 database (Mallick et al., 2024), followed by downloading genomes from the ENA resource (ENA Browser, 2025).

Principal component analysis (PCA) was performed using the smartpca function of the EIGENSOFT package (Pat-

erson et al., 2006). The genotypes of the studied samples were projected onto the genetic variation space of more than 1,300 individuals from 80 present-day European and Caucasian populations from the Human Origin panel, comprising approximately 600 thousand SNPs (Lazaridis et al., 2016). To estimate the proportions of ancestral origins in the genome structure, we used the ADMIXTURE v1.3 software package with the same Human Origin panel (Alexander et al., 2009).

All transitional variants that could represent postmortem modifications, as well as loci with a high proportion of missing data ($-geno\ 0.999$), were removed from the dataset. To ensure marker independence, linked SNPs were excluded from the analysis using the PLINK command `--indep-pairwise 50/10/0.1`. To estimate the optimal number of clusters K , 12 independent calculations were performed in a cross-validation mode, and the final value of K was taken based on the minimum cross-validation error. Prior to analysis, samples with fewer than 10 thousand markers from the Human Origin panel were excluded.

Results

In total, we analyzed 48 published ancient genomes from this region, which were grouped based on origin, geography, time period, and archaeological culture (Fig. 1 and see the Table).

Systematic review of published data

This section presents a systematic review of published archaeogenetic studies, grouped by archaeological culture.

List of published whole-genome data for ancient individuals from the Northern Black Sea region

Ancient DNA study	Number of ancient human genomes	Culture	Dating	Geographic location
Andreeva et al., 2025	2	Late Scythian period	1st century BC–2nd century CE	Coastal zone
	1	Alans	2nd–4th centuries CE	Ciscaucasus
Saag et al., 2025	1	Sarmatian period	1st–3rd centuries CE	Forest-steppe zone
	11	Chernyakhov	2nd–5th centuries CE	
	6	Alans	7th–9th centuries CE	
	3	Bulgars	7th–9th centuries CE	
Schütz et al., 2025	9	Sarmatian period	1st–6th centuries CE	Lower Danube region
Sharko et al., 2024*	1	Alans	3rd century CE	Ciscaucasus
Maróti et al., 2022	2	Other nomads	8th–10th centuries CE	
Järve et al., 2019	3	Chernyakhov	3rd–6th centuries CE	Forest-steppe zone
Damgaard et al., 2018	6	Alans	4th–9th centuries CE	Ciscaucasus
	3	Saltovo-Mayaki	7th–10th centuries CE	Forest-steppe zone

* No meta-analysis was conducted for the individual described in this article due to the poor quality of the genomic data.

Late Scythian population. Whole-genome data on the Late Scythian population are presented in only one published study (Andreeva et al., 2025), which includes two individuals from the 1st century BC to the 2nd century CE from the most representative sites of the Late Scythian period – Neapolis Scythian and the Belyaus settlement. Archaeological data from Belyaus indicate a transformation of the site from a Greek settlement into a center of Late Scythian culture, likely accompanied by a change in the population (Dashevskaya, 2014).

The results of principal component analysis (PCA) reveal differences in the genetic profiles of individuals from these sites. One of them (AS10) is similar to other Scythian samples from the Crimea, dating from both the Late Scythian period (1st century BC–1st century CE) and the earlier period (6th–3rd centuries BC), indicating a probable continuity of the local population (Andreeva et al., 2025). According to the results of *qpAdm* modelling, his genetic profile, unlike that of other Scythian groups, includes a high proportion (around 40 %) of the Anatolian component, suggesting a connection between the Crimean Scythians and the ancient population of Anatolia. The second individual (AS1) is clustered in the PCA with the classical Scythians of the Northern Black Sea region from the 4th–3rd centuries BC, and his genetic profile contains a significant proportion of an ancient Siberian component (Andreeva et al., 2025).

Collectively, these archaeogenetic data indicate pronounced genetic heterogeneity of the Late Scythian population of the Northern Black Sea region (Andreeva et al., 2025). However, the limited sample size and the isolated nature of research into this period do not allow for generalisations at the level of archaeological culture. It is noteworthy that this study represents the first attempt to predict phenotypic traits for the Late Scythian population of this period using markers from

the HIRISplex-S system (Andreeva et al., 2025). The results indicate a probable brown (chestnut) hair color, intermediate skin tone and brown eyes (Andreeva et al., 2025). Although genomic coverage allowed phenotypic characteristics to be determined for only one individual, these results provide a basis for a subsequent description of the physical appearance of the entire group.

Sarmatian period. In the forest-steppe zone of the Northern Donets, a single individual from the Sarmatian period (1st–3rd centuries CE) was excavated from the “Lyubovka” burial ground. The burial was located within the cultural layer of an earlier Scythian settlement, but was attributed to the Sarmatian period on the basis of the artefacts found and dated to the 1st–3rd centuries. According to the results of PCA and ADMIXTURE, this individual proved to be similar to the population of the Caucasus 7th–9th centuries CE, as were the later groups of Alans and Steppe Bulgars of the Saltovo-Mayaki culture discussed in this article (Saag et al., 2025).

Interesting findings have emerged from the analysis of genomes from individuals dating to the Sarmatian period (1st–4th centuries CE) from the Lower Danube region (Romania) (Schütz et al., 2025). Analysis using *f4*, PCA and ADMIXTURE revealed their marked genetic similarity to the steppe Sarmatians of the Urals and Central Asia, and *qpAdm* modeling results also indicated a high proportion of the steppe component. It is noteworthy that the Sarmatians from the territory of Romania and the Carpathian Basin demonstrate an extremely low proportion of common ancestry with the preceding Scythian-era population from the same territory.

Analysis of chromosomal segments of shared origin (IBD) confirms this picture: extensive shared chromosomal regions were identified among carriers of the Sarmatian culture in the European and Asian parts of the Black Sea–Caspian steppe.

In particular, shared IBD segments of up to 64 cM in length were identified between individuals from the Carpathian Basin (POG-10, male, 1st century CE) and from the Rostov Oblast (DA139, female, 8th century BCE–1st century CE). Moreover, individual DA139 shared 88 cM of IBD with a Sarmatian from the Orenburg Oblast (chy001, female, 1st–3rd centuries CE) (Schütz et al., 2025). These data indicate recent common ancestry among these individuals and active migration processes between the eastern and western regions of the Sarmatian world in the early centuries of the 1st millennium CE.

Alans. The Alan population of the Northern Black Sea region studied to date is represented by two main groups, differing both geographically and culturally-archaeologically (see the Table). The first group includes eight Alans from the Ciscaucasus – the population of the mountainous and foothill areas of the North Caucasus. The second group is represented by six Alans from the forest-steppe zone, who inhabited the interfluvium of the Seversky Donets and Don rivers and are archaeologically attributed to the Saltovo-Mayaki culture (7th–9th centuries CE).

The earliest archaeogenetic data were obtained from the Ciscaucasia (North Ossetia – Alania) using a sample of six individuals (Damgaard et al., 2018). The authors discussed the origins of the Alans and their presumed Iranian-speaking affiliation. The results revealed no significant Sarmatian contribution to the gene pool of the 5th-century Alans of the Ciscaucasus. Statistical analysis indicated that the North Caucasian Alans form a clade with the Iron Age population from Armenia, demonstrating significant genetic differences between Sarmatians and Alans. The authors suggest that the Alans likely inherited the Iranian language from the steppe Sarmatians, but by the 5th–6th centuries CE, the genetic structure of their population was already virtually indistinguishable from the local Caucasian population.

In the study where the genome of one Alan individual from the kurgan burial ground “Bratskie I Kurgans” (the Chechen Republic) was sequenced (Sharko et al., 2024), different conclusions were drawn. According to PCA results, this individual shows genetic similarity to Alans from the Ciscaucasus studied in an earlier work (Damgaard et al., 2018). ADMIXTURE analysis and f_3 and f_4 statistics results showed that his genome is structurally close to populations of the Koban culture; however, unlike the highland Caucasian groups of the Bronze Age (Maykop and Kura-Araxes cultures), it exhibits the presence of a steppe component. In this study, signs of contamination were identified in some samples, therefore the obtained data require caution in interpretation and subsequent confirmation using other samples and datasets.

The genetic profile of one Alan individual from the site “Solnechnodolsk-4” (Stavropol region) (radiocarbon age 132–329 cal CE) includes steppe components of Bronze Age populations; according to PCA and ADMIXTURE results, it was found to be similar to the profiles of representatives of the Sarmatian and Koban populations from the same site (Andreeva et al., 2025), which may also indicate population continuity in this region during the transition of archaeological cultures.

Pronounced genetic heterogeneity was documented in the study of a sample of six Alans, bearers of the Saltovo-Mayaki culture, from the forest-steppe zone of the Northern Donets region at the archaeological complex “Verkhniy Saltov” (7th–9th centuries CE) (Saag et al., 2025). Four individuals from this site demonstrate maximum genetic similarity to populations of the North Caucasus, consistent with the studies described above. Two other individuals exhibit the greatest genetic similarity to populations of Central Asia (Saag et al., 2025).

Taken together, the available data point to a complex and heterogeneous genetic structure within the Alanian population, comprising steppe, North Caucasian and Central Asian genetic components. The inconsistency of some of the results highlights the need to expand the sample sizes and conduct further research on Alanian groups.

Bulgars. Currently, the Bulgar population of the Black Sea region in the 1st millennium CE is represented in genetic studies by only three individuals from the site “Bochkov” in the Seversky Donets basin (Saag et al., 2025). In the genetic profile of these Bulgars, admixture from populations of the Caucasus and Central Asia was identified. The high degree of genetic similarity between the Alans and Bulgars, who are bearers of the same Saltovo-Mayaki culture but exhibit differences in burial rites (catacomb and pit burials, respectively), leads to the conclusion that both groups may have originated from a single genetically homogeneous group. This conclusion aligns with archaeological interpretations, according to which both catacomb and pit burials of the Saltovo-Mayaki culture in the forest-steppe zone are associated with migrations of Alans from the Caucasus (Berezina et al., 2012; Afanasyev et al., 2015b; Korobov, 2019; Malashev, Krivosheev, 2024).

According to *qpAdm* proximal modeling results, the Sarmatian, Alan, and Bulgar groups from the forest-steppe zone that demonstrate genetic connections to the Caucasus can be modeled by the contribution of a preceding group of elite Scythian nomads from the Seversky Donets basin (Saag et al., 2025).

Three samples of representatives of the Saltovo-Mayaki culture, associated with one of the population groups of the Khazar Khaganate, were studied in an earlier work (Damgaard et al., 2018). These individuals, although heterogeneous, also cluster in the principal component space with Caucasian Alans and other samples from the North Caucasus, and according to a later study, they were found to be related to the forest-steppe zone Alans based on f_4 test results (Saag et al., 2025).

Chernyakhov culture. Archaeogenetic studies of the Chernyakhov culture population cover samples from the forest-steppe zone of the region, dating to the 3rd–5th centuries CE. In 2018, the genomes of three Chernyakhov individuals were presented (Järve et al., 2019), and later the sample was supplemented with an additional 11 specimens (Saag et al., 2025).

Results of principal component analysis and *qpAdm* indicate the identification of at least three genetic groups among representatives of the Chernyakhov culture. One of them includes individuals with a predominance of Eastern and Central Euro-

pean components, the second consists of bearers of a genetic profile similar to continental Southern European populations (close to Thracians) (Järve et al., 2019; Saag et al., 2025).

f4 tests show that the first Eastern/Central European group forms clades with preceding Illyrian-Thracian populations on the right bank of the Dnieper. *qpAdm* results indicate that this group can be modeled based on ancestral groups of the Yamnaya culture, the Trypillia culture, and a minor proportion of East Asian populations. The other two groups of Chernyakhov individuals demonstrate a larger proportion of Trypillia genetic heritage (up to 70 %). One representative of the Chernyakhov culture has a pronounced Southern component and clusters in PCA with modern Cypriots (Saag et al., 2025).

Data were also indicate probable connections between bearers of the Chernyakhov culture and early Slavic groups. According to *qpAdm* modeling data, medieval groups of the Golden Horde period, including presumed Slavs and nomads from the Seversky Donets basin, are formed 100 % from the genetic pool of Group 1 of the Chernyakhov culture, suggesting direct continuity between the Late Antique population of the Chernyakhov culture and the early Slavs (Saag et al., 2025). Data were also obtained indicating genetic continuity through the maternal line between a representative of the Chernyakhov culture from the 3rd century CE (Odessa region) and an individual of Slavic origin from the medieval burial ground “Minino II” from the 12th–13th centuries CE (Vologda region). The individuals were found to have a complete match of the mitochondrial sequence (Rozhdestvenskikh et al., 2025).

Genomic data from individual representatives of Northern Black Sea cultures. The genetic profiles of two individuals from the 8th–9th centuries CE from burials of nomads of unclear cultural attribution at the site “Andreevskaya Shchel” (Anapa) contain, according to *qpAdm* modeling, a dominant contribution from the population of Armenia of the Middle Bronze Age (76–96 %) (Maróti et al., 2022).

Within the scope of this review, we have systematised archaeogenetic data from individual studies and small samples. Nevertheless, even with limited material, genetic connections are clearly evident both within individual cultures (as in the case of the Sarmatians) and between cultures—among the Alans, Bulgars, and bearers of the Saltovo-Mayaki culture.

Meta-analysis

To obtain a comprehensive picture of the population genetic structure of the Northern Black Sea region in the 1st millennium CE, we conducted a meta-analysis of the previously published and above-mentioned whole-genome data using PCA and ADMIXTURE methods. The results of the analysis are presented in Fig. 2.

In the projection of the first two principal components, formed on the basis of modern populations of Western Eurasia and the Caucasus, the earliest samples of the 1st millennium CE (Late Scythian and Sarmatian individuals) occupy an intermediate position between European and Caucasian groups. The Sarmatians form a distinct genetic cluster, consistent with published archaeogenetic data (Saag et al., 2025; Schütz et al., 2025). The only exceptions are two samples: one clusters with the Alans, Bulgars, and bearers of the Saltovo-Mayaki culture,

while the other, the latest from this sample (5th–6th centuries CE), clusters with modern Southern European populations.

Representatives of the Chernyakhov culture do not form a single cluster, but are predominantly located near modern Central European and Balkan populations, with one Chernyakhov individual clustering with the Late Scythian and Sarmatian samples (Fig. 2a).

The Alans and Bulgars – representatives of the Saltovo-Mayaki culture – demonstrate pronounced intragroup heterogeneity. A significant portion of these samples cluster within Caucasian populations. Three samples – the Alan from the Ciscaucasus DA161, along with the Alan UKR135 and Bulgar UKR147 from the forest-steppe zone – are located near Central Asian populations. Another three samples – the Alans UKR134, UKR144, and the Bulgar UKR137 from the forest-steppe zone – are notably shifted on the PCA plot towards Eastern populations. It should be noted that the earliest studied Alan from the Ciscaucasus (AB105; 2nd–4th centuries CE) falls within the cluster of Late Scythian and Sarmatian samples, consistent with the hypothesis that the Alans emerged from Sarmatian groups. However, all later representatives of the Alan population show greater similarity to Caucasian populations rather than to steppe Sarmatian groups.

Similar results were obtained from ADMIXTURE analysis (Fig. 2b). The genetic profiles of representatives from each studied culture (Late Scythians/Sarmatians, Alans, Bulgars, and Chernyakhov) exhibit clearly distinguishable differences. In the genetic profiles of the earliest steppe groups (Scythians and Sarmatians), as well as in the early Alans, a small proportion of the Western Hunter-Gatherer component is present which is completely absent in later representatives of the Alans and other bearers of the Saltovo-Mayaki culture. However, this component is present in significant proportions in the Chernyakhov individuals, which testifies to different origins for all these groups inhabiting the Northern Black Sea region during the 1st millennium CE and the absence of direct continuity between them. The genetic profile of the Alans and Bulgars shows contributions from Caucasus Hunter-Gatherers and/or ancient Iranian Neolithic farmers, indicating a connection of these groups with the ancient population of the Caucasus. Several Bulgar and Alan individuals possess a pronounced Asian component, explaining their shift on the PCA plot towards Eastern populations and suggesting the involvement of migrants from Eastern Eurasia in the formation of these groups and certain representatives of the Northern Black Sea population.

The bearers of the Chernyakhov culture are characterized by significant diversity in genetic profiles, with a common predominance of the Anatolian Neolithic component combined with Western and Eastern Hunter-Gatherer components, which generally reflects their predominantly European genetic profile.

Mitochondrial DNA and Y-chromosome analysis. Additionally, to identify potential migration routes of the ancient population of the Northern Black Sea region, we compiled all available data on mitochondrial DNA and Y-chromosomes for the analysis of maternal and paternal lineages. A total of

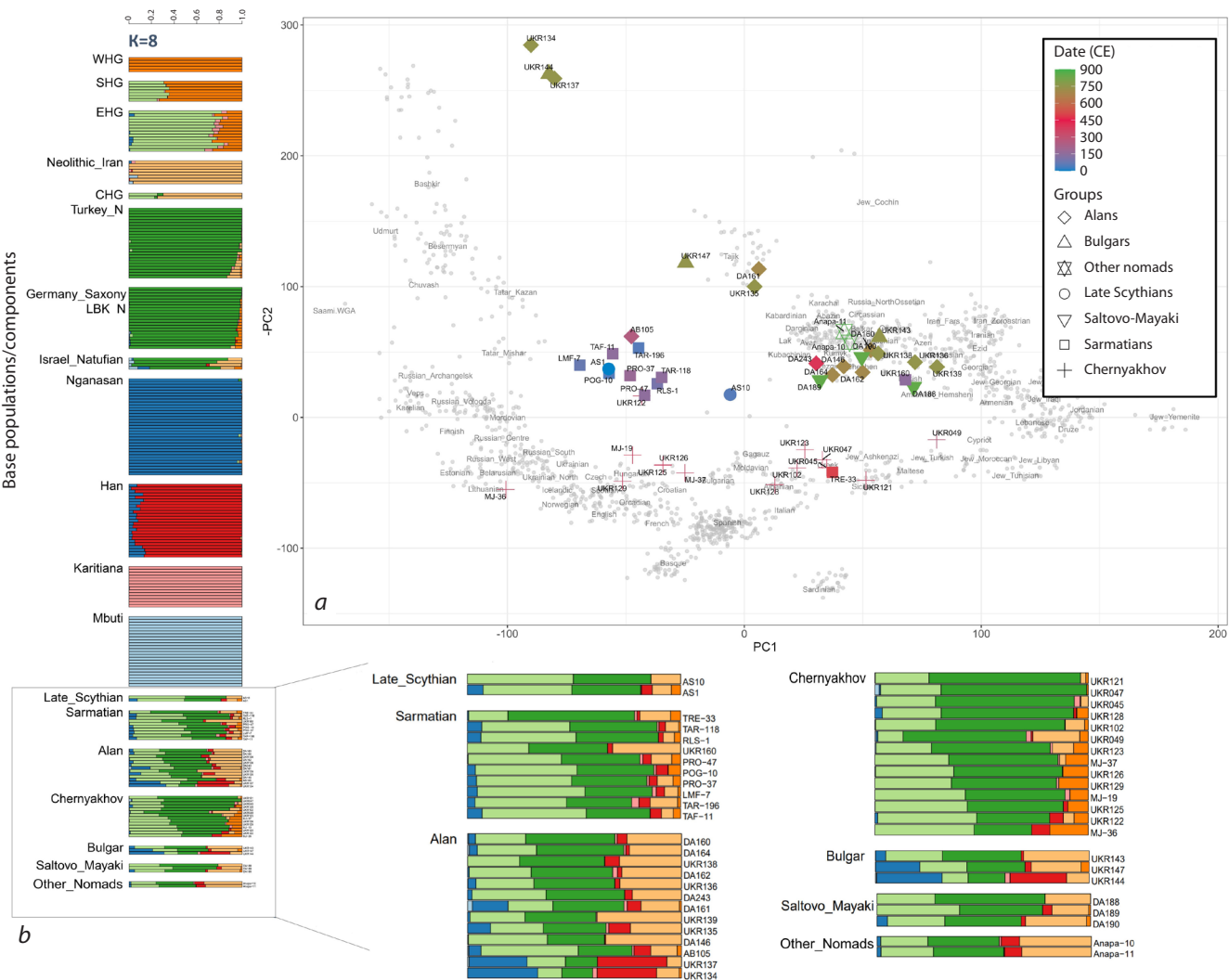


Fig. 2. Results of the meta-analysis for individuals from the Northern Black Sea region.

a – principal component analysis (PCA); *b* – population structure analysis (ADMIXTURE) with *K* = 8 inferred ancestral components.

Component (population) designations: WHG – Western Hunter-Gatherer (9,107–5,986 BC); SHG – Scandinavian Hunter-Gatherer (5,967–5,484 BC); EHG – Eastern Hunter-Gatherer (9,649–5,535 BC); Neolithic_Iran – individuals from Neolithic Iran (8,295–7,976 BC); CHG – Caucasus Hunter-Gatherer (11,961–7,599 BC); Turkey_N – individuals from Neolithic Anatolia (9,656–5,835 BC); Germany_Saxony_LBK_N – individuals of the Linear Pottery culture (5,400–4,900 BC); Israel_Natufian – individuals of the Natufian culture (12,000–9,500 BC); Nganasan – present-day Nganasan individuals (Siberia); Han – present-day Han individuals (China); Karitiana – present-day Karitiana individuals (Brazil); Mbuti – present-day Mbuti individuals (Central Africa).

324 individuals from the Northern Black Sea region across different chronological periods were included in the analyses (Fig. 3, see Supplementary Material)¹.

Western Eurasian mitochondrial lineages predominate among the Late Scythians (Juras et al., 2017); among the Sarmatians, these are supplemented by the Eastern haplogroup F, potentially indicating a Central Asian maternal contribution (Comas et al., 2004). There is significant genetic diversity among the Alans and Bulgars, including the presence of East Eurasian mitochondrial haplogroups (Comas et al., 2004; Derenko et al., 2010, 2012): subclades A and F among the Alans and B and D among the Bulgars. This confirms the

contribution of Central and East Asian components to their genetic profile. Among representatives of the Chernyakhov culture, haplogroups characteristic of West Eurasian populations dominate (H, K, T2, HV, U4).

It should be noted that, unlike data on autosomal markers and mitochondrial haplogroups, the number of available Y-chromosome data for analysis from individuals of the Northern Black Sea region and adjacent regions remains limited.

From the Bronze Age onwards, a sharp increase in the proportion of haplogroups R1a and R1b has been observed in the Black Sea region, which is thought to correlate with Indo-European expansion and the formation of the main branches of Indo-European languages (Allentoft et al., 2015; Haak et al.,

¹ Supplementary Material is available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Aitug_Engl_30_4.pdf

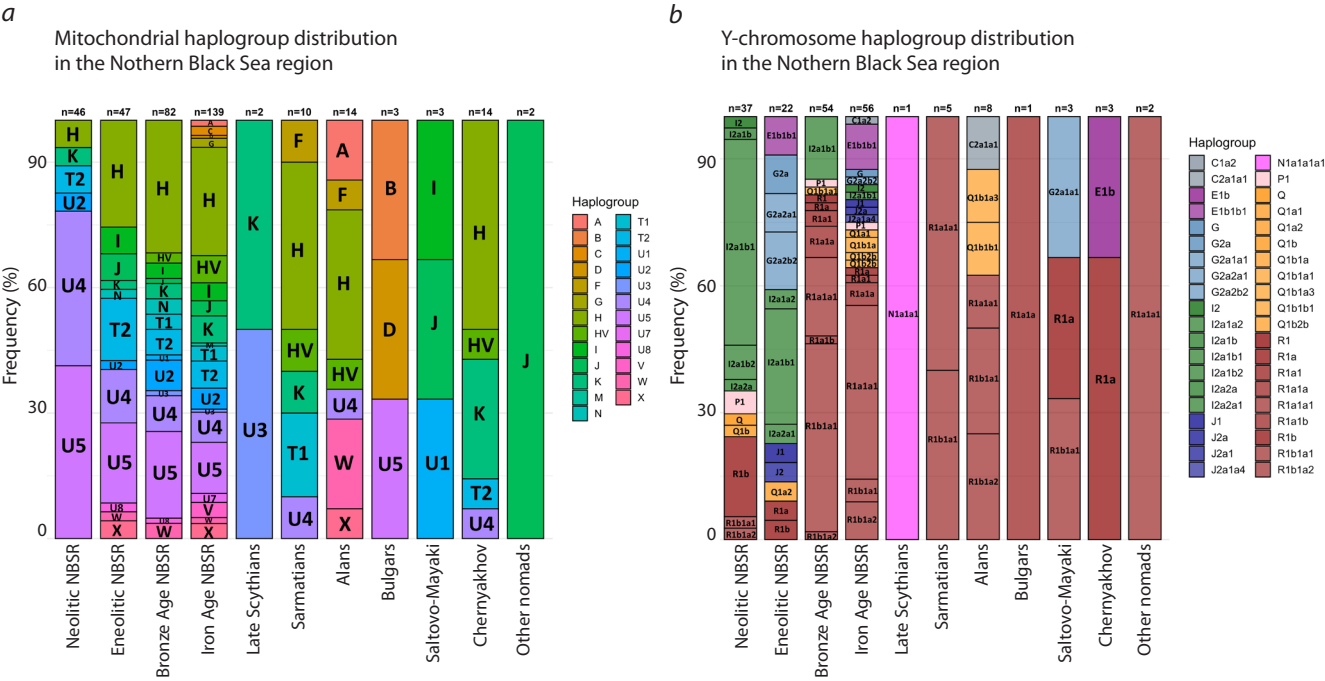


Fig. 3. Distribution of mitochondrial (a) and Y-chromosomal haplogroups (b) in the Northern Black Sea region.
*The figure shows abbreviated haplogroup names for clarity.

2015). Of particular interest is the Late Scythian sample from Crimea with the Y-chromosome haplogroup N1a1a1a2a (N-Z1934) – a lineage absent in earlier complexes of the Northern Black Sea region and widely distributed among contemporary Finno-Ugric populations (Post et al., 2019). This Y-chromosome lineage is currently found only among Russians from the central and Volga-Perm regions (YTree, 2025).

The Alans are characterized by a more heterogeneous Y-chromosome composition. Alongside the haplotypes R1b-Z2106, R1b-L23, R1b-M269, and R1b-Z2105, previously identified among Scythian groups of the Early Iron Age in the Northern Black Sea region, they also possess haplotypes Q1b-BZ5214, found in modern Chechens, and C2a-Y11606, found in modern populations of Mongolia and China, as well as in ancient elite groups of the Xiongnu and Avars (YTree, 2025) (see Supplementary Material).

The Bulgars and the people of the Saltovo-Mayaki culture are represented by haplogroup R1a-M417, which had previously been recorded in the Northern Black Sea region. In one individual from the Saltovo-Mayaki culture, a complete match of subclade R1b-M269 was identified with an Alan sample and with several individuals from the Bronze Age of Moldova and the North Caucasus, as well as with Early and Late Scythians. In another individual, lineage G2a-Z6653 was found, matching a Late Antique sample from Georgia (YTree, 2025).

The Chernyakhov culture is represented by haplogroups R1a-Z93, R1a-Z645, and E1b-V13, previously identified among Scythian groups of the Northern Black Sea region, populations of the Late Bronze Age of Moldova, as well as among Thracian and, probably, certain Classical Greek groups (Saag et al., 2025; YTree, 2025).

Discussion

Despite its key role in trans-Eurasian migration processes during the first millennium CE, the Northern Black Sea region remains, from an archaeogenetic perspective, still poorly and fragmentarily characterized to date. While for certain groups, such as the Sarmatians, the Alans of the mountainous and forest-steppe zones, the Bulgars, and to some extent the population of the Chernyakhov culture, it is possible to provide a preliminary description of their genetic characteristics, for a number of archaeological cultures that inhabited this territory, no data are currently available.

The small sample sizes and limited number of whole-genome samples covering various chronological periods and archaeological cultures significantly impede the reconstruction of demographic processes and reduce the reliability of overarching conclusions. Nevertheless, the systematic review and meta-analysis of the available data allowed us to identify the main features of the genetic structure of the population of the Northern Black Sea region during the 1st millennium CE.

In several cases, the transition of archaeological culture was not accompanied by complete population replacement, indicating local genetic continuity. This is clearly demonstrated by data from the site “Solnechnodolsk-4” (Ciscaucasus), where an Alan individual was genetically close to Sarmatians and bearers of the Koban culture from the same site (Andreeva et al., 2025). Similarly, a Late Scythian individual from the Belyaus settlement exhibits a genetic profile similar to the earlier Scythian population of the same region (Andreeva et al., 2025).

At the same time, instances of long-distance migration have been documented: genetic links have been identified between the Sarmatians from the Carpathian Basin and the Southern

Urals, providing direct evidence of large-scale migrations across the Northern Black Sea region in the first centuries CE (Schütz et al., 2025).

Our meta-analysis revealed genetic connections between certain representatives of the Sarmatians, Alans, Bulgars, and bearers of the Saltovo-Mayaki culture. The groups from the forest-steppe zone and the Caucasian Alans demonstrate the greatest similarity to each other, which may indicate a common origin or prolonged contacts. At the same time, within the structure of the studied groups, we were able to document the main directions of genetic influences: Caucasian and East Eurasian, which is consistent with our results from the analysis of uniparental markers (mitochondrial DNA and Y-chromosome).

The genetic diversity identified among individuals of the Chernyakhov culture is complemented by data from the sites “Shishaki” and “Minino II”, where genetic connections with later Slavic groups are observed (Rozhdestvenskikh et al., 2025; Saag et al., 2025). These results provide a foundation for further, more in-depth genetic analysis of the processes underlying the formation of Slavic groups in Eastern Europe.

Finally, it was demonstrated that the Sarmatians who arrived in the Black Sea region were not genetically close to the local Scythians (Andreeva et al., 2025). This suggests that the Scythian population was largely replaced by the Sarmatians without genetic admixture.

It should be noted that, as a result of our analysis using previously published genomic data, we did not identify traces of admixture between the various ethnic and cultural groups inhabiting the Northern Black Sea region during the 1st millennium CE.

Conclusion

The systematic review and meta-analysis of archaeogenetic data on the population of the Northern Black Sea region during the 1st millennium CE allow us to view the region not as a sequence of successive cultures, but rather as a long-functioning contact zone. The obtained data have created a solid foundation for subsequent expansion of sample sets and deepening of research. A promising direction is the acquisition of whole-genome data for the key populations of the Northern Black Sea region for which genomic data are still absent, including the population of the Bosporan Kingdom, the nomadic confederations of the early Middle Ages (Goths, Huns, etc.), early Slavic groups, as well as the populations of Bulgaria and the Khazar Khaganate.

Investigating the connections between the genetic characteristics of the population of this region, the transmission of cultural traditions, and the spread of languages is particularly important for clarifying its demographic history in subsequent historical periods (the Middle Ages). This requires a comprehensive interdisciplinary approach involving geneticists, archaeologists, anthropologists, and linguists.

References

- Afanasyev G.E., Ven' Sh., Tun S., Van L., Veil L., Dobrovolskaya M.V., Korobov D.S., Reshetova I.K., Li Kh. Khazar Confederates in the Don region. In: Natural Scientific Methods and the Paradigm of Modern Archaeology. Moscow: IA RAN, 2015a;146-154 (in Russian)
- Afanasyev G.E., Dobrovolskaya M.V., Korobov D.S., Reshetova I.K. New archaeological, anthropological and genetic aspects in the study of the Alans from the Don region. *Kratkiye Soobshcheniya Instituta Arkheologii = Brief Communications of the Institute of Archaeology*. 2015b;(237):64-79 (in Russian)
- Alekseeva E.M. Greek Colonization of the Northwestern Caucasus. Moscow, 1991 (in Russian)
- Alexander D.H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655-1664. doi 10.1101/gr.094052.109
- Allentoft M.E., Sikora M., Sjögren K.-G., Rasmussen S., Rasmussen M., Stenderup J., Damgaard P.B., ... Sicheritz-Pontén T., Brunak S., Nielsen R., Kristiansen K., Willerslev E. Population genomics of Bronze Age Eurasia. *Nature*. 2015;522(7555):167-172. doi 10.1038/nature14507
- Allentoft M.E., Sikora M., Refoyo-Martínez A., Irving-Pease E.K., Fischer A., Barrie W., Ingason A., ... Delaneau O., Werge T., Racimo F., Kristiansen K., Willerslev E. Population genomics of post-glacial western Eurasia. *Nature*. 2024;625(7994):301-311. doi 10.1038/s41586-023-06865-0
- Andreeva T.V., Soshkina A.D., Gusev F.E., Malyarchuk A.B., Dotsenko G.S., Dudko N.A., Plotnikova M.Yu., ... Kantorovich A.R., Maslov V.E., Barinov D.G., Dobrovolskaya M.V., Rogaev E.I. Genetic history of Scythia. *Sci Adv*. 2025;11(30):eads8179. doi 10.1126/sciadv.ads8179
- Berezina N.Y., Buzhilova A.P., Reshetova I.K. New craniological materials revisited anthropological background of medieval Alans. *Vestnik MGU. Seriya XXIII. Antropologiya = Lomonosov J Anthropol*. 2012;(4):18-36 (in Russian)
- Bubenok O.B. To the Question about the Quantity of Tribes of the Nomads of North Black Sea Region in the late 5–first half of 7 century AD. *Drinovs'kij Zbirnik*. Kharkiv; Sofiya. 2014, V. IX;29-35
- Comas F., Plaza S., Wells R.S., Yuldaseva N., Lao O., Calafell F., Bertranpetit J. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet*. 2004;12(6):495-504. doi 10.1038/sj.ejhg.5201160
- Damgaard P.B., Marchi N., Rasmussen S., Peyrot M., Renaud G., Korneliussen T., Moreno-Mayar J.V., ... Nielsen R., Sikora M., Heyer E., Kristiansen K., Willerslev E. 137 ancient human genomes from across the Eurasian steppes. *Nature*. 2018;557(7705):369-374. doi 10.1038/s41586-018-0094-2
- Dashevskaya O.D. The Belyaus Necropolis. Simferopol, 2014 (in Russian)
- Derenko M., Malyarchuk B., Grzybowski T., Denisova G., Rogalla U., Perkova M., Dambueva I., Zakharov I. Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PLoS One*. 2010;5(12):e15214. doi 10.1371/journal.pone.0015214
- Derenko M., Malyarchuk B., Denisova G., Perkova M., Rogalla U., Grzybowski T., Khusnutdinova E., Dambueva I., Zakharov I. Complete mitochondrial DNA analysis of eastern Eurasian haplogroups rarely found in populations of Northern Asia and Eastern Europe. *PLoS One*. 2012;7(2):e32179. doi 10.1371/journal.pone.0032179
- ENA Browser, 2025. Available online: <https://www.ebi.ac.uk/ena/browser/home>
- Gelabert P., Schmidt R.W., Fernandes D.M., Karsten J.K., Harper T.K., Madden G.D., Ledogar S.H., Sokhatsky M., Oota H., Kennett D.J., Pinhasi R. Genomes from Verteba cave suggest diversity within the Trypillians in Ukraine. *Sci Rep*. 2022;12(1):7242. doi 10.1038/s41598-022-11117-8
- Haak W., Lazaridis I., Patterson N., Rohland N., Mallick S., Llamas B., Brandt G., ... Brown D., Anthony D., Cooper A., Alt K.W., Reich D. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 2015;522(7555):207-211. doi 10.1038/nature14317
- Harney É., Patterson N., Reich D., Wakeley J. Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics*. 2021;217(4):iyaa045. doi 10.1093/genetics/iyaa045

- Järve M., Saag L., Scheib C.L., Pathak A.K., Montinaro F., Pagani L., Flores R., ... Metspalu M., Savelev N., Kriiska A., Kivisild T., Villems R. Shifts in the genetic landscape of the Western Eurasian Steppe associated with the beginning and end of the Scythian dominance. *Curr Biol.* 2019;29(14):2430-2441.e10. doi 10.1016/j.cub.2019.06.019
- Jones E.R., Zarina G., Moiseyev V., Lightfoot E., Nigst P.R., Manica A., Pinhasi R., Bradley D.G. The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr Biol.* 2017;27(4):576-582. doi 10.1016/j.cub.2016.12.060
- Juras A., Krzewińska M., Nikitin A.G., Ehler E., Chyleński M., Łukas S., Krenz-Niedbala M., Sinika V., Piontek J., Ivanova S., Dabert M., Götherström A. Diverse origin of mitochondrial lineages in Iron Age Black Sea Scythians. *Sci Rep.* 2017;7(1):43950. doi 10.1038/srep43950
- Kazanski M.M. Radagaïs and the End of the Chernyakhov Culture (Radagaïs et la fin de la civilisation de Černjahov). *OIUM.* Kyiv; Luts'k. Vol. 1. 2011;1:22-32 (in Russian)
- Kazanski M.M. the antiquities of the steppe nomads of the Post-Hunnic period (mid-fifth to mid-sixth century in Eastern Europe. *Materialy po Arkheologii, Istorii i Etnografii Tavrii = Materials in Archaeology, History and Ethnography of Tauria.* 2020;25:90-167 (in Russian)
- Korobov D.S. The Alans of the North Caucasus: Ethnos, Archaeology, Paleogenetics. Saint Petersburg, 2019 (in Russian)
- Koryakova L. Europe to Asia. In: The Oxford Handbook of the European Iron Age. Oxford University Press, 2018;575-618. doi 10.1093/oxfordhb/9780199696826.013.29
- Koshelenko G.A., Kruglikova I.T., Dolgorukov V.S. Ancient States of the Northern Black Sea Region. Moscow, 1984 (in Russian)
- Kovács P. Fontes Sarmatarum in Hungaria Habitantium – A Magyarországi Szarmatákra Vonatkozó Antik Források. Budapest: Magyar-ságtudat Intézet, 2023
- Krzewińska M., Kılınc G.M., Juras A., Koptekin D., Chyleński M., Nikitin A.G., Shcherbakov N., ... Dalén L., Sinika V., Jakobsson M., Storå J., Götherström A. Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. *Sci Adv.* 2018;4(10):eaat4457. doi 10.1126/sciadv.aat4457
- Lazaridis I., Nadel L., Rollefson G., Merrett D.C., Rohland N., Mallick S., Fernandes D., ... Yengo L., Hovhannysyan N.A., Patterson N., Pinhasi R., Reich D. Genomic insights into the origin of farming in the ancient Near East. *Nature.* 2016;536(7617):419-424. doi 10.1038/nature19310
- Lazaridis I., Alpaslan-Roodenberg S., Acar A., Açıkkol A., Agelarakis A., Aghikyan L., Akyüz U., ... Çavuşoğlu R., Rohland N., Pinhasi R., Reich D., Davtyan R. The genetic history of the Southern Arc: a bridge between West Asia and Europe. *Science.* 2022;377(6609):eabm4247. doi 10.1126/science.abm4247
- Malashev V.Yu., Krivosheev M.V. Revisiting the issue of attribution of some catacomb monuments groups from the south of Eastern Europe in the late 3rd–early 5th centuries. *Nizhnevolzhskii Arheologicheskii Vestnik = Lower Volga Archaeological Bulletin.* 2024; 23(3):52-78. doi 10.15688/nav.jvolsu.2024.3.4 (in Russian)
- Mallick S., Micco A., Mah M., Ringbauer H., Lazaridis I., Olalde I., Patterson N., Reich D. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci Data.* 2024; 11(1):182. doi 10.1038/s41597-024-03031-7
- Maróti Z., Neparáczki E., Schütz O., Maár K., Varga G.I.B., Kovács B., Kalmár T., ... Gulyás B., Kovácsóczy B.N., Gál S.S., Tomka P., Török T. The genetic origin of Huns, Avars, and conquering Hungarians. *Curr Biol.* 2022;32(13):2858-2870.e7. doi 10.1016/j.cub.2022.04.093
- Mathieson I., Alpaslan-Roodenberg S., Posth C., Szécsényi-Nagy A., Rohland N., Mallick S., Olalde I., ... Borić D., Bonsall C., Krause J., Pinhasi R., Reich D. The genomic history of southeastern Europe. *Nature.* 2018;555(7695):197-203. doi 10.1038/nature25778
- Mattila T.M., Svensson E.M., Juras A., Günther T., Kashuba N., Ala-Hulkko T., Chyleński M., ... Netea M.G., Nikitin A.G., Persson P., Malmström H., Jakobsson M. Genetic continuity, isolation, and gene flow in Stone Age Central and Eastern Europe. *Commun Biol.* 2023; 6(1):793. doi 10.1038/s42003-023-05131-3
- Matveev S.V. Barbarized Romans or Romanized Barbarians in the Chernyakhov Community of the Prut-Dniester Interfluvium. In: Eastern Europe in Antiquity and the Middle Ages: Ancient and Medieval Communities. Moscow, 2017;147-151 (in Russian)
- McVean G. A genealogical interpretation of principal components analysis. *PLoS Genet.* 2009;5(10):e1000686. doi 10.1371/journal.pgen.1000686
- Mordvintseva V. The Sarmatians: the creation of archaeological evidence. *Oxford J. Archaeol.* 2013;32(2):203-219. doi 10.1111/oja.12010
- Nikitin A.G., Lazaridis I., Patterson N., Ivanova S., Videiko M., Der-gachev V., Kotova N., ... Workman J.N., Zalzal F., Mallick S., Rohland N., Reich D. A genomic history of the North Pontic Region from the Neolithic to the Bronze Age. *Nature.* 2025;639(8053):124-131. doi 10.1038/s41586-024-08372-2
- Onishchuk Ja.I. Relationship of the populations of the Velbar, Sarmatian, and Chernyakhov cultures in the territory of Western Podolia in the second quarter of the 1st millennium AD. *Zhurnal Belorusskogo Gosudarstvennogo Universiteta. Istorija = Journal of the Belarusian State University. History.* 2018;3:120-129 (in Russian)
- Patterson N., Price A.L., Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2(12):e190. doi 10.1371/journal.pgen.0020190
- Patterson N., Moorjani P., Luo Y., Mallick S., Rohland N., Zhan Y., Genschoreck T., Webster T., Reich D. Ancient admixture in human history. *Genetics.* 2012;192(3):1065-1093. doi 10.1534/genetics.112.145037
- Patterson N., Isakov M., Booth T., Büster L., Fischer C.E., Olalde I., Ringbauer H., ... Lillie M., Rohland N., Pinhasi R., Armit I., Reich D. Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature.* 2021;601(7894):588-594. doi 10.1038/s41586-021-04287-4
- Penske S., Rohrlach A.B., Childebayeva A., Gneccchi-Ruscione G., Schmid C., Spyrou M.A., Neumann G.U., ... Ringbauer H., Stockhammer P.W., Hansen S., Krause J., Haak W. Early contact between late farming and pastoralist societies in southeastern Europe. *Nature.* 2023;620(7973):358-365. doi 10.1038/s41586-023-06334-8
- Perevalov S.M. Alans: a Mirage of Nomadic Empire. *Vestnik Vladikavkazskogo Nauchnogo Tsentra = Vestnik of Vladikavkaz Scientific Centre.* 2014;14(2):2-9 (in Russian)
- Pletneva S.A. Nomads of the South Russian Steppes in the Middle Ages (4th–13th Centuries). Voronezh, 2003 (in Russian)
- Post H., Németh E., Klima L., Flores R., Fehér T., Türk A., Székely G., ... Khusnutdinova E.K., Metspalu E., Villems R., Tambets K., Rootsi S. Y-chromosomal connection between Hungarians and geographically distant populations of the Ural Mountain region and West Siberia. *Sci Rep.* 2019;9(1):7786. doi 10.1038/s41598-019-44272-6
- Posth C., Yu H., Ghalichi A., Rougier H., Crevecoeur I., Huang Y., Ringbauer H., ... Terberger T., Caramelli D., Altina E., Haak W., Krause J. Palaeogenomics of Upper Palaeolithic to Neolithic European hunter-gatherers. *Nature.* 2023;615(7950):117-126. doi 10.1038/s41586-023-05726-0
- Reshetova I.K. New anthropological materials of the Saltovo-Mayaki culture from the Verkhny Saltov-IV burial ground. *Rossiyskaya Archeologiya = Russian Archaeology.* 2012;(3):129-136 (in Russian)
- Ringbauer H., Huang Y., Akbari A., Mallick S., Olalde I., Patterson N., Reich D. Accurate detection of identity-by-descent segments in human ancient DNA. *Nat Genet.* 2024;56(1):143-151. doi 10.1038/s41588-023-01582-w
- Rozhdestvenskikh E.V., Andreeva T.V., Malyarchuk A.B., Adriano-va I.Yu., Khodyreva D.S., Evteev A.A., Buzhilova A.P., Rogayev E.I. Mitogenomic analysis of a representative of the Chernyakhov culture in the Middle Dniester and their genetic relationship with the Slavs in the context of paleoanthropological data. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed.* 2025;29(5):722-731. doi 10.18699/vjgb-25-79

- Saag L., Utevska O., Zadnikov S., Shramko I., Gorbenko K., Bandrivskyi M., Pavliv D., ... Barrington C., Gilardet A., Macleod R., Skoglund P., Thomas M.G. North Pontic crossroads: mobility in Ukraine from the Bronze Age to the early modern period. *Sci Adv.* 2025;11(2):eadr0695. doi 10.1126/sciadv.adr0695
- Schütz O., Maróti Z., Tihanyi B., Kiss A.P., Nyerki E., Gînguță A., Kiss P., ... Soficaru A.D., Spekker O., Varga S., Neparáczi E., Török T. Unveiling the origins and genetic makeup of the “forgotten people”: a study of the Sarmatian-period population in the Carpathian Basin. *Cell.* 2025;188(15):4074-4090.e11. doi 10.1016/j.cell.2025.05.009
- Sedov V.V. The Origin and Early History of the Slavs. Moscow, 1979 (in Russian)
- Sharko F.S., Boulygina E.S., Tsygankova S.V., Slobodova N.V., Rastorguev S.M., Krasivskaya A.A., Belinsky A.B., ... Shvedchikova T.Yu., Dobrovolskaya M.V., Reshetova I.K., Korobov D.S., Nedoluzhko A.V. Koban culture genome-wide and archeological data open the bridge between Bronze and Iron Ages in the North Caucasus. *Eur J Hum Genet.* 2024;32(11):1483-1491. doi 10.1038/s41431-023-01524-4
- Shushunova E.V., Yartsev S.V. The Ancestor Cult of the Ancient Population of the Northern Black Sea Region in the 1st Century BC – 4th Century AD and Features of its Further Development in the Way of Transformation from Paganism to Christianity. Simferopol, 2024 (in Russian)
- Wang C.C., Reinhold S., Kalmykov A., Wissgott A., Brandt G., Jeong C., Cheronet O., ... Belinskij A.B., Reich D., Hansen S., Krause J., Haak W. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat Commun.* 2019;10(1):590. doi 10.1038/s41467-018-08220-8
- YTree, 2025. Available online: <https://www.yfull.com/tree/>

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

Received December 23, 2025. Revised December 29, 2025. Accepted December 29, 2025.

doi 10.18699/vjgb-26-67

Selection and evaluation of DNase I hypersensitive sites for prenatal screening of trisomy 21 in the fetus

A.M. Mazur ¹, A.S. Starshin ¹, N.V. Bogush², E.B. Prokhortchouk ^{1,3}¹ Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia² N.I. Pirogov Russian National Research Medical University, Moscow, Russia³ Sirius University of Science and Technology, Sirius Federal Territory, Krasnodar region, Russia mazur.am@gmail.com

Abstract. This study introduces a novel approach for noninvasive prenatal testing (NIPT) of chromosomal abnormalities, based on analysis of epigenetic features in circulating cell-free DNA (cfDNA). The core innovation of our method leverages fundamental differences in chromatin organization between maternal and fetal cells. Specifically, we focused on genomic regions that exhibit open chromatin configuration in maternal blood cells but remain tightly packed in fetal tissues (DNase I hypersensitive sites or DHSs). These epigenetic differences create distinct cfDNA fragmentation signatures that allow selective identification of fetal DNA within the maternal cfDNA pool. The study workflow comprised several key steps: performing genome-wide screening to identify differentially accessible chromatin regions, selecting the most informative markers using a machine learning algorithm, and targeted sequencing of the selected epigenetic markers using molecular barcodes. Subsequently, a LASSO regression model was constructed and validated. As a proof of concept, the study demonstrates the method's efficacy in identifying trisomy 21 (Down syndrome), though the underlying principles can be readily adapted to other abnormalities. Complementing its robust performance, the technique offers practical advantages in terms of platform compatibility – the same epigenetic markers can be assessed using either next-generation sequencing or simpler, more cost-efficient methods like digital PCR. With further refinement, the approach could be extended to screen for additional aneuploidies (trisomies 13 and 18) and microdeletion syndromes. Therefore, this approach offers new opportunities for developing cost-effective testing systems suitable for widespread routine clinical implementation, combining high diagnostic accuracy with reduced analysis costs.

Key words: chromatin; cell free DNA; aneuploidy; trisomy; NIPT

For citation: Mazur A.M., Starshin A.S., Bogush N.V., Prokhortchouk E.B. Selection and evaluation of DNase I hypersensitive sites for prenatal screening of trisomy 21 in the fetus. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed.* 2026;30(4): 669-675. doi 10.18699/vjgb-26-67

Funding. This work was supported by a grant from the state program of the "Sirius" Federal Territory "Scientific and Technological Development of the "Sirius" Federal Territory" (Agreement No. 21-03 dated September 27, 2024) (for the section "Whole-genome sequencing data analysis algorithm") and the Ministry of Science and Higher Education of the Russian Federation.

Отбор и анализ DNase I-гиперчувствительных сайтов для пренатального определения трисомии 21 у плода

A.M. Мазур ¹, А.С. Старшин ¹, Н.В. Богущ², Е.Б. Прохорчук ^{1,3}¹ Институт биоинженерии, Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук, Москва, Россия² Российский национальный исследовательский медицинский университет им. Н.И. Пирогова Министерства здравоохранения Российской Федерации (Пироговский Университет), Москва, Россия³ Научно-технологический университет «Сириус», федеральная территория «Сириус», Краснодарский край, Россия mazur.am@gmail.com

Аннотация. В данной работе представлен новый подход к неинвазивной пренатальной диагностике хромосомных аномалий, основанный на анализе эпигенетических особенностей циркулирующей внеклеточной ДНК (вкДНК). Ключевая идея метода заключается в использовании фундаментальных различий в организации хроматина между материнскими и фетальными клетками. В частности, мы сосредоточились на участках генома, которые в материнских клетках крови находятся в открытом состоянии (DHS), но остаются закрытыми в тканях плода. Эти эпигенетические различия создают уникальные паттерны вкДНК, позволяющие выделить фетальную ДНК из общего пула материнской вкДНК. Работа включает в себя несколько основных шагов: проведение полногеномного скрининга для выявления

дифференциально доступных участков хроматина, отбор наиболее информативных маркеров при помощи алгоритма машинного обучения, таргетное секвенирование отобранных эпигенетических маркеров с применением молекулярных баркодов и построение модели на основе LASSO-регрессии с последующей ее валидацией. Особое внимание уделено диагностике трисомии хромосомы 21 (синдрома Дауна), как наиболее клинически значимого хромосомного нарушения. Разработанный алгоритм демонстрирует высокую точность при анализе отобранных эпигенетических маркеров. Важное преимущество метода – его технологическая гибкость: отобранные маркеры можно анализировать как с помощью NGS, так и более доступными методами, включая цифровую ПЦР. Перспективы дальнейшего развития метода включают расширение диагностической панели для выявления других частых анеуплоидий (трисомии хромосом 13, 18), а также микроделеционных синдромов. Предлагаемый подход открывает новые возможности для создания экономически эффективных тест-систем, которые могут быть внедрены в широкую клиническую практику, обеспечивая высокую точность диагностики при снижении стоимости анализа.

Ключевые слова: хроматин; внеклеточная ДНК; анеуплоидия; трисомия; НИПТ

Introduction

The Non-Invasive Prenatal Test (NIPT), which employs next-generation sequencing (NGS) technology, represents a revolutionary approach to detecting fetal chromosomal abnormalities using only a sample of maternal blood (Fan et al., 2008). Prior to the development of NGS, this type of analysis was not feasible, as there was no reliable method to isolate the small amount of fetal DNA present amidst the vast excess of maternal DNA in the bloodstream (Dhallan et al., 2007). Due to the high throughput and sensitivity of NGS, it is now possible to accurately detect and quantify these trace amounts of fetal cell-free DNA (cfDNA) directly from maternal plasma (Alberry et al., 2021). Consequently, NIPT has emerged as a rapid, safe, and reliable screening tool for assessing the risk of common aneuploidies, including Down, Edwards, and Patau syndromes (Gil et al., 2017). The genomics revolution driven by NGS has therefore paved the way for the widespread clinical implementation of NIPT (Satam et al., 2023).

However, despite the impressive capabilities of NGS, the effectiveness of NIPT still depends on selecting optimal biological markers to distinguish between maternal and fetal DNA (Hui, Bianchi, 2020). In this context, epigenetic differences – particularly chromatin accessibility profiles – are of special interest.

DNase I-hypersensitive sites (DHSs) are genomic regions with an open chromatin structure, exhibiting increased accessibility to the DNase I enzyme. These sites are frequently associated with gene regulatory elements and display cell-specific expression profiles (Thurman et al., 2012).

In the previous stage of this study, the use of DHS clusters significantly restricted the number of available candidate regions, thereby reducing the sensitivity and specificity of the test system. To overcome this limitation, we transitioned from cluster analysis to the investigation of individual DHS. This approach substantially expanded the pool of candidate regions, enabling more precise filtering and ultimately yielding a sufficient number of high-quality loci for reliable diagnostics.

Our study focuses on developing an enhanced NIPT approach based on the analysis of individual DHSs – genomic regions with chromatin structures characteristic of the fetus

and mother. The proposed method is expected to ensure high accuracy in fetal DNA detection by leveraging the epigenetic features of DHSs, while also reducing assay costs through their targeted analysis.

Material and methods

Study cohort. A total of 1,012 blood samples from pregnant women were collected between April 2014 and April 2015. The patients were referred from 47 medical institutions across the Russian Federation or were self-referred. The mean age of participants was 35 years, with the youngest being 20 and the oldest 51 years. Gestational age at the time of blood collection ranged from 10 to 24 weeks, with a mean of 14 weeks. The mean fetal cell-free DNA (cfDNA) concentration was 11 % (Pantiuhk et al., 2016). All women provided informed consent for the use of their samples for scientific purposes. The study was approved by the Committee of the Academic Council of the Faculty of Biomedical Sciences, Pirogov Russian National Research Medical University, Ministry of Health of the Russian Federation (protocol No. 4 dated December 26, 2022).

Primer design. Primers were designed using the Primer3 software (Untergasser et al., 2012) with the following parameters: PRIMER_OPT_SIZE = 21, PRIMER_MIN_SIZE = 18, PRIMER_MAX_SIZE = 24, PRIMER_MAX_NS_ACCEPTED = 1, and PRIMER_PRODUCT_SIZE_RANGE = 100–140 bp.

Target enrichment and library preparation. Targeted sequencing was performed using single-end enrichment with duplex unique molecular identifiers (UMIs), following the protocol described by Q. Peng et al. (2019).

Library sequencing. Sequencing was carried out using Illumina SBS technology with single-end reads of at least 80 bp.

Model development and evaluation. In the first step, the input data were prepared by dividing samples into two categories: those with trisomy and those without it. Samples with a total molecule count below 15,000 were filtered out.

Subsequently, 100 independent trials were conducted using the R packages rsample and caret (Kuhn, 2008). In each trial, the data were randomly split into training (60 % of samples),

validation (20 %), and test (20 %) sets, while preserving the original class ratio.

To compensate for class imbalance in the training sets, observation weighting was applied. Class weights were calculated separately for each training set according to the formulas:

$w_t = N_{\text{major}} / N_{\text{minor}}$ – for samples of the minority class (trisomy),

$w_n = 1$ – for samples of the majority class (normal),

where N_{major} and N_{minor} are the number of samples in the majority and minority classes, respectively, within the training set. This weighting effectively increases the contribution of rare trisomy cases to the loss function, preventing the model from biasing toward the more common class.

The optimal classification threshold was determined on the validation set using Youden's index (Youden, 1950), which maximizes the J statistic ($J = \text{sensitivity} + \text{specificity} - 1$). This criterion aims to find a threshold that balances the method's ability to correctly identify both positive cases (sensitivity) and negative cases (specificity). This approach ensures that the selected threshold balances the two types of errors: false negatives (where aneuploidy is present but the test fails to detect it) and false positives (where the test incorrectly indicates aneuploidy). Three key performance metrics were calculated on the test data: the area under the ROC curve (AUC), sensitivity, and specificity. Model training was performed using the glmnet package (Friedman et al., 2010), and all results were visualized using ggplot2 (Wickham, 2016) and ggpubr.

Results

Identification of candidate regions in the human genome

At the first stage, DNase I-hypersensitive sites (DHSs) were selected from an open-access database (Sheffield et al., 2013). For maternal tissue representation, all hematopoietic and endothelial cell lines were used, as these are typical sources of freely circulating maternal cfDNA (Lui et al., 2002).

The selection criterion was based on differential accessibility of these sites: an open chromatin state in cell lines corresponding to maternal tissues (31 lines of hematopoietic and endothelial origin) and a closed state in fetal cell lines (Chorion). This approach enabled the identification of 13,637 individual DHSs (Fig. S1)¹.

The coordinates of all selected sites were converted to the human genome version GRCh38 (Table S1). For subsequent analysis, a 1,000 bp region was extracted around each 150 bp DHS, with the DHS positioned at its center. Although the DHSs themselves did not overlap, the flanking regions could partially overlap when the distance between central sites was less than 500 bp – a factor that was taken into account during final selection.

¹ Supplementary Figures S1–S3 and Tables S1–S3 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Mazur_Engl_30_4.xlsx

To construct the diagnostic panel, we selected a set of regions containing DHSs on: chromosome 21 – as the direct target for Down syndrome diagnosis; chromosome Y – as a negative control; and chromosomes 1, 2, 3, 4, 5, 6, 8, 10, and 12 – as markers, the relative coverage decrease of which in trisomy cases reflects the redistribution of reads in favor of chromosome 21.

Sample set formation

The study utilized cfDNA sequencing data from a previous work focused on non-invasive prenatal diagnostic methods for aneuploidies (Pantiukh et al., 2016).

To determine fetal cfDNA concentration, two complementary methods were applied: quantitative analysis of the Y chromosome and the SeqFF method, which is based on regression analysis of overrepresented fetal sequences (Kim et al., 2015; van Beek et al., 2017).

The need to group samples arose from the characteristics of the original data, where the mean coverage was only 0.3×. At this coverage level, a standard 150 bp DHS received on average only one read, which made reliable quantitative analysis at the individual level impossible. To address this low coverage issue, samples were sorted by increasing fetal DNA concentration and evenly distributed into four groups (P1–P4) (Fig. S2).

The resulting clinical sample set included data from 958 cases with normal fetal karyotype and 54 samples with trisomy 21 (Down syndrome).

Whole-genome sequencing data analysis algorithm

All selected cfDNA samples were aligned to the human reference genome GRCh38 using bowtie2 (Langmead, Salzberg, 2012) in global alignment mode (--end-to-end) with the --sensitive parameter. The mean percentage of successfully mapped reads was 97.05 %, with unmapped reads excluded from further analysis. For each sample, coverage was calculated for every nucleotide within the pre-selected regions.

Aggregated data were divided into groups: non-pregnant women, pregnant women with normal fetal karyotype (further subdivided by fetal DNA concentration), and pregnant women with fetal trisomies. A statistically significant increase in coverage within DHSs was observed between groups of pregnant women with a normal fetal karyotype (Fig. 1a) and between pregnant women compared to non-pregnant women – the latter explained by the closed chromatin state of fetal DNA in these genomic regions, rendering it inaccessible to DNase I (Fig. 1b). Differences were also noted for pregnant women carrying fetuses with trisomy 21 (Fig. 1b). This increase in DHS coverage confirmed that the proposed method enables the detection of fetal aneuploidies.

Interestingly, multiple regions with increased coverage were identified in pregnant women. Verification using the UCSC Genome Browser confirmed the presence of additional DHSs

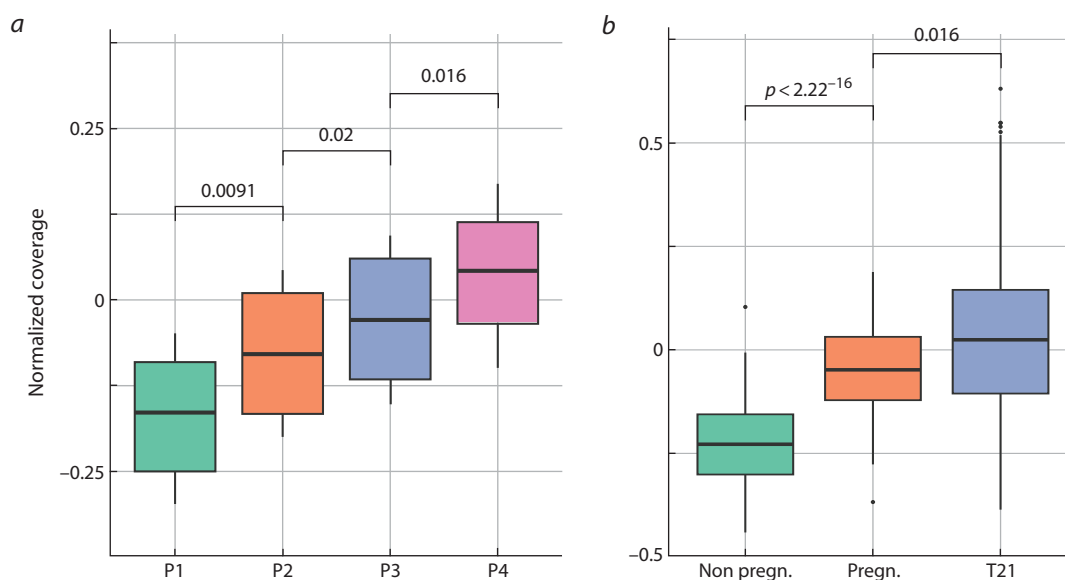


Fig. 1. Normalized coverage (Z-score) in target regions using chromosome 21 as an example: *a* – in subgroups of pregnant women with normal fetal karyotype; *b* – in groups of non-pregnant women, pregnant women with normal fetal karyotype, and pregnant women with fetal trisomy.

Statistical significance of differences between the groups was assessed using the Mann–Whitney U-test (Wilcoxon test for independent samples).

within such regions (Fig. S3), which is consistent with known data on the clustering of regulatory elements in the genome (Thurman et al., 2012).

Establishing criteria and selecting regions

The primary objective was to identify regions most effective at distinguishing between pregnant and non-pregnant women. To this end, the following criteria were proposed: the length of the maximum window where the difference between mean coverage in non-pregnant women and mean coverage in pregnant women is less than zero; the maximum area under the difference curve within a 100-nucleotide window; the relationship between coverage level and fetal cfDNA concentration; and functional significance. Functional significance was assessed based on criteria proposed in R.E. Thurman et al. (2012). Promoter sites were defined as those the coordinates of which intersected with transcription start sites from the GENCODE human genome reference annotation (Harrow et al., 2012), or as the DHS nearest to a transcription start site in the 5' direction. Enhancers were defined as those located within 500,000 nucleotides of a promoter site and showing high correlation ($R \geq 0.7$) with that promoter site based on DNase-seq data.

To determine which of the proposed criteria was most suitable for region selection, a binary classifier based on random forest was constructed to distinguish samples from pregnant women versus non-pregnant women, using the mean coverage in the 20 top-ranking regions according to each criterion. We used the Random Forest Classifier algorithm from the scikit-

learn library in its default configuration (Pedregosa et al., 2011). Prediction quality was assessed using the F-measure for non-pregnant sample predictions, as the number of such samples was smaller than that of pregnant samples. To identify the most effective region selection rule, we evaluated the predictive performance of each criterion individually and in various combinations. The best F-measure was achieved using regions selected by a combination of three criteria: the area under the difference curve criterion, the presence of a relationship between coverage level and fetal DNA concentration within the region, and the criterion of whether the site was a promoter or enhancer.

Next, for each nucleotide coordinate, the mean normalized coverage (Z-score) was calculated across all classes: for non-pregnant samples; for pregnant samples (by groups with different cfDNA concentrations); and for samples from pregnant women carrying fetuses with trisomy 21. Subsequently, all regions on a given chromosome were sorted in descending order according to the value determined by the selected criterion, and a specified number of regions were selected.

Sequencing of the selected regions

For the analysis of the 40 selected genomic regions (Table S2) across 283 samples, targeted sequencing based on multiplex PCR was employed. Specific primers were designed for each region (Table S3), enabling simultaneous amplification of all target sites in a single reaction. Prior to sequencing, molecular barcoding of the original DNA molecules was performed during the amplification step, ensuring accurate counting of

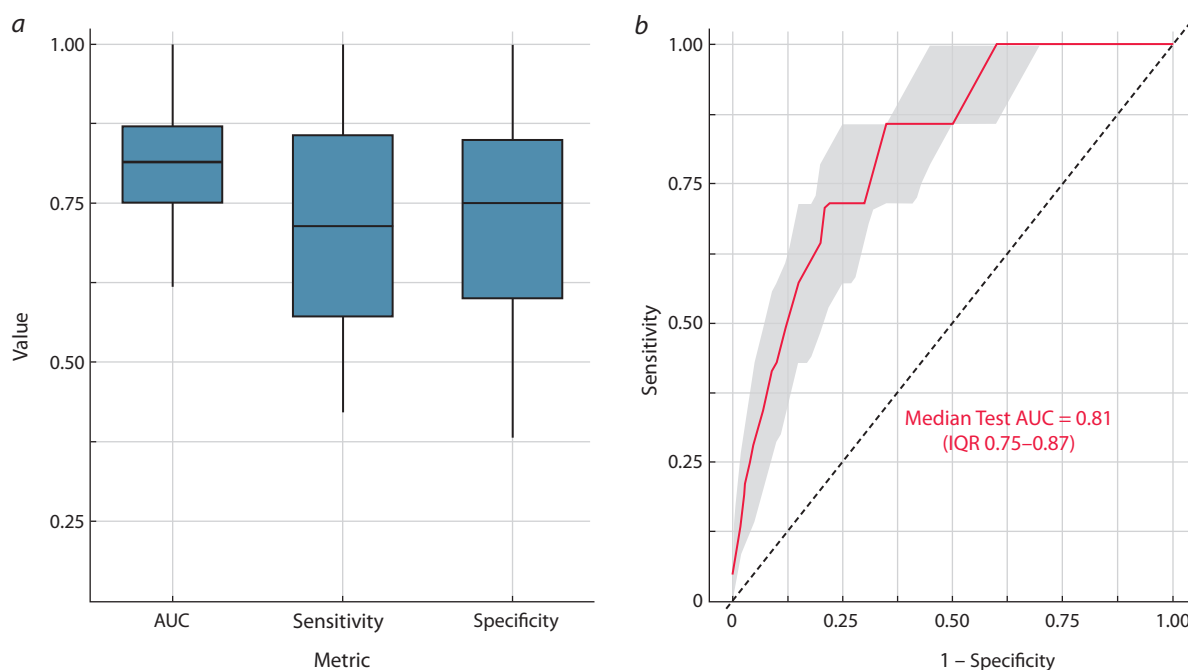


Fig. 2. Model performance: *a* – distribution of metrics (AUC, sensitivity, and specificity); *b* – ROC curve based on 100 independent splits.

initial molecule numbers and minimizing the impact of PCR artifacts. Read mapping was conducted in the same manner as described in the “Whole-genome sequencing data analysis algorithm” section.

Model development and validation

For 201 samples with a normal fetal karyotype and 82 samples with trisomy, counts of cfDNA molecules corresponding to the 40 selected regions were obtained. To reliably assess the suitability of these regions for use in a test system, we trained and tested a LASSO regression model using stratified repeated random subsampling with class weighting (see Material and methods).

The high sensitivity and specificity values (Fig. 2*a*), along with the AUC (Fig. 2*b*) achieved by the model, allow us to conclude that it is possible to predict the presence of fetal trisomy based on sequencing data from DHS regions of circulating cfDNA.

Discussion

Before the analysis of plasma cfDNA became the dominant approach in non-invasive prenatal testing, the scientific community was actively searching for alternative methods to obtain fetal genetic material without invasive procedures (Simpson, Elias, 1993).

Two main cell types served as primary targets in these studies: trophoblasts (Bruch et al., 1991) and nucleated red blood cells (fetal erythroblasts) (Simpson, Elias, 1994). Trophoblasts, being placental derivatives, seemed a logical target as they contain the complete fetal genome and could theoretically serve

as ideal material for analysis (Covone et al., 1984). Concurrently, research was conducted on isolating fetal erythroblasts (Bianchi et al., 1990). The advantage of this approach lay in the fact that immature nucleated red blood cells are normally virtually absent from the peripheral blood of healthy adults, theoretically allowing them to be distinguished against the background of maternal cells.

However, practical application of the cellular approach encountered serious limitations (Holzgreve et al., 1992). The main problem was the extremely low concentration of fetal cells in maternal blood, making their isolation technically challenging and unreliable.

Against this backdrop, the discovery made by Dennis Lo and colleagues in 1997 – demonstrating the presence of significant amounts of fetal cfDNA in maternal plasma – marked the beginning of a fundamentally different path (Lo et al., 1997). Despite the fact that cfDNA represented fragmented placental DNA rather than the complete fetal genome, this approach had decisive practical advantages: a relatively high concentration of genetic material and technological accessibility for analysis using PCR (Lo et al., 1998).

Current commercial NIPTs utilizing high-throughput sequencing methods demonstrate high accuracy, but their widespread adoption is still constrained by relatively high costs (Gil et al., 2017).

Our method, based on targeted analysis of 40 epigenetic markers (DHSs), offers a cost-effective solution. Such approaches require sequencing less than 1 % of the genome, substantially reducing reagent and data processing costs (Schmitt et al., 2015). It is worth noting that not only NGS platforms but

also the more accessible digital PCR (dPCR) technology can be used for this analysis, further lowering costs and simplifying implementation into routine clinical practice (Li et al., 2022). This is particularly relevant for smaller laboratories without access to expensive sequencing systems. At the same time, all key advantages of NIPT are preserved: non-invasiveness, the possibility of early diagnosis, and high accuracy in detecting chromosomal abnormalities (Zhang et al., 2026).

An important aspect of this study was the development of rigorous criteria for selecting informative regions. Our method incorporates a comprehensive set of parameters, including not only statistical measures of differential accessibility but also the functional significance of sites and the dependence of the signal on fetal DNA concentration. Such an integrated approach enables the creation of a robust diagnostic panel resilient to technical noise. Furthermore, the application of molecular barcoding addresses a crucial challenge in quantitative cfDNA analysis – accurate counting of original molecules in the presence of PCR amplification.

Conclusion

The prospects for practical application of the developed approach appear promising. First, the method fundamentally allows for expanding the range of detectable chromosomal abnormalities without substantially increasing analysis costs. Second, its adaptability to various technological platforms opens up possibilities for creating flexible diagnostic systems that can be optimized for specific clinical tasks and budget constraints.

Future research may focus on optimizing the marker panel for simultaneous detection of a wide range of chromosomal abnormalities and standardizing protocols for use in routine clinical practice. The proposed method has every reason to become the foundation for a new generation of screening tests that combine high reliability with cost-effectiveness.

References

- Alberry M.S., Aziz E., Ahmed S.R., Abdel-Fattah S. Non invasive prenatal testing (NIPT) for common aneuploidies and beyond. *Eur J Obstet Gynecol Reprod Biol.* 2021;258:424-429. doi 10.1016/j.ejogrb.2021.01.008
- Bianchi D.W., Flint A.F., Pizzimenti M.F., Knoll J.H., Latt S.A. Isolation of fetal DNA from nucleated erythrocytes in maternal blood. *Proc Natl Acad Sci USA.* 1990;87(9):3279-3283. doi 10.1073/pnas.87.9.3279
- Bruch J.F., Metzeau P., Garcia-Fonknechten N., Richard Y., Tricottet V., Hsi B.L., Kitzis A., Julien C., Papiernik E. Trophoblast-like cells sorted from peripheral maternal blood using flow cytometry: a multiparametric study involving transmission electron microscopy and fetal DNA amplification. *Prenat Diagn.* 1991;11(10):787-798. doi 10.1002/pd.1970111007
- Covone A.E., Mutton D., Johnson P.M., Adinolfi M. Trophoblast cells in peripheral blood from pregnant women. *Lancet.* 1984;2(8407):841-843. doi 10.1016/s0140-6736(84)90875-4
- Dhallan R., Guo X., Emche S., Damewood M., Bayliss P., Cronin M., Barry J., Betz J., Franz K., Gold K., Vallecillo B., Varney J. A non-invasive test for prenatal diagnosis based on fetal DNA present in maternal blood: a preliminary study. *Lancet.* 2007;369(9560):474-481. doi 10.1016/S0140-6736(07)60115-9
- Fan H.C., Blumenfeld Y.J., Chitkara U., Hudgins L., Quake S.R. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc Natl Acad Sci USA.* 2008;105(42):16266-16271. doi 10.1073/pnas.0808319105
- Friedman J., Hastie T., Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Statist Software.* 2010;33(1):1-22. doi 10.18637/jss.v033.i01
- Gil M.M., Accurti V., Santacruz B., Plana M.N., Nicolaides K.H. Analysis of cell-free DNA in maternal blood in screening for aneuploidies: updated meta-analysis. *Ultrasound Obstet Gynecol.* 2017;50(3):302-314. doi 10.1002/uog.17484
- Harrow J., Frankish A., Gonzalez J.M., Tapanari E., Diekhans M., Kokocinski F., Aken B.L., ... Valencia A., Reymond A., Gerstein M., Guigó R., Hubbard T.J. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 2012;22(9):1760-1774. doi 10.1101/gr.135350.111
- Holzgreve W., Garritsen H.S., Ganshirt-Ahlert D. Fetal cells in the maternal circulation. *J Reprod Med.* 1992;37(5):410-418
- Hui L., Bianchi D.W. Fetal fraction and noninvasive prenatal testing: what clinicians need to know. *Prenat Diagn.* 2020;40(2):155-163. doi 10.1002/pd.5620
- Kim S.K., Hannum G., Geis J., Tynan J., Hogg G., Zhao C., Jensen T.J., Mazloom A.R., Oeth P., Ehrich M., van den Boom D., Deciu C. Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. *Prenat Diagn.* 2015;35(8):810-815. doi 10.1002/pd.4615
- Kuhn M. Building predictive models in R using the caret package. *J Stat Softw.* 2008;28(5):1-26. doi 10.18637/jss.v028.i05
- Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357-359. doi 10.1038/nmeth.1923
- Li Y.Q., Tan G.J., Zhou Y.Q. Digital PCR and its applications in noninvasive prenatal testing. *Brief Funct Genomics.* 2022;21(5):376-386. doi 10.1093/bfpg/elac024
- Lo Y.M., Corbetta N., Chamberlain P.F., Rai V., Sargent I.L., Redman C.W., Wainscoat J.S. Presence of fetal DNA in maternal plasma and serum. *Lancet.* 1997;350(9076):485-487. doi 10.1016/S0140-6736(97)02174-0
- Lo Y.M., Tein M.S., Lau T.K., Haines C.J., Leung T.N., Poon P.M., Wainscoat J.S., Johnson P.J., Chang A.M., Hjelm N.M. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am J Hum Genet.* 1998;62(4):768-775. doi 10.1086/301800
- Lui Y.Y., Chik K.W., Chiu R.W., Ho C.Y., Lam C.W., Lo Y.M. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin Chem.* 2002;48(3):421-427. doi 10.1093/clinchem/48.3.421
- Pantiukh K.S., Chekanov N.N., Zaigrin I.V., Zotov A.M., Mazur A.M., Prokhortchouk E.B. Report on noninvasive prenatal testing: classical and alternative approaches. *F1000Res.* 2016;5:722. doi 10.12688/f1000research.8243.1
- Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., ... Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12(85):2825-2830
- Peng Q., Xu C., Kim D., Lewis M., DiCarlo J., Wang Y. Targeted single primer enrichment sequencing with single end duplex-UMI. *Sci Rep.* 2019;9(1):4810. doi 10.1038/s41598-019-41215-z
- Satam H., Joshi K., Mangrolia U., Waghoo S., Zaidi G., Rawool S., Thakare R.P., Banday S., Mishra A.K., Das G., Malonia S.K. Next-

- generation sequencing technology: current trends and advancements. *Biology*. 2023;12(7):997. doi 10.3390/biology12070997
- Schmitt M.W., Fox E.J., Prindle M.J., Reid-Bayliss K.S., True L.D., Radich J.P., Loeb L.A. Sequencing small genomic targets with high efficiency and extreme accuracy. *Nat Methods*. 2015;12(5):423-425. doi 10.1038/nmeth.3351
- Sheffield N.C., Thurman R.E., Song L., Safi A., Stamatoyannopoulos J.A., Lenhard B., Crawford G.E., Furey T.S. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res*. 2013;23(5):777-788. doi 10.1101/gr.152140.112
- Simpson J.L., Elias S. Isolating fetal cells from maternal blood. *JAMA*. 1993;270(19):2357-2361
- Simpson J.L., Elias S. Isolating fetal cells in maternal circulation for prenatal diagnosis. *Prenat Diagn*. 1994;14(13):1229-1242. doi 10.1002/pd.1970141308
- Thurman R.E., Rynes E., Humbert R., Vierstra J., Maurano M.T., Haugen E., Sheffield N.C., ... Kaul R., Furey T.S., Dekker J., Crawford G.E., Stamatoyannopoulos J.A. The accessible chromatin landscape of the human genome. *Nature*. 2012;489(7414):75-82. doi 10.1038/nature11232
- Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M., Rozen S.G. Primer3 – new capabilities and interfaces. *Nucleic Acids Res*. 2012;40(15):e115. doi 10.1093/nar/gks596
- van Beek D.M., Straver R., Weiss M.M., Boon E.M.J., Huijsdens-van Amsterdam K., Oudejans C.B.M., Reinders M.J.T., Sijtermans E.A. Comparing methods for fetal fraction determination and quality control of NIPT samples. *Prenat Diagn*. 2017;37(8):769-773. doi 10.1002/pd.5079
- Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag, 2016
- Youden W.J. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-35. doi 10.1002/1097-0142(1950)3:1%3C32::aid-cnrcr2820030106%3E3.0.co;2-3
- Zhang L., Chen L., Yu Y., Wu Z., Zhao M., Wang P., Liu Z., Dai J., Li Y., Wang M. Digital PCR in prenatal diagnosis. *Clin Chim Acta*. 2026;578:120535. doi 10.1016/j.cca.2025.120535

Conflict of interest. Mazur A.M. is the science director at Genoanalytica LLC, a company involved in NIPT development. However, this study was not funded by Genoanalytica LLC and is not related to the company's commercial interests.

Received December 18, 2025. Revised February 6, 2026. Accepted February 9, 2026.

doi 10.18699/vjgb-26-68

The paradigm of somatic mosaicism in complex diseases

A.A. Sleptcov ^{1, 2} , M.S. Nazarenko ^{1, 2} , V.P. Puzyrev ¹¹ Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia² Tyumen Cardiology Research Center – Branch of Tomsk National Research Medical Center of the Russian Academy of Sciences, Tyumen, Russia alexei.sleptcov@medgenetics.ru

Abstract. The multifactorial etiology of complex diseases involves the interplay of polygenic/oligogenic susceptibility loci and environmental factors. Complex diseases are characterized by pronounced phenotypic variability, genetic heterogeneity, pleiotropy of genetic variants, variable penetrance, and expressivity. Advances in population-wide genomic sequencing have expanded our understanding of the genetic architecture of complex diseases significantly; however, germline variants explain only a fraction of the observed phenotypic variability. The introduction of deep DNA sequencing and single-cell multi-omics analysis has revealed an additional, previously underestimated category of risk factors: somatic mutations that continuously accumulate in cells throughout an individual's lifespan, giving rise to genetic mosaicism. This review considers the role of somatic mutations in the pathogenesis of complex diseases in the context of their combined contribution with germline determinants to the formation of disease phenotypes. Particular attention is paid to clonal hematopoiesis of indeterminate potential as the best-studied model of somatic mosaicism associated with age-related conditions. It is demonstrated that the interaction of somatic and germline variants occurs within specific tissue contexts through mechanisms of clonal selection, stochastic clonal drift, and epigenetic dysregulation, causing organ dysfunction. Within the concept of somatic mosaicism, this work suggests a selection mechanism alternative to classical oncogenesis. This pathway, defined as “passive clonal dominance”, describes selection through the persistence of clones that gain advantage not via proliferation but through resistance to apoptosis under chronic stress, which may be particularly relevant to post-mitotic tissues such as the heart muscle and nervous tissue. From a practical standpoint, the somatic variants discussed herein are of interest as candidate biomarkers for predictive diagnostics and risk stratification of complex diseases, pending clinical validation. Moreover, they point to pathophysiological pathways that may reveal targets for therapies aimed at modulating clonal composition and preventing disease progression. The paper also discusses prospects for studying somatic mosaicism in the context of complex diseases, including the potential of dynamic lineage tracing technologies for experimental verification of the proposed clonal selection mechanisms, as well as the need to integrate somatic and germline genetic variant data into unified models for individual disease risk assessment.

Key words: somatic mutations; cardiovascular diseases; clonal hematopoiesis; complex/common diseases

For citation: Sleptcov A.A., Nazarenko M.S., Puzyrev V.P. The paradigm of somatic mosaicism in complex diseases. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2026;30(4):676-684. doi 10.18699/vjgb-26-68

Funding. This work was partially supported by an internal grant from Tomsk National Research Medical Center “Clonal hematopoiesis of uncertain potential as a risk factor for chronic heart failure: the role of somatic mutations and systemic inflammation”.

Парадигма соматического мозаицизма при многофакторных заболеваниях

А.А. Слепцов ^{1, 2} , М.С. Назаренко ^{1, 2} , В.П. Пузырев ¹¹ Научно-исследовательский институт медицинской генетики Томского национального исследовательского медицинского центра Российской академии наук, Томск, Россия² Тюменский кардиологический научный центр – филиал Томского национального исследовательского медицинского центра Российской академии наук, Тюмень, Россия alexei.sleptcov@medgenetics.ru

Аннотация. Многофакторные заболевания (МФЗ) характеризуются сложной этиологией, включающей взаимодействие полигенных/олигогенных наследственных детерминант и факторов среды, а также выраженным клиническим полиморфизмом, генетической гетерогенностью, плейотропией генетических вариантов, вариабельной пенетрантностью и экспрессивностью. Достижения в области секвенирования геномов популяций существенно расширили понимание наследственной архитектуры МФЗ, однако герминативные варианты объясняют лишь часть наблюдаемой фенотипической вариабельности. Внедрение методов глубокого секвенирования ДНК и мультиомиксного анализа отдельных клеток позволило выявить дополнительную, ранее

недооцененную категорию факторов риска, а именно соматические мутации, непрерывно накапливающиеся в клетках на протяжении онтогенеза и формирующие мозаичную генетическую конституцию организма (соматический мозаицизм). В настоящем обзоре рассматривается роль соматических мутаций в этиопатогенезе МФЗ с позиции их сочетанного вклада с герминативными детерминантами в формирование патологического фенотипа. Особое внимание уделено феномену клонального гемопоэза неопределенного потенциала как наиболее изученной модели соматического мозаицизма, ассоциированного с возраст-зависимой патологией. Показано, что взаимодействие соматических и герминативных вариантов реализуется в клетках в специфических тканевых контекстах через механизмы клональной селекции, стохастического клонального дрейфа и эпигенетической дисрегуляции, приводящих к дисфункции органов. В рамках концепции соматического мозаицизма в работе обосновывается механизм селекции, альтернативный классическому онкогенезу. Данный путь определяется как «пассивная клональная доминантность» и описывает селекцию как персистенцию клонов, получивших преимущество не за счет пролиферации, а благодаря резистентности к апоптозу в условиях хронического стресса. Такой механизм может быть особенно важен для постмитотических тканей, таких как миокард и нервная ткань. С практической точки зрения рассмотренные соматические варианты представляют потенциальный интерес в качестве кандидатных биомаркеров для предиктивной диагностики и стратификации риска МФЗ, что требует дальнейшей клинической валидации. Более того, они указывают на патофизиологические пути, потенциально открывающие мишени для таргетной терапии, направленной на модуляцию клонального состава и предотвращение прогрессирования заболеваний. В заключение даны перспективы и направления изучения соматического мозаицизма в контексте МФЗ, включая потенциал технологий динамического отслеживания клеточных линий для экспериментальной верификации предложенных механизмов клональной селекции. Отмечена важность интеграции данных о соматических и герминативных генетических вариантах в единые модели оценки индивидуального риска заболеваний.

Ключевые слова: соматические мутации; сердечно-сосудистые заболевания; клональный гемопоэз; многофакторные заболевания

Introduction

Complex diseases (CDs) are characterized by high prevalence and broad phenotypic diversity across populations. Although these disorders do not follow classic Mendelian inheritance patterns, familial clustering is well documented and is best explained by polygenic or oligogenic models incorporating substantial environmental contributions. Importantly, the balance between genetic susceptibility and environmental exposures varies across individuals, even within the same diagnostic category.

The phenotypic manifestation of CDs reflects core genetic principles, including clinical heterogeneity, locus heterogeneity, extensive pleiotropy of risk variants, and variable penetrance and expressivity, each shaped by the broad polygenic architecture and gene-environment interactions.

Beyond this genetic complexity, the disease phenotype in complex disorders exhibits multilayered heterogeneity spanning molecular, cellular, and tissue levels. Disease trajectories are dynamic: they evolve over time, and display spatial specificity driven by variation in local tissue microenvironments and cell-type-specific transcriptional programs. Substantial variability is observed among patients diagnosed with the same CDs, both in clinical presentation and in underlying pathogenic mechanisms.

Unlike inherited germline variants, somatic mutations arise *de novo* and accumulate continuously throughout the lifespan. The dynamics of this accumulation ranges from approximately linear in postmitotic tissues to markedly nonlinear in the setting of clonal selection (Lodato et al., 2018; Brunner et al., 2019). This ongoing muta-

genesis generates intraorganismal genetic diversity, or *somatic mosaicism*, which is a form of postzygotic variation that can shape individual phenotypic outcomes. The accompanying epigenetic drift across clonal populations may further amplify phenotypic heterogeneity.

This review addresses the role of somatic mosaicism in CD pathogenesis. We focus on somatic variants as tissue-specific modifiers of hereditary risk and as autonomous drivers of disease initiation and progression via clonal selection, stochastic clonal drift, and epigenetic dysregulation, which remodel the clonal architecture of affected tissues.

Somatic genetic variants and complex diseases

During embryogenesis, the high rate of cell divisions, coupled with shortened cell cycles and relaxed checkpoint control, leads to the accumulation of *de novo* mutations. When a progenitor cell carrying such a mutation remains viable and maintains its proliferative potential, it gives rise to a clone whose genome diverges from the constitutional germline genome.

In adult organisms, somatic mutations arise ubiquitously, but their accumulation and clonal expansion occur predominantly in tissues with high cellular turnover, such as the hematopoietic system and barrier epithelia, or in parenchymal organs under conditions of chronic injury and regeneration. Recurrent cycles of cell division increase the likelihood of acquiring somatic mutations, among which driver mutations in genes that directly regulate the cell cycle, proliferation, and cell survival are of particular importance.

Concurrently, neutral passenger mutations accumulate without directly affecting cell fitness. However, some mutations that do not meet the classic definition of oncogenic drivers may still exert indirect effects on cell survival through modulation of metabolic pathways, stress responses, or intercellular interactions within the microenvironment, thereby conferring a *selective advantage* in specific tissue environments.

Age-related decline in DNA repair efficiency contributes to the accumulation of somatic genetic variants, a subset of which confers a selective advantage to cells. As the regenerative capacity of tissues becomes depleted, this process drives clonal expansion of the most viable cell populations. Consequently, mosaic clonal popula-

tions emerge within tissues and bring the interindividual genetic diversity to such a degree that in some tissues it may exceed that observed at the population level (Harris, 2025).

Somatic mutagenesis has historically been a focus of cancer genomics research, as the accumulation of driver mutations in oncogenes and tumor-suppressor genes is considered pivotal in the pathogenesis of malignancy.

Advances in genomic technologies, including next-generation sequencing and single-cell genomics, have enabled the identification of nonmalignant complex diseases in which somatic mutations of various classes accumulate (see the Table).

Somatic genetic variants in complex diseases

Somatic mutation type	Complex disease	Tissue, organ	PMID*
Structural and numerical chromosomal abnormalities	Spontaneous abortion (miscarriage)	Fetal tissues and parental leukocytes	11675616
Large clonal mosaic chromosomal rearrangements (>2 Mb); copy number variants (CNVs); loss of heterozygosity (LOH); single-nucleotide substitutions	Cancer	Leukocytes, tumor	22561519, 22561516
	Neuropsychiatric disorders: intellectual disability, autism, epilepsy, bipolar disorder, schizophrenia, behavioral abnormalities	Leukocytes, brain	23828942, 23888031
Large clonal mosaic chromosomal rearrangements (>2 Mb)	Type 2 diabetes mellitus with vascular complications	Leukocytes	23852171
Copy number variants (CNVs)	Childhood asthma	Leukocytes	19728390, 30262839
Single-nucleotide substitutions	Early-onset Alzheimer disease	Leukocytes, brain	30670873, 31300647, 35444284
	Atrial fibrillation	Leukocytes, myocardium	16790700, 20606116
	Congenital heart defects	Leukocytes, saliva, heart muscle	29729027, 32349777
	Venous malformations	Leukocytes, vascular tissue	19079259
	Cerebral cavernous malformations	Leukocytes, vascular tissue	33729480, 34496175, 36995941, 41366709, 35852613
	Endometriosis	Endometrium	30110635
	Inflammatory bowel disease	Intestinal crypts	32697969
	Atherosclerosis, vascular malformations, aortic aneurysm, heart failure, Alzheimer disease, type 2 diabetes mellitus, and chronic liver and kidney diseases	Leukocytes	40175709, 31406340, 34298011, 33731931, 31672865, 33057201, 28636844, 37046084, 37197843
	Autism	Leukocytes from proband-parent trios and other family members	33432195

* Pubmed ID.

A substantial fraction of postzygotic mutations likely remains undetected due to *negative selection*. Deleterious loss-of-function (LoF) variants in essential genes trigger apoptosis or irreversible cell cycle arrest in affected cells, rendering them invisible to standard genomic analysis methods.

Trio-based DNA sequencing (mother-father-child) enables the identification of *de novo* germline mutations. In contrast, detection of somatic variants requires comparative analysis of different tissues from the same individual, as well as the use of deep sequencing or single-cell sequencing technologies to detect low-level mosaicism (Shao et al., 2025).

Recent studies using single-cell genome sequencing have revised traditional views on the burden of somatic mutations in postmitotic tissues. Neurons from healthy adults have been shown to carry more than 1,000 somatic single-nucleotide variants (SNVs) per cell, markedly exceeding earlier estimates (Lodato et al., 2018).

Similarly, an analysis of 48 cardiac muscle cells from 10 healthy subjects of different ages has demonstrated that a single cardiac cell contains 4,000 to 30,000 somatic SNVs, with mutation burden showing a strong correlation with subject age (Choudhury et al., 2022). These findings, obtained through direct mutation counting in individual cells, indicate that somatic SNV accumulation is a widespread and substantial process in nonproliferating cells, potentially contributing to physiological aging and disease susceptibility.

Postzygotic mutations represent a fundamentally distinct, non-Mendelian source of phenotypic variability. They can not only modify the penetrance and expressivity of inherited predisposition but also initiate pathological processes independently through clonal selection mechanisms within target tissues. Thus, somatic mosaicism may function not only as a modifier of inherited risk but also as an independent pathogenic factor contributing to disease development in specific clinical contexts.

Below, we discuss the role of somatic mutations in CDs using cardiovascular disorders as an example.

Somatic genomic variability in cardiovascular diseases

Somatic mutagenesis, which gives rise to tissue mosaicism, affects not only parenchymal cells of target organs but also blood cells. In the hematopoietic system, this process manifests itself as clonal hematopoiesis (CH), in which a hematopoietic stem cell (HSC) harboring a driver mutation acquires a selective advantage, initiating clonal expansion.

Of particular importance is the phenomenon of clonal hematopoiesis of indeterminate potential (CHIP), in

which a clonal population of cells carrying somatic mutations in genes associated with hematologic malignancies is present in peripheral blood, in the absence of evidence for myelodysplasia, leukemia, or other hematologic neoplasm in the individual (Jaiswal et al., 2014; Khoury et al., 2022).

Extensive research of CHIP in recent years has not only established its status as an independent risk factor for CDs but also added much to our understanding of the pathogenic role of somatic mosaicism by demonstrating that mutant hematopoietic clones can induce a state of chronic sterile systemic inflammation.

The pathogenic significance of this phenomenon lies in the acquisition of aberrant proinflammatory activity by mutant immune cell clones (primarily macrophages). Their ability to overproduce cytokines establishes a state of low-grade chronic systemic inflammation. This process is considered the key mechanism through which CHIP contributes to increased risk of atherosclerosis, coronary artery disease, heart failure, and other age-related disorders (Jaiswal et al., 2017).

It is important to note that the pathogenic impact of somatic mosaicism on the cardiovascular system is not confined to the hematopoietic compartment. An equally significant parallel mechanism is oligoclonal expansion of resident cells of the vascular wall, particularly smooth muscle cells. In response to injury or atherogenic stimuli, individual progenitor cells acquire a competitive advantage, forming clonal clusters within atherosclerotic plaques. This process fundamentally affects plaque stability and progression (Lin et al., 2023).

Clonal hematopoiesis of indeterminate potential in cardiovascular disease: molecular mechanisms and key determinants

According to the consensus criteria proposed by D.P. Steensma et al. (2015), mutations in CHIP are defined by a variant allele frequency (VAF) of 2 % or greater. They occur in a relatively restricted set of hematopoietic driver genes, including *DNMT3A*, *TET2*, *ASXL1*, *JAK2*, and others (Steensma et al., 2015). We should note that this threshold reflects the sensitivity limit of standard next-generation sequencing and does not preclude the pathogenic relevance of clones with lower VAFs.

The prevalence of CHIP correlates with age. Using the standard detection threshold ($VAF \geq 2\%$), prevalence reaches 15–20 % among individuals older than 70 years, whereas deep sequencing approaches with higher sensitivity reveal substantially higher frequencies of low-level clones. Although CHIP is not associated with clinically overt cytopenias, it shows associations with subcli-

nical alterations in erythrocyte indices, most notably wider red blood cell distribution and larger mean corpuscular volume (Acuna-Hidalgo et al., 2017; Walsh et al., 2022).

In individuals with CHIP, the relative risk of developing hematologic malignancies increases approximately 10- to 13-fold across multiple cohort studies, although this estimate varies considerably depending on clone size and the spectrum of driver mutations (Genovese et al., 2014; Jaiswal et al., 2014).

The risk of clinically overt coronary artery disease (CAD) depends on a specific mutated driver gene in CHIP. Mutations in *DNMT3A*, *TET2*, or *ASXL1* are associated with a double risk of CAD, whereas the carriership of the *JAK2* p.Val617Phe mutation is associated with a 12-fold increase in relative risk, an effect presumably mediated by a concomitant subclinical myeloproliferative phenotype (Jaiswal et al., 2017).

A direct correlation has been established between mutant clone size (VAF) and the atherosclerotic burden, as assessed by the coronary artery calcium score (Agatston score). Among patients with CHIP, a high frequency of ischemic stroke and other cardiovascular events is observed, which increases all-cause mortality in this group.

Although the panel of genes associated with CHIP includes up to approximately 100 loci, mutations in *DNMT3A*, *TET2*, and *ASXL1* account for the majority of cases. They are identified in a substantial proportion of CAD patients with CHIP (Jaiswal et al., 2017). Loss-of-function mutations in *DNMT3A* and *TET2* in mouse models lead to disruption of epigenetic control, resulting in enhanced HSC self-renewal, impaired differentiation, and subsequent clonal dominance in the hematopoietic system (Challen et al., 2011; Moran-Crusio et al., 2011).

The molecular pathogenesis of *DNMT3A* mutations involves focal hypomethylation of specific genomic loci and subsequent aberrant derepression of genes that regulate HSC self-renewal, including the *Hox* cluster (*HOXA5*, *HOXA7*, and *HOXA9*). Loss of *TET2* function impairs oxidative demethylation, leading to focal persistence of 5-methylcytosine in regulatory regions of the genome, which disrupts the enhancer landscape and blocks terminal differentiation of myeloid progenitors (Challen et al., 2011; Moran-Crusio et al., 2011).

Although histone modification mediated by *ASXL1* protein is recognized as a fundamental process required for physiological hematopoiesis, the detailed molecular mechanisms that confer a selective advantage to mutant cells in the setting of age-related CH remain to be insufficiently elucidated (Yamamoto et al., 2021).

Whereas wild-type *ASXL1* maintains precise epigenetic control through regulation of repressive complexes, the pathogenesis of CHIP is closely linked to specific mutations, predominantly frameshifts in exon 12. Mutations in *ASXL1* are known to produce a stable truncated protein (*ASXL1-trunc*) with gain-of-function properties. This protein intensely destabilizes the PRC2 complex while simultaneously enhancing the interaction with BAP1 deubiquitinase within the PR-DUB complex. The synergistic effect of these activities leads to global reduction in H3K27me3 levels and aberrant derepression of critical leukemogenic loci, including *HOXA* cluster genes, which may confer a selective advantage to mutant cells (Inoue et al., 2016; Nagase et al., 2018; Yang et al., 2018).

Somatic mutations in *JAK2*, *PPM1D*, and *TP53* associated with CHIP contribute to pathogenesis when occur in patients with cardiovascular diseases. In contrast to *TET2* and *DNMT3A*, mutations in *TP53* and *PPM1D* often emerge under selective pressure, such as chemotherapy, and they are associated with a more aggressive clone. The *JAK2* tyrosine kinase acts as a key component of the JAK–STAT signaling pathway. The presence of somatic mutations induces constitutive kinase domain activity, resulting in cytokine-independent proliferation and conferring a selective advantage to the mutant clone (Kralovics et al., 2005).

The *TP53*-encoded p53 protein serves as a central regulator of the response to DNA damage (DDR) and cellular stress. Hematopoietic stem cells carrying loss-of-function mutations in *TP53* acquire a selective advantage by evading apoptosis and escaping cellular senescence, leading to clonal dominance under conditions of competitive interaction with wild-type cells (Bondar, Medzhitov, 2010). An additional factor contributing to the clonal advantage of *TP53*-mutant cells may be higher tolerance to oxidative stress and genotoxic impacts. *PPM1D* also functions within the DDR pathway as a downregulator of p53, and truncating mutations in exon 6 provide a proliferative advantage to HSCs through enhanced resistance to p53-mediated apoptosis (Hsu et al., 2018).

Although CHIP mutations are detected in somatic cells, mounting evidence demonstrates the existence of inherited predisposition to their acquisition in HSCs. In 2009, a genome-wide association study (GWAS) identified an association between a single-nucleotide variant (rs10974944) marking the *JAK2* 46/1 haplotype and predisposition to acquiring the somatic *JAK2* p.Val617Phe mutation, as well as subsequent development of myeloproliferative neoplasms (Kilpivaara et al., 2009).

It was found later that the risk of CHIP is tightly associated with genetic variants in the *TERT* gene (in particular,

rs144418061), which encodes the catalytic component of the telomerase complex required for telomere maintenance, and with a polymorphism at the *TCL1A* locus (rs2887399) (Bick et al., 2020).

Furthermore, genetic analysis has revealed partial overlap between loci that predispose to acquisition of CHIP mutations (termed *heritability of somatic mutation predisposition*) and loci associated with cardiovascular disease risk in GWAS. This observation supports the existence of shared germline determinants that modulate individual risk profiles by simultaneously affecting the kinetics of mutant HSC clonal expansion and vascular resilience through independent pathogenic pathways, exemplifying genetic pleiotropy.

The kinetics of clonal expansion of mutant HSCs is tightly controlled by epigenetic regulatory mechanisms. A strong correlation between the growth rate of mutant clones and measures of epigenetic aging, assessed both through direct DNA methylation profiling and by using genetic predictors, has been confirmed. Markers of biological age, including the PhenoAge, GrimAge, and HannumAge algorithms, show positive associations with both the presence of CHIP and the rate of clonal hematopoietic cell proliferation (Nachun et al., 2021; Mack et al., 2024).

In addition to constitutional and endogenous determinants, exogenous stressors and proinflammatory settings significantly influence the dynamics of mutant HSC clonal expansion. Cigarette smoking and iatrogenic interventions, particularly cytotoxic therapy, increase the frequency of somatic mutations. Chronic systemic inflammation acts primarily as a selective pressure driving clonal outgrowth. Collectively, these processes create an adaptive advantage for clones that are resistant to apoptosis and cytotoxic effects.

The phenomenon of CH shows significant associations not only with progression of atherosclerosis and thromboembolic complications but also with a broad range of heterogeneous age-related conditions, positioning it as a key driver of the chronic systemic inflammation that accompanies aging (inflammaging). The range of comorbid disorders includes atrial fibrillation, heart failure, abdominal aortic aneurysm, venous thromboembolism, as well as Alzheimer's disease-related neurodegeneration and metabolic disorders, particularly type 2 diabetes mellitus (Schuermans, Honigberg, 2025).

Studies in mouse models, first of all, bone marrow transplantation from *Tet2* or *Dnmt3a* knockout mice, demonstrate that CHIP-associated clones of macrophages and other myeloid derivatives produce excessive amounts of proinflammatory cytokines and chemokines, such as IL-1 β , IL-6, and CXCL2. This persistent proin-

flammatory signaling induces dysfunction in peripheral tissues and target organs, initiating the development of insulin resistance and disruption of tissue homeostasis. These pathophysiological shifts recapitulate the multisystem phenotype of accelerated aging observed in clinical practice (Schuermans, Honigberg, 2025).

A specific form of clonal mosaicism is mosaic loss of the Y chromosome (mLOY) in peripheral blood leukocytes, which represents the most common somatic event in men. This phenomenon is associated with increased risk of cardiovascular disease, particularly the development of cardiac fibrosis and heart failure (Sano et al., 2022), as well as with severe progression of neurodegenerative disorders (Ljungström et al., 2022). In men with pronounced mLOY, a trend toward increased VAF of CHIP-associated mutations is observed, suggesting that these events may coexist within the same clone. The available evidence indicates that CHIP and mLOY frequently co-occur during aging, supporting their consideration as potentially interconnected manifestations of common age-related changes in the clonal architecture of bone marrow (Ljungström et al., 2022).

Despite these associations, detailed elucidation of the molecular cascades linking CHIP to the broad spectrum of CDs requires further investigation. To explain mechanisms of cell selection, R. Happle (1993) put forward the concept of paradiplomancy, which presumes that a somatic mutation arising in the setting of germline heterozygosity at the same locus leads to biallelic inactivation and local manifestation of a pathological phenotype during early development. However, the applicability of this model to CHIP driver genes such as *TET2* or *DNMT3A* remains a matter of ongoing debate, because in this context the clonally significant phenotype often arises from monoallelic inactivation.

Nevertheless, the theoretical potential of this model lies in its ability to provide a mechanistic explanation for the phenomena of incomplete penetrance and variable expressivity in CDs (Puzyrev et al., 2014). This is particularly relevant for driver genes in which constitutional homozygous inactivation often results in severe developmental defects or is embryonic lethal.

Conclusion and prospects of studying somatic mosaicism in complex diseases

Detailed research of the contribution of somatic mosaicism to CD etiopathogenesis opens prospects for the development of fundamentally new strategies for molecular diagnosis and therapeutic intervention. Such approaches should target not only the affected organ but also the selective elimination or reprogramming of specific pathogenic clones that contribute to disease pro-

gression, therapy resistance, and clinical heterogeneity. A fundamental requirement of this therapeutic paradigm is the necessity of accounting for the temporal dynamics of clonal architecture. The composition and population of pathogenic clones exhibit temporal instability, being subject to modification under selective pressures, including the cytotoxic effects of therapeutic agents.

Mechanisms of clonal dynamics are primarily considered through the lens of *neutral drift*, defined as stochastic replacement in the absence of selective pressure, as described in intestinal epithelium (Klein, Simons, 2011), or *positive selection*, which confers a proliferative advantage to mutant clones in renewing tissues such as the hematopoietic system. However, somatic mosaicism in postmitotic tissues, particularly the accumulation of mutations in neurons and cardiac muscle cells, is often interpreted merely as a passive process of *genetic noise* accumulation accompanying aging (Lodato et al., 2018). However, these frameworks do not fully account for the potentially nonrandom nature of mosaic landscape formation in tissues with extremely low proliferative capacity.

In this review, we propose a conceptual model of an alternative mechanism for clonal selection in postmitotic tissues, which we term *passive clonal dominance*, realized through selection for persistence. This term refers to the increase in relative fraction of clones that do not possess a direct proliferative advantage but instead exhibit enhanced resistance to elimination under conditions of chronic stress.

In proliferating tissues such as the hematopoietic system or epithelium, Darwinian selection favors clones harboring mutations that enhance proliferation and survival, commonly referred to as driver mutations. However, a fundamentally different scenario may occur in functionally specialized tissues with low or absent regenerative capacity, including populations of cardiac muscle cells and neurons. Under these conditions, the advantage is conferred not to proliferating clones but to populations of resilient cell populations characterized by enhanced resistance to elimination under chronic stress, so that the mosaic tissue architecture is remodeled, as cells with standard apoptotic thresholds are eliminated.

Such cellular selection may be driven by mutations that, within the specific microenvironment of postmitotic tissues, do not confer a direct proliferative advantage but instead promote cell survival. These include inactivation of genes controlling apoptosis, the integrated stress response (ISR), or metabolic homeostasis. In this model, the increased representation of a mutant clone is achieved not through mitotic activity but through its prolonged

persistence, while neighboring wild-type cells with lower stress resistance undergo accelerated apoptosis.

A logical consequence of this process is the gradual accumulation of functionally impaired but elimination-resistant cellular clones, representing a distinct form of somatic evolution shaped by selective pressure from the microenvironment. This hypothetical process, through which tissues become enriched in clones with preserved viability but compromised function, may contribute to age-related tissue dysfunction and clinical manifestation of CDs alongside other well-established mechanisms. Experimental validation will require longitudinal studies and approaches for tracking clonal dynamics *in vivo*.

A key tool for such validation is dynamic lineage tracing, an approach that has already transformed our understanding of clonal architecture in hematopoiesis and can be adapted for the study of postmitotic tissues. Based on site-specific recombination systems (Cre-lox) combined with polychromatic reporter constructs, such as the Confetti or Rainbow systems, and high-throughput single-cell sequencing technologies (scRNA-seq), this approach has enabled direct observation of clonal expansion kinetics within tissue architecture (McKenna, Gagnon, 2019).

Our initial understanding of CHIP was shaped by large-scale population studies using next-generation sequencing. The study of CHIP by lineage-tracing technologies have facilitated a shift from static detection of mutant clones to analysis of their dynamic history (Haghverdi, Ludwig, 2023; Jindal et al., 2024). In particular, it has become possible to determine the hierarchical origin of a clone by identifying the specific hematopoietic stem or progenitor cell that initiated expansion with mutations in *DNMT3A*, *TET2*, or *ASXL1*, and to reconstruct the complete phylogenetic tree of the clone within the bone marrow niche.

In addition, these methods enable real-time observation of competitive cellular dynamics, such as tracking how a clone carrying a driver mutation, e. g., *JAK2* p.Val617Phe, outcompetes wild-type cells during aging or in response to proinflammatory stimuli (Haghverdi, Ludwig, 2023; Jindal et al., 2024). Finally, it has become possible to link clonal identity to its functional status, transcriptional profile, and epigenome at the single-cell level, thereby identifying determinants that account for the quiescence of some clones and the pronounced proinflammatory potential of others (Xie, Zeidan, 2023; Jindal et al., 2024).

Application of such integrated approaches to the study of somatic mosaicism is promising for transforming our knowledge of the molecular events through which

somatic mutations reprogram the function of hematopoietic and other somatic cells and thereby induce systemic inflammation and increase the risk of CDs (Jindal et al., 2024). Elucidating the mechanisms of clonal selection, including the proposed model of *passive clonal dominance*, provides grounds for the development of targeted interventions aimed at modulating tissue clonal architecture, which in turn may contribute to the prevention of a broad range of age-related conditions.

References

- Acuna-Hidalgo R., Sengul H., Steehouwer M., van de Vorst M., Vermeulen S.H., Kiemeny L.A.L.M., Veltman J.A., Gilissen C., Hoeschen A. Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. *Am J Hum Genet.* 2017;101(1):50-64. doi 10.1016/j.ajhg.2017.05.013
- Bick A.G., Weinstock J.S., Nandakumar S.K., Fulco C.P., Bao E.L., Zekavat S.M., Szeto M.D., ... Natarajan P. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature.* 2020;586(7831):763-768. doi 10.1038/s41586-020-2819-2
- Bondar T., Medzhitov R. p53-mediated hematopoietic stem and progenitor cell competition. *Cell Stem Cell.* 2010;6(4):309-322. doi 10.1016/j.stem.2010.03.002
- Brunner S.F., Roberts N.D., Wylie L.A., Moore L., Aitken S.J., Davies S.E., Sanders M.A., ... Abascal F., Stratton M.R., Martincorena I., Hoare M., Campbell P.J. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature.* 2019;574(7779):538-542. doi 10.1038/s41586-019-1670-9
- Challen G.A., Sun D., Jeong M., Luo M., Jelinek J., Berg J.S., Bock C., ... Meissner A., Issa J.P., Godley L.A., Li W., Goodell M.A. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet.* 2011;44(1):23-31. doi 10.1038/ng.1009
- Choudhury S., Huang A.Y., Kim J., Zhou Z., Morillo K., Maury E.A., Tsai J.W., ... Araten S., Hilal N., Lee E.A., Chen M.H., Walsh C.A. Somatic mutations in single human cardiomyocytes reveal age-associated DNA damage and widespread oxidative genotoxicity. *Nat Aging.* 2022;2(8):714-725. doi 10.1038/s43587-022-00261-5
- Genovese G., Kähler A.K., Handsaker R.E., Lindberg J., Rose S.A., Bakhoum S.F., Chambert K., ... Sullivan P.F., Sklar P., Grönberg H., Hultman C.M., McCarroll S.A. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med.* 2014;371(26):2477-2487. doi 10.1056/NEJMoa1409405
- Haghverdi L., Ludwig L.S. Single-cell multi-omics and lineage tracing to dissect cell fate decision-making. *Stem Cell Rep.* 2023;18(1):13-25. doi 10.1016/j.stemcr.2022.12.003
- Happle R. Mosaicism in human skin. Understanding the patterns and mechanisms. *Arch Dermatol.* 1993;129(11):1460-1470
- Harris K. An uneasy truce between population health and the gene pools within our bodies. *Nat Rev Genet.* 2025;26(7):441. doi 10.1038/s41576-025-00851-0
- Hsu J.I., Dayaram T., Tovy A., De Braekeleer E., Jeong M., Wang F., Zhang J., ... Vassiliou G., Futreal P.A., Donehower L.A., Takahashi K., Goodell M.A. PPMID mutations drive clonal hematopoiesis in response to cytotoxic chemotherapy. *Cell Stem Cell.* 2018;23(5):700-713.e6. doi 10.1016/j.stem.2018.10.004
- Inoue D., Matsumoto M., Nagase R., Saika M., Fujino T., Nakayama K.I., Kitamura T. Truncation mutants of ASXL1 observed in myeloid malignancies are expressed at detectable protein levels. *Exp Hematol.* 2016;44(3):172-176.e1. doi 10.1016/j.exphem.2015.11.011
- Jaiswal S., Fontanillas P., Flannick J., Manning A., Grauman P.V., Mar B.G., Lindsley R.C., ... Koistinen H.A., Ladenvall C., Getz G., Correa A., Ebert B.L. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med.* 2014;371(26):2488-2498. doi 10.1056/NEJMoa1408617
- Jaiswal S., Natarajan P., Silver A.J., Gibson C.J., Bick A.G., Shvartz E., McConkey M., ... Melander O., Sukhova G.K., Neuberg D., Libby P., Ebert B.L. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N Engl J Med.* 2017;377(2):111-121. doi 10.1056/NEJMoa1701719
- Jindal K., Adil M.T., Yamaguchi N., Yang X., Wang H.C., Kamimoto K., Rivera-Gonzalez G.C., Morris S.A. Single-cell lineage capture across genomic modalities with CellTag-multi reveals fate-specific gene regulatory changes. *Nat Biotechnol.* 2024;42(6):946-959. doi 10.1038/s41587-023-01931-4
- Khoury J.D., Solary E., Abla O., Akkari Y., Alaggio R., Apperley J.F., Bejar R., ... Wang W., Wood B., Xiao W., Yeung C., Hochhaus A. The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. *Leukemia.* 2022;36(7):1703-1719. doi 10.1038/s41375-022-01613-1
- Kilpivaara O., Mukherjee S., Schram A.M., Wadleigh M., Mullally A., Ebert B.L., Bass A., ... Offit K., Stone R.M., Gilliland D.G., Klein R.J., Levine R.L. A germline JAK2 SNP is associated with predisposition to the development of JAK2^{V617F}-positive myeloproliferative neoplasms. *Nat Genet.* 2009;41(4):455-459. doi 10.1038/ng.342
- Klein A.M., Simons B.D. Universal patterns of stem cell fate in cycling adult tissues. *Development.* 2011;138(15):3103-3111. doi 10.1242/dev.060103
- Kralovics R., Passamonti F., Buser A.S., Teo S.S., Tiedt R., Passweg J.R., Tichelli A., Cazzola M., Skoda R.C. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med.* 2005;352(17):1779-1790. doi 10.1056/NEJMoa051113
- Lin A., Brittan M., Baker A.H., Dimmeler S., Fisher E.A., Sluimer J.C., Misra A. Clonal expansion in cardiovascular pathology. *JACC Basic Transl Sci.* 2023;9(1):120-144. doi 10.1016/j.jacbs.2023.04.008
- Ljungström V., Mattisson J., Halvardson J., Pandzic T., Davies H., Rychlicka-Buniowska E., Danielsson M., Lacaze P., Cavelier L., Dumanski J.P., Baliakas P., Forsberg L.A. Loss of Y and clonal hematopoiesis in blood-two sides of the same coin? *Leukemia.* 2022;36(3):889-891. doi 10.1038/s41375-021-01456-2
- Lodato M.A., Rodin R.E., Bohrsen C.L., Coulter M.E., Barton A.R., Kwon M., Sherman M.A., ... Hatem N.E., Ryu S.C., Woodworth M.B., Park P.J., Walsh C.A. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science.* 2018;359(6375):555-559. doi 10.1126/science.aao4426
- Mack T.M., Raddatz M.A., Pershad Y., Nachun D.C., Taylor K.D., Guo X., Shuldiner A.R., ... Smith A.V., Reiner A.P., Jaiswal S., Weinstock J.S., Bick A.G. Epigenetic and proteomic signatures associate with clonal hematopoiesis expansion rate. *Nat Aging.* 2024;4(8):1043-1052. doi 10.1038/s43587-024-00647-7
- McKenna A., Gagnon J.A. Recording development with single cell dynamic lineage tracing. *Development.* 2019;146(12):dev169730. doi 10.1242/dev.169730
- Moran-Crusio K., Reavie L., Shih A., Abdel-Wahab O., Ndiaye-Lobry D., Lobry C., Figueroa M.E., ... Nimer S.D., Melnick A., Godley L.A., Aifantis I., Levine R.L. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell.* 2011;20(1):11-24. doi 10.1016/j.ccr.2011.06.001
- Nachun D., Lu A.T., Bick A.G., Natarajan P., Weinstock J., Szeto M.D., Kathiresan S., ... Whitsel E.A., Wilson J.G., Horvath S., Jaiswal S.; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium. Clonal hematopoiesis associated with epigenetic aging and clinical outcomes. *Aging Cell.* 2021;20(6):e13366. doi 10.1111/accel.13366
- Nagase R., Inoue D., Pastore A., Fujino T., Hou H.A., Yamasaki N., Goyama S., ... Takeda R., Tien H.F., Honda H., Abdel-Wahab O., Kitamura T. Expression of mutant Asxl1 perturbs hematopoiesis

- and promotes susceptibility to leukemic transformation. *J Exp Med.* 2018;215(6):1729-1747. doi 10.1084/jem.20171151
- Puzyrev V.P., Nazarenko M.S., Lebedev I.N., Markov A.V., Sleptsov A.A., Kashevarova A.A., Tolmacheva E.N., Frolov A.V., Popov V.A., Barbarash O.L., Barbarash L.S. Phenomenon of para-dominant inheritance in atherosclerosis. *Medical Genetics.* 2014; 13(10):10-17 (in Russian)
- Sano S., Horitani K., Ogawa H., Halvardson J., Chavkin N.W., Wang Y., Sano M., ... Arai Y., Mychaleckyj J.C., Hirschi K.K., Forsberg L.A., Walsh K. Hematopoietic loss of Y chromosome leads to cardiac fibrosis and heart failure mortality. *Science.* 2022;377(6603):292-297. doi 10.1126/science.abn3100
- Schuermans A., Honigberg M.C. Clonal haematopoiesis in cardiovascular disease: prognostic role and novel therapeutic target. *Nat Rev Cardiol.* 2025;22(11):845-856. doi 10.1038/s41569-025-01148-9
- Shao D.D., Kriz A.J., Snellings D.A., Zhou Z., Zhao Y., Enyenihi L., Walsh C. Advances in single-cell DNA sequencing enable insights into human somatic mosaicism. *Nat Rev Genet.* 2025;26(11):761-774. doi 10.1038/s41576-025-00832-3
- Steensma D.P., Bejar R., Jaiswal S., Lindsley R.C., Sekeres M.A., Has-serjian R.P., Ebert B.L. Clonal hematopoiesis of indeterminate po-tential and its distinction from myelodysplastic syndromes. *Blood.* 2015;126(1):9-16. doi 10.1182/blood-2015-03-631747
- Walsh K., Raghavachari N., Kerr C., Bick A.G., Cummings S.R., Druley T., Dunbar C.E., ... Natarajan P., Shindyapina A.V., Shuldiner A.R., Van Den Akker E.B., Vijg J. Clonal hematopoiesis analyses in clinical, epidemiologic, and genetic aging studies to unravel underlying mechanisms of age-related dysfunction in humans. *Front Aging.* 2022;3:841796. doi 10.3389/fragi.2022.841796
- Xie Z., Zeidan A.M. CHIPing away the progression potential of CHIP: a new reality in the making. *Blood Rev.* 2023;58:101001. doi 10.1016/j.blre.2022.101001
- Yamamoto K., Goyama S., Asada S., Fujino T., Yonezawa T., Sato N., Takeda R., ... Nakagawa S., Hirose T., Ogawa S., Akimitsu N., Kitamura T. A histone modifier, ASXL1, interacts with NONO and is involved in paraspeckle formation in hematopoietic cells. *Cell Rep.* 2021;36(8):109576. doi 10.1016/j.celrep.2021.109576
- Yang H., Kurtenbach S., Guo Y., Lohse I., Durante M.A., Li J., Li Z., ... Nimer S.D., Harbour J.W., Wahlestedt C., Xu M., Yang F.C. Gain of function of ASXL1 truncating protein in the pathogenesis of myeloid malignancies. *Blood.* 2018;131(3):328-341. doi 10.1182/blood-2017-06-789669

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

Received February 26, 2025. Revised March 16, 2026. Accepted March 16, 2026.

doi 10.18699/vjgb-26-69

Mobile genetic elements in the pathogenesis of human reproductive disorders

A.A. Babovskaya , E.A. Trifonova , V.A. Stepanov 

Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia

 anastasia.babovskaya@medgenetics.ru

Abstract. Mobile genetic elements (MGEs), or transposons, are autonomous DNA sequences capable of moving and proliferating within the genome. Long considered “selfish” or “junk” DNA, MGEs are now recognized as key components involved in genome evolution, the regulation of gene expression, and the pathogenesis of various diseases. In humans, MGEs are divided into two main classes: retrotransposons (Class I), which replicate via an RNA intermediate through a “copy-and-paste” mechanism, and DNA transposons (Class II), which move via a “cut-and-paste” mechanism without an RNA intermediate. According to the Human Genome Project, retrotransposons constitute the majority (approximately 42 %) of the MGE fraction within the human genome. The most abundant are the non-long terminal repeat (non-LTR) retrotransposons, dominated by autonomous LINE-1 elements. Although approximately 500,000 LINE-1 copies are present in the genome, the vast majority are defective, and only a small fraction (<100) retain the capacity for transposition in modern humans. The second most prevalent group (about 10.6 %) within the retrotransposon family is the short interspersed nuclear elements (SINEs), specifically Alu elements, which are non-autonomous and hijack the LINE-1 molecular machinery for their mobilization and integration. MGE activity is a tightly regulated process in somatic tissues. Epigenetic mechanisms, particularly DNA methylation, normally effectively suppress MGE expression and mobility. Disruption of this control is associated with a wide range of pathologies. For instance, hypomethylation and reactivation of retrotransposons, notably LINE-1, have been demonstrated in various cancers, as well as in neurodegenerative and autoimmune diseases. The aim of this review is to provide a systematic analysis of the current understanding of the role of mobile genetic elements, particularly LINE-1, Alu, and HERV retrotransposons, in the development of human reproductive system disorders. This also includes diseases associated with placental pathology, an area that remains insufficiently studied to date, despite a growing body of data.

Key words: mobile genetic elements; transposons; reproduction; placenta; human diseases

For citation: Babovskaya A.A., Trifonova E.A., Stepanov V.A. Mobile genetic elements in the pathogenesis of human reproductive disorders. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov J Genet Breed.* 2026;30(4):685-695. doi 10.18699/vjgb-26-69

Funding. This research was funded by a state grant (Basic Scientific Research No. 122020200083-8).

Мобильные генетические элементы в патогенезе репродуктивных нарушений человека

A.A. Бабовская , E.A. Трифонова , В.А. Степанов 

Научно-исследовательский институт медицинской генетики Томского национального исследовательского медицинского центра, Томск

 anastasia.babovskaya@medgenetics.ru

Аннотация. Мобильные генетические элементы (МГЭ), или транспозоны, представляют собой автономные последовательности ДНК, способные к перемещению и распространению в пределах генома. Долгое время рассматривавшиеся как «эгоистичная» или «мусорная» ДНК в настоящее время МГЭ считаются одними из компонентов, принимающих участие в эволюции генома, регуляции экспрессии генов и патогенезе ряда заболеваний. У человека МГЭ подразделяются на два основных класса: ретротранспозоны (I класс), реплицирующиеся через промежуточную стадию РНК по механизму «копирования–вставки» (copy-and-paste), и ДНК-транспозоны (II класс), перемещающиеся по механизму «вырезания–вставки» (cut-and-paste) без РНК-интермедиата. Ретротранспозоны, по данным проекта «Геном человека», составляют примерно 42 % от доли МГЭ в человеческом геноме. Наиболее многочисленны ретротранспозоны без длинных концевых повторов (non-LTR), среди которых доминируют автономные элементы LINE-1. Несмотря на то что в геноме присутствует около 500 000 копий LINE-1, большинство из них дефектно и лишь малая часть (<100) сохраняет возможность к транспозиции у современного человека. Второй по распространенности группой (около 10.6 %) семейства ретротранспозонов являются короткие диспергированные повторы Alu (SINE-элементы),

которые, будучи неавтономными, используют для перемещения и встраивания молекулярный аппарат LINE-1. Активность МГЭ – строго регулируемый процесс в соматических тканях. Эпигенетические механизмы, в частности метилирование ДНК, в норме эффективно подавляют экспрессию и мобильность мобильных генетических элементов. Нарушение этого контроля ассоциировано с широким спектром патологий. Так, гипометилирование и реактивация ретротранспозонов, в особенности LINE-1, продемонстрированы при различных типах рака, нейродегенеративных и аутоиммунных заболеваниях. Цель настоящего обзора – систематический анализ современных представлений о роли мобильных генетических элементов, в частности ретротранспозонов LINE-1, Alu и HERV, в развитии заболеваний репродуктивной системы человека, также включая заболевания, связанные с патологией плаценты, область которых остается недостаточно изученной, несмотря на растущий объем данных.

Ключевые слова: мобильные элементы генома; транспозоны; репродукция; плацента; заболевания человека

Introduction

Contemporary research on mobile genetic elements (MGEs) and their role in disease development represents a rapidly advancing field of molecular biology and genetics. In humans, MGEs are classified into two major classes: retrotransposons, which replicate via an mRNA intermediate through a “copy-and-paste” mechanism, and DNA transposons, which utilize a DNA segment as an intermediate during transposition. The most abundant group in the human genome consists of retrotransposons, which are further subdivided into elements containing long terminal repeats (LTRs) and those lacking LTRs (non-LTR elements). The latter predominate, accounting for approximately 75.2 % of all mobile elements and about 33.6 % of the entire genome (Lander et al., 2001; Chénais et al., 2012). Among non-LTR retrotransposons, long interspersed nuclear elements-1 (LINE-1; Fig. 1) have undergone extensive expansion, reaching approximately 500,000 copies in the human genome, the majority of which are represented by defective or fragmented sequences. The LINE-1 family accounts for about 38 % of all MGEs and 16.9 % of the genome (Fig. 1). However, only a small fraction of LINE-1 copies retain transpositional activity, that is, the capacity for autonomous retrotransposition via reverse transcription of their mRNA and integration of the resulting cDNA into new genomic loci.

The second most abundant category of MGEs comprises non-autonomous short interspersed nuclear elements (SINEs), particularly Alu repeats, which constitute approximately 24 % of all MGEs and 10.6 % of the genome. The human reference genome contains around one million copies of Alu elements derived from several evolutionarily young subfamilies (e. g., AluY). Their extensive amplification is associated with a burst of retrotranspositional activity in ancestral primates approximately 35–40 million years ago. For their retrotransposition, Alu elements rely on the enzymatic machinery of LINE-1 elements, including endonuclease and reverse transcriptase. Other representatives of SINEs in humans include MIR repeats (Mammalian-wide Interspersed Repeats). In addition, composite SVA elements composed of homologous SINE regions, variable number tandem repeats (VNTRs), and Alu-like sequences also belong to the group of non-autonomous elements mobilized by LINE-1. Only certain

subfamilies are considered transpositionally active in the modern human genome, including Alu Ya5, Alu Yb8, Alu Yb9, SVA-E, and SVA-F. LINE-1 activity in the germ line leads to the emergence of new LINE-1, Alu, and SVA insertions, which serve as population genetic markers (Khitrinskaya et al., 2003, 2014).

LTR retrotransposons are present in smaller numbers in the human genome and are mainly represented by human endogenous retroviruses (HERVs). In terms of their structure and mechanism of retrotransposition, HERVs are similar to retroviruses but have lost functional envelope (*env*) genes. The activity of LTR retrotransposons in humans has markedly declined over the past several million years; nevertheless, their sequences still constitute approximately 8 % of the genome (Fig. 1). The classification of HERVs is based on the type of tRNA used to initiate reverse transcription; the evolutionarily youngest and most active subfamily is HERV-K (which uses lysine tRNA). Other groups include HERV-I (isoleucine tRNA) and HERV-L (leucine tRNA). Class II elements (DNA transposons) are considerably less common in the human genome, accounting for only about 6 % of all MGEs and 2.8 % of the genome (Fig. 1). The most represented are three superfamilies: TC1/mariner (including mariner, MER2-Tigger, and Tc2), hAT (including MER1-Charlie and Zaphod), and PiggyBac (Lander et al., 2001; Brouha et al., 2003).

In recent years, the implementation of whole-genome and transcriptome analysis methods, as well as single-cell sequencing technologies, has substantially expanded our understanding of the role of MGEs in the regulation of gene expression. These elements have been reported to participate in a wide range of biological processes, both under normal physiological conditions and in pathological states. The investigation of the molecular mechanisms through which MGEs modulate gene activity and contribute to the pathogenesis of various diseases has therefore emerged as an important new direction in biomedical research. In particular, the role of MGEs has been examined in the development of oncological diseases (Chénais et al., 2015; Kong et al., 2019; McKerrow et al., 2022), atherosclerosis (Hueso et al., 2018), infectious diseases (Chen et al., 2023), and neurodegenerative disorders (Ravel-Godreuil et al., 2021; Roy et al., 2024; Zhang W.

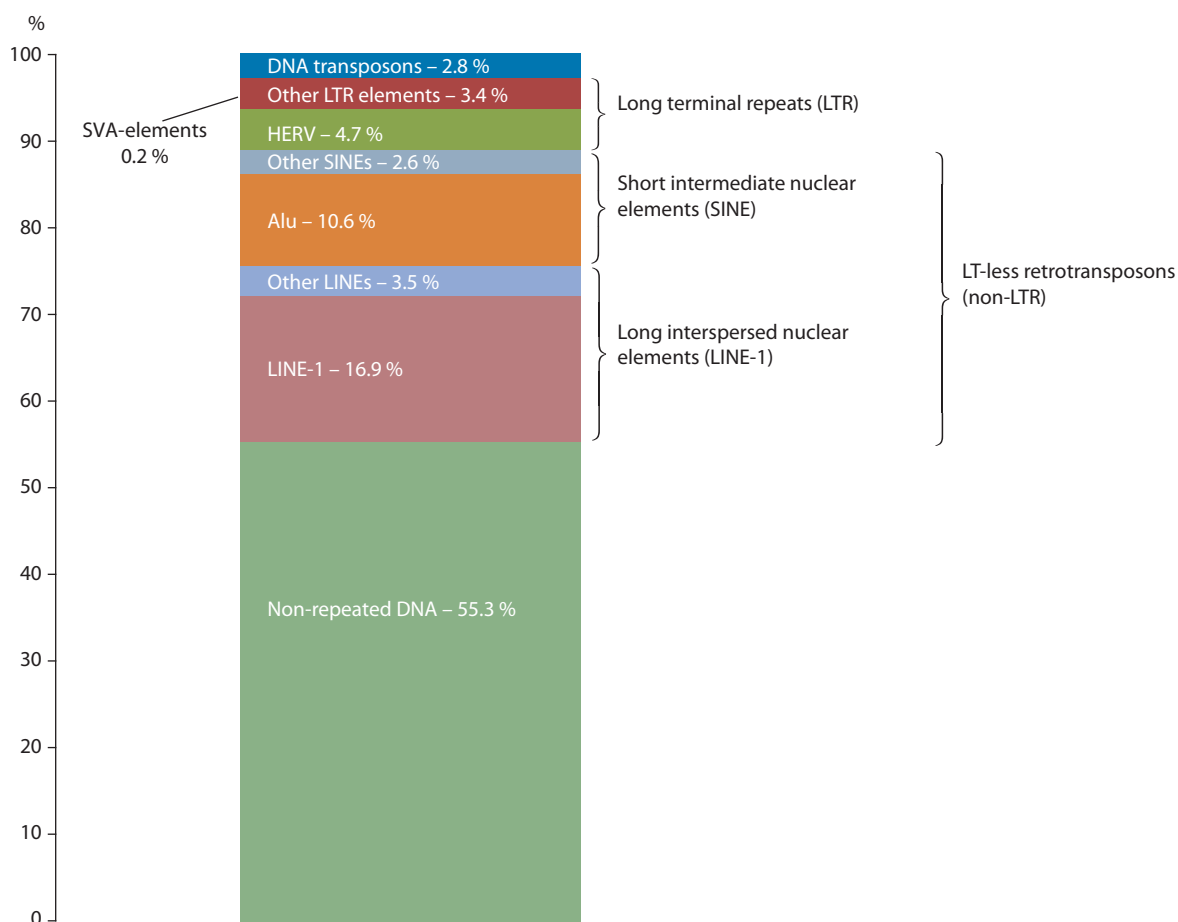


Fig. 1. Proportions of mobile genetic elements (MGEs) in the human genome and their distribution among different families and subfamilies according to (Lander et al., 2001).

et al., 2025). Retrotransposons can influence gene expression both directly and indirectly. Direct effects occur when these elements integrate into genes and disrupt the reading frame, leading to premature termination of protein synthesis. Indirect effects include the provision of novel promoters, splice sites, and transcription factor binding sites. In addition, MGEs may serve as sources of enhancer and promoter sequences, influence the three-dimensional architecture of chromatin, and contribute to the emergence of new regulatory genes, including non-coding RNAs and transcription factors (Bourque et al., 2018).

Accumulating evidence suggests that MGEs have acted as key drivers of evolutionary changes underlying the emergence of placental pregnancy in mammals (Lynch et al., 2015). In this context, it is of particular interest to extend the analysis of the literature beyond somatic structures of the reproductive system to include the placenta – a unique transient organ that forms during pregnancy. Studies of the placental methylome have shown that certain repetitive elements, such as LINE-1 and human endogenous retroviruses (HERVs), function as key regulators of gene expression in placental tissue. Their methylation in the placenta is believed to regulate placenta-

specific functions. For example, some retrotransposons act as alternative promoters for placenta-specific transcripts, including *KCNH5* and *IL2RB*. The expression of an alternative form of *KCNH5* may contribute to the process of trophoblast invasion.

Several studies have also investigated the activity of the LINE-1 retrotransposon during the first trimester of pregnancy (Vasilyev et al., 2021; Demeneva et al., 2023). However, despite significant advances in understanding the role of MGEs in various pathological conditions, their contribution to the development of diseases of the reproductive system remains insufficiently explored to date.

Overview of methods for studying the impact of mobile genetic elements on the reproductive system

The reproductive system represents a unique target for the activity of mobile genetic elements due to the intensive processes of cell division, recombination, and epigenetic reprogramming that occur within it. Investigating their influence requires an integrated methodological approach combining classical genetic techniques with modern high-throughput sequen-

cing technologies and bioinformatic analysis. Contemporary studies examining the impact of mobile genetic elements on reproductive function can be broadly divided into three key methodological approaches, each characterized by specific advantages and limitations.

Genomic and transcriptomic approaches enable the precise identification of MGE insertions and the assessment of their activity. For example, whole-genome analysis has revealed 88 tumor-specific LINE-1 insertions in high-grade serous ovarian carcinomas, which were associated with poorer patient survival (Nguyen et al., 2018).

Transcriptomic analysis (RNA-Seq), combined with specialized bioinformatic algorithms such as TETranscripts, RE-discoverTE, SalmonTE, ExplorATE, and SQuIRE, makes it possible to quantitatively assess MGE expression and identify its association with the development of pathological conditions (Grow et al., 2015; Bourque et al., 2018).

Epigenetic approaches focus on the mechanisms controlling MGE activity, primarily DNA methylation and histone modifications. Using bisulfite sequencing, it has been demonstrated that during spermatogenesis in mice, extensive demethylation followed by remethylation of LINE-1 promoters occurs, creating a window of vulnerability for retrotransposition (Molaro et al., 2011). In humans, age-related decreases in LINE-1 methylation in sperm have been shown to correlate with reduced sperm quality (Jenkins et al., 2014). ChIP-seq studies in mouse models have confirmed the critical role of the PIWI pathway, demonstrating that piRNA-associated proteins (e. g., MIWI2) direct the establishment of the repressive H3K9me3 mark at MGE loci in germ cells. Knockout of the corresponding genes leads to widespread reactivation of transposons, disruption of meiosis, and sterility (Di Giacomo et al., 2013).

Functional studies *in vitro* and *in vivo* aim to establish causal relationships between genetic alterations resulting from the insertion of mobile genetic elements and the subsequent development of phenotypic traits. Cellular models, such as human embryonic stem cells (ESCs) equipped with reporter systems, allow real-time monitoring of LINE-1 retrotransposition and the assessment of the effects of various factors on this process (Macia et al., 2017). An example of functional *in vivo* validation is provided by mouse models with knockout of genes involved in the PIWI pathway (e. g., *Mov10LINE-1*), in which derepression of MGEs, arrest of spermatogenesis, and complete sterility have been observed (Zheng, Wang, 2012).

However, despite significant progress, the study of MGEs remains associated with several methodological challenges. In particular, the accurate analysis of repetitive sequences using short-read sequencing technologies continues to present difficulties, although improved algorithms are being developed to address this issue and reduce the occurrence of false-positive results (Gardner et al., 2017).

Molecular mechanisms of the pathogenic impact of mobile genetic elements on the human genome

Mobile genetic elements exert multifaceted effects on the structure and function of the genome. The pathogenic potential of MGEs is realized through a variety of molecular mechanisms that disrupt genomic stability, normal gene expression, and epigenetic regulation. These mechanisms can be broadly categorized into three main groups: insertional mutagenesis, induction of genomic instability, and disruption of the epigenetic landscape (Chénais et al., 2015; Goodier et al., 2016).

Insertional mutagenesis. The most common mechanism of pathogenesis is insertional mutagenesis – the integration of a new MGE copy into a functionally significant region of the genome. More than 120 cases of hereditary diseases caused by *de novo* or inherited pathogenic MGE insertions have been described (Callinan, Batzer, 2006; Hancks, Kazazian, 2016). Inactivation of the target gene generally occurs via two principal mechanisms: direct disruption of the coding sequence or interference with RNA processing. In the first case, insertion of an MGE into an exon disrupts the open reading frame, often resulting in the formation of a premature termination codon and subsequent degradation of the transcript through nonsense-mediated mRNA decay (NMD) (Miki et al., 1992). In the second case, insertion into intronic or exonic regions may create cryptic splice sites, leading to aberrant pre-mRNA splicing, inclusion of pseudoexons, or exon skipping, and consequently to the synthesis of defective protein isoforms. A striking example of this mechanism is the insertion of an SVA element into the 3' untranslated region (3'-UTR) of the *FKTN* gene, which is associated with Fukuyama muscular dystrophy. This insertion does not cause complete knockout inactivation of the gene but instead induces alternative splicing, resulting in the synthesis of a C-terminally truncated protein. This truncated protein accumulates in the Golgi apparatus and the endoplasmic reticulum, disrupting post-translational modification processes and thereby contributing to disease development (Taniguchi-Ikeda et al., 2011). In addition, the process of retrotransposition may be accompanied by deletion at the integration site. The insertion of retrotransposon cDNA (LINE-1, Alu, SVA) is associated with the formation of double-strand DNA breaks, which can lead to the loss of sequences ranging from several base pairs to several megabases. Such *de novo* deletions associated with MGE insertions have been identified in human genomes and may directly eliminate functionally important genetic elements (Gilbert et al., 2002; van den Hurk et al., 2007).

The induction of genomic instability and chromosomal rearrangements. The second mechanism underlying the pathogenicity of MGEs is the induction of genomic instability and chromosomal rearrangements. The high degree of sequence homology among numerous copies of MGEs (e. g.,

Alu repeats) creates a substrate for non-allelic homologous recombination. Incorrect pairing of homologous sequences located at different genomic loci during meiosis can lead to deletions, duplications, inversions, or translocations (Sen et al., 2006). This mechanism underlies several microdeletion syndromes, such as hereditary neuropathy with liability to pressure palsies (HNPP), which is caused by reciprocal recombination between repeated sequences in the 17p11.2–12 region of chromosome 17, encompassing the gene encoding peripheral myelin protein 22 (PMP22) (Polynnikova et al., 2021).

Disruption of epigenetic regulation represents the third mechanism of the pathogenic effects of MGEs. This mechanism is associated with the ability of MGEs to alter the normal DNA methylation status of genomic regions. Since the transcriptional activity of MGEs poses a potential threat to cellular integrity, eukaryotic cells have evolved mechanisms to repress them, primarily through methylation of CpG dinucleotides within their promoter regions (Figueiredo et al., 2009). It has been reported that the insertion of MGEs may also influence adjacent DNA regions. Highly methylated promoters of integrated MGEs (such as LINE-1 or SVA) can serve as targets for the recruitment of DNA methyltransferases and histone-modifying complexes, leading to the spread of heterochromatin into nearby regulatory regions of genes and resulting in their transcriptional silencing (Slotkin, Martienssen, 2007).

Why is the reproductive system particularly vulnerable?

The human reproductive system exhibits increased susceptibility to dysregulation of mobile genetic elements due to its unique biological features, including extensive epigenetic reprogramming, evolutionary co-option of MGEs, and stringent requirements for genomic stability (Smith et al., 2012).

A key factor underlying this vulnerability is the large-scale epigenetic reprogramming that occurs during gametogenesis and early embryogenesis. This process involves global DNA demethylation, which is necessary to erase parental epigenetic marks and establish totipotency (Smith et al., 2012). However, the transient removal of repressive methylation marks, particularly at MGE promoter regions such as LINE-1 and Alu, creates a temporary window of vulnerability. During this period, the transcriptional potential of MGEs is reactivated, substantially increasing the risk of *de novo* insertional mutagenesis in genes critical for gamete and embryonic development (van den Hurk et al., 2007; Grow et al., 2015). In this way, the very biological program that ensures totipotency and developmental progression also elevates genomic instability.

It should be noted that several studies indicate that a certain level of MGE activity is physiologically necessary for normal reproductive function. Throughout evolution, MGE sequences have been co-opted into regulatory networks controlling embryogenesis and placentation (Lynch et al., 2015). For instance, the expression of specific retrotransposon families – MERVL

in mice and HERVH in humans – is required for activation of the totipotency genetic program at the 2-cell embryonic stage, functioning as alternative enhancers and promoters (Macfarlan et al., 2012; Grow et al., 2015). Furthermore, evidence suggests that some proteins mediating cytotrophoblast fusion and the formation of the syncytiotrophoblast in the placenta are of retroviral origin. For example, syncytin genes, originally derived from the envelope (env) genes of endogenous retroviruses (HERV-W, HERV-FRD), are essential for proper placental development (Mi et al., 2000; Blaise et al., 2003).

At the same time, disruption of the epigenetic control mechanisms regulating MGEs can lead to uncontrolled transcription and retrotransposition, resulting in damage to genes essential for reproduction (Vasilyev et al., 2021). Such dysregulation is thought to be associated with a broad spectrum of reproductive pathologies, including idiopathic infertility, implantation failure, pregnancy loss, and complications related to placental dysfunction, such as preeclampsia and fetal growth restriction. The placenta is particularly vulnerable because, unlike somatic tissues, it maintains global hypomethylation of the genome, including LINE-1 promoters (He, Ecker, 2014). Under normal conditions, this hypomethylated state supports the unique functions of the trophoblast; however, aberrant hyperactivation of retrotransposons can impair trophoblast invasiveness and spiral artery remodeling, directly contributing to the pathogenesis of preeclampsia and fetal growth restriction (Vasiliev et al., 2015).

MGEs in the pathogenesis of gametogenesis disorders and infertility

Dysregulation of mobile genetic elements, particularly LINE-1 retrotransposons, represents a key molecular mechanism associated with impaired fertility and gametogenic pathology. The pathogenic effects are mediated through disruption of epigenetic control, induction of genomic instability, and aberrant regulation of genes critical for reproduction.

Epigenetic dysregulation of MGEs in germ cells. Unlike somatic cells, where LINE-1 promoters are constitutively hypermethylated, the epigenetic control of these elements in germ cells is highly dynamic and tissue-specific. One major risk factor for impaired gamete maturation is age-dependent alteration of LINE-1 methylation status. In spermatozoa, LINE-1 promoter methylation increases with paternal age (Jenkins et al., 2014), whereas in human oocytes at the diplotene stage and in ovulated secondary oocytes, LINE-1 elements exhibit relative hypomethylation (Smith et al., 2012). Such epigenetic imbalance creates conditions for reactivation of MGE transpositional activity, potentially leading to insertional mutagenesis in key genes that regulate meiosis, gamete differentiation, and early embryonic development. Studies in mice have shown that spermatozoa from individuals with idiopathic infertility exhibit significant hypomethylation of

LINE-1 promoters compared to fertile controls (Jachowicz et al., 2017). This epigenetic disruption is accompanied by elevated transcriptional and retrotranspositional activity of LINE-1. Notably, excessive LINE-1 expression correlates with increased sperm DNA fragmentation, indicating a direct link between retrotransposon derepression and genomic damage in male gametes. LINE-1 hypomethylation thus serves as an independent prognostic marker of impaired spermatogenesis and reduced ejaculate quality (Jenkins et al., 2014).

Insertional mutagenesis in genes critical for spermatogenesis. The pathogenic impact of mobile genetic elements on male reproductive function is primarily mediated through insertional mutagenesis. It has been demonstrated that *de novo* LINE-1 retrotransposon insertions can disrupt the integrity of genes essential for normal spermatogenesis. Of particular significance are lesions within the AZF (Azoospermia Factor) locus of the Y chromosome, which contains a cluster of genes directly responsible for sperm production. LINE-1 insertions in DAZ (Deleted in Azoospermia) gene family members located within the AZF region have been identified as a direct cause of gene inactivation, leading to impaired spermatogenesis and the development of azoospermia (Hancks, Kazazian, 2016).

Beyond effects on Y-chromosome genes, retrotranspositional activity can also impact critical autosomal loci. For instance, LINE-1 insertions in the *FKBP6* gene, which encodes a protein involved in meiotic division, have been reported. Mutations in *FKBP6* are associated with defects in chromosome pairing and synaptonemal complex formation during prophase I of meiosis, ultimately resulting in male infertility (Wyrwoll et al., 2022).

Disruption of early embryogenesis and implantation. The second wave of extensive epigenetic reprogramming, which begins after fertilization and continues until the blastocyst stage, is accompanied by passive genome-wide demethylation and a marked increase in the transcriptional potential of LINE-1 elements. Under normal conditions, this process is tightly regulated by RNA interference mechanisms and is essential for chromatin remodeling and activation of the embryonic genome (Jachowicz et al., 2017). However, disruption of the delicate balance between MGE activation and repression can lead to pathological outcomes. It has been reported that artificial suppression of LINE-1 transcription in mouse zygotes impairs cleavage and arrests development at early stages (Percharde et al., 2018). This is because LINE-1 sequences act as alternative promoters for genes critical for cell division (e. g., *TP53*) and participate in the regulation of higher-order chromatin organization (Chow et al., 2010).

Pregnancy loss associated with chromosomal abnormalities

An important aspect of reproductive pathology is the interplay between MGE activity and chromosomal instability. Aneuploidy, in turn, can induce global changes in the methylome.

It has been shown that in placental tissues with mosaic forms of aneuploidy, LINE-1 methylation levels are significantly increased (Vasiliev et al., 2015). This may reflect a compensatory mechanism, whereby promoter hypermethylation reduces transcriptional load on the genome under conditions of disrupted gene dosage balance.

Moreover, MGE activation in cases of pregnancy loss may also be secondary, serving as a marker of general cellular stress. Oxidative stress, which is characteristic of placentas with chromosomal abnormalities, induces global DNA demethylation, relieving repression of LINE-1 and Alu promoters (Lou et al., 2020). Subsequent retrotransposition and accumulation of double-strand DNA breaks further destabilize the trophoblast genome, reinforcing the pathological state and rendering continued pregnancy development unviable (Garcia-Perez et al., 2016). Thus, MGE dysregulation acts not only as a trigger but also as a universal amplifier of pathological processes, contributing to reproductive failure at the earliest stages of gestation.

The contribution of MGEs to the pathogenesis of preeclampsia

Previous studies have shown that hundreds of genes in the endometrium gain or lose expression coinciding with the onset of pregnancy and decidualization in mammals (Lynch et al., 2015; Mika et al., 2021). To prepare the uterus for embryo implantation, the endometrium undergoes decidualization, during which stromal cells transform into decidual stromal cells (DSCs). The decidual layer formed as a result of this process supports embryo implantation and protects the fetus from maternal immune rejection. Decidualization is regulated by progesterone through its receptor (PGR), the secondary messenger cAMP, protein kinase A (PKA), and the transcription factor FOXO1 (Kajihara et al., 2013). This evolutionarily acquired process in eutherian mammals is accompanied by extensive changes in gene regulatory activity, cellular organization, and endometrial physiology, all of which are essential for successful implantation and pregnancy (Mess, Carter, 2006; Kin et al., 2014).

MGEs are believed to have contributed substantially to the development and function of DSCs during this adaptive process. Two major waves of transposon activity and genomic integration have remodeled the transcriptome and regulatory landscape of DSCs. Current evidence indicates that genes located near regulatory elements derived from transposons exhibit high progesterone sensitivity and appear to play key roles in mediating the hormone's signaling response, providing binding sites for progesterone receptors (Mika, Lynch, 2022). Defective decidualization during early pregnancy can lead to fetal growth restriction, pregnancy loss, or obstetric complications such as preeclampsia (PE).

Preeclampsia is one of the most severe pregnancy complications, characterized by defective placentation, endothelial dysfunction, and a systemic inflammatory response. The cen-

tral pathogenic hypothesis of PE is incomplete trophoblast invasion of the uterine spiral arteries, resulting in placental hypoxia. The role of MGEs – such as LINE-1, Alu, and SVA elements – in regulating gene expression and epigenetic mechanisms in decidual cells remains insufficiently explored.

Preeclampsia is accompanied by systemic inflammation, in which decidual cells play a pivotal role through the secretion of pro-inflammatory cytokines. Studies have shown that PSG1, PSG6, and PSG11 increase secretion of anti-inflammatory cytokines IL-10 and IL-6, while all PSG genes activate the TGF β receptor in immune cells (Warren et al., 2018). MGE activation, particularly LINE-1, may induce inflammatory pathways via the production of double-stranded RNA (dsRNA), which is recognized by innate immune receptors such as TLR3. Similar mechanisms are observed in autoimmune diseases, where transposons trigger an interferon response (Crow, 2014). However, direct evidence supporting this mechanism in PE pathogenesis within decidual cells remains limited and requires further investigation. Integration of transposons into the human genome has also influenced the regulation of key placental hormones, including leptin (LEP), insulin-like protein 4 (INSL4), and corticotropin-releasing hormone (CRH), highlighting the significant role of MGEs in modulating processes critical for pregnancy development.

Furthermore, some studies suggest that hyperactivation of MGEs can directly impair trophoblast differentiation and function. Successful placentation requires cytotrophoblast differentiation along two main pathways: into syncytiotrophoblast and invasive extravillous trophoblast. Disruption of this finely tuned differentiation process, resulting in incomplete extravillous trophoblast invasion and failure to remodel spiral arteries, is directly associated with PE development. LINE-1 and other retrotransposons have been shown to interfere with key signaling pathways controlling trophoblast differentiation, which may underlie the placentation defects central to the pathogenesis of preeclampsia (Demeneva et al., 2024).

Mobile genetic elements and their role in the pathogenesis of fetal growth restriction

Fetal growth restriction (FGR), also referred to as intrauterine growth restriction, is defined as the birth of a fetus with a body weight and/or length at or below the 10th percentile, or ≥ 2 standard deviations below the mean for a given gestational age (Li et al., 2019). FGR is associated with an increased risk of perinatal mortality, neurological deficits, delayed psychomotor development, metabolic syndrome, and cardiovascular diseases in adulthood (Entringer et al., 2012). The pathogenesis of FGR is multifactorial, involving maternal, placental, and fetal determinants, among which dysregulation of mobile genetic elements plays a significant role.

LINE-1 elements are the most abundant autonomous retrotransposons in the human genome. Under normal conditions,

LINE-1 sequences are deeply methylated and transcriptionally silenced in somatic tissues of adults (Figueiredo et al., 2009). However, under conditions of epigenetic instability – such as methyl donor deficiency – promoter hypomethylation can occur, leading to LINE-1 activation, transcription, and retrotransposition (Howard et al., 2008). Folic acid, as a key methyl group donor, plays a central role in maintaining global DNA methylation. *In vitro* studies have shown that culturing mouse embryonic stem cells (mESCs) under folate-deficient conditions induces intracellular depletion of 5-methyltetrahydrofolate, reduces levels of S-adenosylmethionine (SAM), and subsequently leads to LINE-1 hypomethylation (Chang et al., 2013). Similar observations have been made *in vivo*: pregnant women carrying fetuses with neural tube defects (NTDs) exhibited significantly reduced LINE-1 methylation, which correlated with low maternal serum folate levels (Li et al., 2019).

Mouse model studies have demonstrated that LINE-1 hypomethylation and activation in the placenta correlate with reduced fetal mass and disrupted placental architecture, including thinning of the syncytiotrophoblast layer and decreased vascularization. In human cohort studies, LINE-1 methylation levels in cord blood and placental tissue of infants with FGR were significantly lower than in infants with normal birth weight. Moreover, mothers with low folate levels during the first trimester exhibited an increased risk of giving birth to infants with FGR, accompanied by reduced LINE-1 methylation in the placenta (Joubert et al., 2016; Li et al., 2019).

Oncogenic role of mobile genetic elements

A review of experimental studies highlights the significant role of mobile genetic elements, particularly LINE-1 retrotransposons, in the pathogenesis of malignancies of the reproductive system. In epithelial ovarian cancer, team of S. Sato (Sato et al., 2023) demonstrated that the LINE-1 ORF1p protein is not only expressed but also secreted by tumor cells, detectable in patient plasma and ascitic fluid, suggesting its potential as a biomarker. Earlier, a whole-genome analysis by the group of T.H.M. Nguyen (Nguyen et al., 2018) identified 88 tumor-specific LINE-1 retrotransposition events in high-grade serous ovarian carcinomas, which correlated with poorer patient survival. *In vitro* experiments showed that inhibiting LINE-1 expression reduced cell invasion and migration while increasing apoptosis (Fu et al., 2023). In endometrial cancer, significant hypomethylation of LINE-1 promoters was observed in tumor tissues compared to normal endometrium, with the most pronounced hypomethylation in papillary serous carcinomas. This epigenetic derepression led to increased LINE-1 transcription, particularly in p53-mutated tumors with high levels of copy number variation (CNV), suggesting a synergistic effect between p53 mutation and MGE activation (Zhang B. et al., 2014). The research group of W. McKerrow further

confirmed that LINE-1 overexpression can induce phosphorylation of RAD50, a key step in the DNA double-strand break response. Statistical analysis revealed a strong correlation between LINE-1 activity, chromosomal aberrations, and replication stress pathway activation (McKerrow et al, 2022). In cervical cancer, RNA-seq analyses of invasive carcinoma samples showed that the highest number of differentially expressed retroelements was associated with co-infection by multiple HPV types, indicating a potential synergistic effect in disease progression (Curty et al., 2021). In prostate cancer, systemic analysis of whole-genome sequencing data revealed significantly higher somatic retrotransposition in metastases compared to primary tumors. Notably, 68 % of these insertions localized to exons, introns, or regulatory regions of genes, including tumor suppressors, directly implicating MGE-mediated inactivation of protective pathways and enhanced genomic instability (Tubio et al., 2014).

These studies, although compelling, represent only a fraction of an actively expanding research field. Further investigations are required to fully elucidate the regulatory networks controlling MGE activity in different tissues and cell types, as well as to evaluate the contribution of other retrotransposon families to oncogenesis.

Conclusion

This review underscores the fundamental role of mobile genetic elements in regulating physiological processes of the reproductive system and their involvement in a broad spectrum of pathological conditions. Among various MGE families, LINE-1 retrotransposons have been studied most extensively. Epigenetic dysregulation, particularly promoter hypomethylation during physiological reprogramming or in response to cellular stress, is a primary trigger for MGE-mediated pathogenic effects. This dysregulation activates LINE-1 transcription and retrotransposition, initiating a cascade of events that includes insertional mutagenesis, induction of genomic instability, and disruption of normal gene regulation through the emergence of alternative promoters (Fig. 2).

Collectively, these molecular disturbances contribute to severe obstetric complications. Preeclampsia and fetal growth restriction arise from impaired trophoblast differentiation and invasion of spiral arteries, processes closely associated with LINE-1 hyperactivation. Pregnancy loss occurs due to the combined effects of insertional mutagenesis, chromosomal instability, and defects in placental development. Future progress in this rapidly evolving field will benefit from advances in methodology. Long-read sequencing technologies enable direct detection of full-length MGE insertions in complex genomic regions (de la Morena-Barrio et al., 2022; Yano et al., 2024). Integrative multi-omics approaches combining ATAC-seq and RNA-seq will allow analysis of MGE influence on three-dimensional chromatin architecture, while CRISPR-Cas genome editing provides unique opportunities to model the

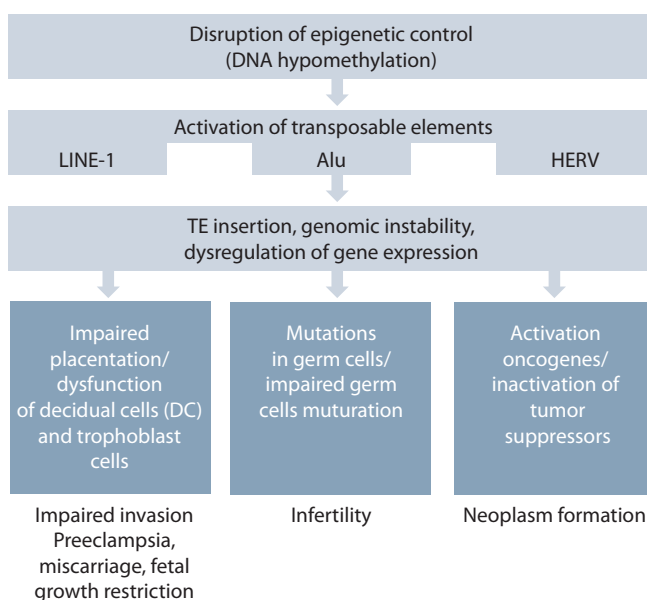


Fig. 2. Key pathways of MGE dysregulation and their contribution to the pathophysiology of reproductive disorders.

effects of specific MGE insertions or deletions in functional studies. Thus, investigating MGEs offers a promising avenue in molecular biology and genetics, enhancing our understanding of mechanisms underlying reproductive system homeostasis and laying the foundation for the development of diagnostic and therapeutic strategies for reproductive disorders.

References

- Blaise S., de Parseval N., Bénit L., Heidmann T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc Natl Acad Sci USA*. 2003;100(22):13013-13018. doi 10.1073/pnas.2132646100
- Bourque G., Burns K.H., Gehring M., Gorbunova V., Seluanov A., Hammell M., Imbeault M., Izsvák Z., Levin H.L., Macfarlan T.S., Mager D.L., Feschotte C. Ten things you should know about transposable elements. *Genome Biol*. 2018;19(1):199. doi 10.1186/s13059-018-1577-z
- Brouha B., Schustak J., Badge R.M., Lutz-Prigge S., Farley A.H., Moran J.V., Kazazian H.H. Jr. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci USA*. 2003; 100(9):5280-5285. doi 10.1073/pnas.0831042100
- Callinan P.A., Batzer M.A. Retrotransposable elements and human disease. *Genome Dyn*. 2006;1:104-115. doi 10.1159/000092503
- Chang S., Wang L., Guan Y., Shanguan S., Du Q., Wang Y., Zhang T., Zhang Y. Long interspersed nucleotide element-1 hypomethylation in folate-deficient mouse embryonic stem cells. *J Cell Biochem*. 2013;114(7):1549-1558. doi 10.1002/jcb.24496
- Chen X., Pacis A., Aracena K.A., Gona S., Kwan T., Groza C., Lin Y.L., Sindeaux R., Yotova V., Pramatarova A., Simon M.-M., Pastinen T., Barreiro L.B., Bourque G. Transposable elements are associated with the variable response to influenza infection. *Cell Genomics*. 2023; 3(5):100292. doi 10.1016/j.xgen.2023.100292
- Chénais B. Transposable elements in cancer and other human diseases. *Curr Cancer Drug Targets*. 2015;15(3):227-242. doi 10.2174/1568009615666150317122506

- Chénaïs B., Caruso A., Hiard S., Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene*. 2012;509(1):7-15. doi 10.1016/j.gene.2012.07.042
- Chow J.C., Ciaudo C., Fazzari M.J., Mise N., Servant N., Glass J.L., Attreed M., Avner P., Wutz A., Barillot E., Greally J.M., Voinnet O., Heard E. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell*. 2010;141(6):956-969. doi 10.1016/j.cell.2010.04.042
- Crow M.K. Type I interferon in the pathogenesis of lupus. *J Immunol*. 2014;192(12):5459-5468. doi 10.4049/jimmunol.1002795
- Curtis G., Menezes A.N., Brant A.C., de Mulder Rougvié M., Moreira M.Â.M., Soares M.A. Expression of retroelements in cervical cancer and their interplay with HPV infection and host gene expression. *Cancers (Basel)*. 2021;13(14):3513. doi 10.3390/cancers13143513
- de la Morena-Barrio B., Stephens J., de la Morena-Barrio M.E., Stefanucci L., Padilla J., Miñano A., Gleadall N., ... Vicente V., Ouwehand W.H., Corral J., Sanchis-Juan A.; NIHR Bio-Resource. Long-read sequencing identifies the first retrotransposon insertion and resolves structural variants causing antithrombin deficiency. *Thromb Haemost*. 2022;122(8):1369-1378. doi 10.1055/s-0042-1749345
- Demeneva V.V., Tolmacheva E.N., Nikitina T.V., Sazhenova E.A., Yuriev S.Y., Makhmutkhodzhaev A.S., Zuev A.S., Filatova S.A., Dmitriev A.E., Darkova Y.A., Nazarenko L.P., Lebedev I.N., Vasilyev S.A. Expression of the NUP153 and YWHAB genes from their canonical promoters and alternative promoters of the LINE-1 retrotransposon in the placenta of the first trimester of pregnancy. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov J Genet Breed*. 2023;27(1):63-71. doi 10.18699/VJGB-23-09
- Demeneva V.V., Tolmacheva E.N., Filatova S.A., Zuev A.S., Ushakova A.S., Vasilyeva O.Yu., Sazhenova E.A., Nikitina T.V., Vasiliev S.A. The level of methylation of LINE-1 retrotransposon in chorionic villi in embryos from families with sporadic and recurrent miscarriage. *Med Genet*. 2024;23(10):50-54. doi 10.25557/2073-7998.2024.10.50-54 (in Russian)
- Di Giacomo M., Comazzetto S., Saini H., De Fazio S., Carrieri C., Morgan M., Vasiliauskaitė L., Benes V., Enright A.J., O'Carroll D. Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Mol Cell*. 2013;50(4):601-608. doi 10.1016/j.molcel.2013.04.026
- Entringer S., Buss C., Swanson J.M., Cooper D.M., Wing D.A., Wafar F., Wadhwa P.D. Fetal programming of body composition, obesity, and metabolic function: the role of intrauterine stress and stress biology. *J Nutr Metab*. 2012;2012:632548. doi 10.1155/2012/632548
- Figueiredo J.C., Grau M.V., Wallace K., Levine A.J., Shen L., Hamdan R., Chen X., Bresalier R.S., McKeown-Eyssen G., Haile R.W., Baron J.A., Issa J.P. Global DNA hypomethylation (LINE-1) in the normal colon and lifestyle characteristics and dietary and genetic factors. *Cancer Epidemiol Biomarkers Prev*. 2009;18(4):1041-1049. doi 10.1158/1055-9965.EPI-08-0926
- Fu J., Qin T., Li C., Zhu J., Ding Y., Zhou M., Yang Q., Liu X., Zhou J., Chen F. Research progress of LINE-1 in the diagnosis, prognosis, and treatment of gynecologic tumors. *Front Oncol*. 2023;13:1201568. doi 10.3389/fonc.2023.1201568
- García-Pérez J.L., Widmann T.J., Adams I.R. The impact of transposable elements on mammalian development. *Development*. 2016;143(22):4101-4114. doi 10.1242/dev.132639
- Gardner E.J., Lam V.K., Harris D.N., Chuang N.T., Scott E.C., Pittard W.S., Mills R.E.; 1000 Genomes Project Consortium; Devine S.E. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res*. 2017;27(11):1916-1929. doi 10.1101/gr.218032.116
- Gilbert N., Lutz-Prigge S., Moran J.V. Genomic deletions created upon LINE-1 retrotransposition. *Cell*. 2002;110(3):315-325. doi 10.1016/S0092-8674(02)00828-0
- Goodier J.L. Restricting retrotransposons: a review. *Mob DNA*. 2016;7:16. doi 10.1186/s13100-016-0070-z
- Grow E.J., Flynn R.A., Chavez S.L., Bayless N.L., Wossidlo M., Wesche D.J., Martin L., Ware C.B., Blish C.A., Chang H.Y., Pera R.A., Wysocka J. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature*. 2015;522(7555):221-225. doi 10.1038/nature14308
- Hancks D.C., Kazazian H.H. Jr. Roles for retrotransposon insertions in human disease. *Mob DNA*. 2016;7:9. doi 10.1186/s13100-016-0065-9
- He X.J., Ecker J.R. Non-CG methylation in the human genome. *Annu Rev Genomics Hum Genet*. 2014;16:55-77. doi 10.1146/annurev-genom-090413-025437
- Howard G., Eiges R., Gaudet F., Jaenisch R., Eden A. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene*. 2008;27(3):404-408. doi 10.1038/sj.onc.1210631
- Hueso M., Cruzado J.M., Torras J., Navarro E. ALUminating the path of atherosclerosis progression: chaos theory suggests a role for alu repeats in the development of atherosclerotic vascular disease. *Int J Mol Sci*. 2018;19(6):1734. doi 10.3390/ijms19061734
- Jachowicz J.W., Bing X., Pontabry J., Bošković A., Rando O.J., Torres-Padilla M.E. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat Genet*. 2017;49(10):1502-1510. doi 10.1038/ng.3945
- Jenkins T.G., Aston K.I., Pflueger C., Cairns B.R., Carrell D.T. Age-associated sperm DNA methylation alterations: possible implications in offspring disease susceptibility. *PLoS Genet*. 2014;10(7):e1004458. doi 10.1371/journal.pgen.1004458
- Joubert B.R., den Dekker H.T., Felix J.F., Bohlin J., Ligthart S., Beckert E., Tiemeier H., ... Peddada S.D., Jaddoe V.W., Nystad W., Duijts L., London S.J. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. *Nat Commun*. 2016;7:10577. doi 10.1038/ncomms10577
- Kajihara T., Brosens J.J., Ishihara O. The role of FOXO1 in the decidual transformation of the endometrium and early pregnancy. *Med Mol Morphol*. 2013;46(2):61-68. doi 10.1007/s00795-013-0018-z
- Khitrinskaya I.Y., Stepanov V.A., Puzyrev V.P. Alu repeats in the human genome. *Mol Biol*. 2003;37(3):325-333. doi 10.1023/A:1024218806634
- Khitrinskaya I.Y., Kharkov V.N., Voevoda M.I., Stepanov V.A. Genetic diversity and relationships of populations of northern Eurasia by polymorphic Alu insertions. *Mol Biol*. 2014;48(1):58-68. doi 10.1134/S0026893314010051
- Kin K., Maziarz J., Wagner G.P. Immunohistological study of the endometrial stromal fibroblasts in the opossum, *Monodelphis domestica*: evidence for homology with eutherian stromal fibroblasts. *Biol Reprod*. 2014;90(5):111. doi 10.1095/biolreprod.113.115139
- Kong Y., Rose C.M., Cass A.A., Williams A.G., Darwish M., Lianoglou S., Haverty P.M., ... Mellman I., Bourgon R., Greally J., Jhunjhunwala S., Chen-Harris H. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nat Commun*. 2019;10(1):5228. doi 10.1038/s41467-019-13035-2
- Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., ... Shizuya H., Choi S., Chen Y.J., Szustakowski J.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860-921. doi 10.1038/35057062
- Lavialle C., Cornelis G., Dupressoir A., Esnault C., Heidmann O., Vernochet C., Heidmann T. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philos Trans R Soc Lond B Biol Sci*. 2013;368(1626):20120507. doi 10.1098/rstb.2012.0507

- Li B., Chang S., Liu C., Zhang M., Zhang L., Liang L., Li R., Wang X., Qin C., Zhang T., Niu B., Wang L. Low maternal dietary folate alters retrotransposon by methylation regulation in Intrauterine Growth Retardation (IUGR) fetuses in a mouse model. *Med Sci Monit.* 2019;25:3354-3365. doi 10.12659/MSM.914292
- Lou C., Goodier J.L., Qiang R. A potential new mechanism for pregnancy loss: considering the role of LINE-1 retrotransposons in early spontaneous miscarriage. *Reprod Biol Endocrinol.* 2020;18(1):6. doi 10.1186/s12958-020-0564-x
- Lynch V.J., Nnamani M.C., Kapusta A., Brayer K., Plaza S.L., Mazur E.C., Emera D., ... Young S.L., Lieb J.D., DeMayo F.J., Feschotte C., Wagner G.P. Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. *Cell Rep.* 2015;10(4):551-561. doi 10.1016/j.celrep.2014.12.052
- Macfarlan T.S., Gifford W.D., Driscoll S., Lettieri K., Rowe H.M., Bonanomi D., Firth A., Singer O., Trono D., Pfaff S.L. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature.* 2012;487(7405):57-63. doi 10.1038/nature11244
- Macia A., Widmann T.J., Heras S.R., Ayllon V., Sanchez L., Benkadour-Boumzaouad M., Muñoz-Lopez M., ... Menendez P., Ng P., Muotri A.R., Goodier J.L., Garcia-Perez J.L. Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Res.* 2017;27(3):335-348. doi 10.1101/gr.206805.116
- McKerrow W., Wang X., Mendez-Dorantes C., Mita P., Cao S., Grivainis M., Ding L., LaCava J., Burns K.H., Boeke J.D., Fenyö D. LINE-1 expression in cancer correlates with p53 mutation, copy number alteration, and S phase checkpoint. *Proc Natl Acad Sci USA.* 2022;119(8):e2115999119. doi 10.1073/pnas.2115999119
- Mess A., Carter A.M. Evolutionary transformations of fetal membrane characters in Eutheria with special reference to Afrotheria. *J Exp Zool B Mol Dev Evol.* 2006;306(6):480-494. doi 10.1002/jez.b.21079
- Mi S., Lee X., Li X., Veldman G.M., Finnerty H., Racie L., LaVallie E., Tang X.Y., Edouard P., Howes S., Keith J.C., McCoy J.M. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature.* 2000;403(6771):785-789. doi 10.1038/35001608
- Mika K., Lynch V.J. Transposable elements continuously remodel the regulatory landscape, transcriptome, and function of decidual stromal cells. *Genome Biol Evol.* 2022;14(12):evac164. doi 10.1093/gbe/evac164
- Mika K., Marinić M., Chigurupati S., Lynch V.J. Evolutionary transcriptomics implicates new genes and pathways in human pregnancy and adverse pregnancy outcomes. *Elife.* 2021;10:e69584. doi 10.7554/eLife.69584
- Miki Y., Nishisho I., Horii A., Miyoshi Y., Utsunomiya J., Kinzler K.W., Vogelstein B., Nakamura Y. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 1992;52(3):643-645
- Molaro A., Hodges E., Fang F., Song Q., McCombie W.R., Hannon G.J., Smith A.D. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell.* 2011;146(6):1029-1041. doi 10.1016/j.cell.2011.08.016
- Nguyen T.H.M., Carreira P.E., Sanchez-Luque F.J., Schauer S.N., Fagg A.C., Richardson S.R., Davies C.M., ... Perrin L.C., Hooper J.D., Ewing A.D., Upton K.R., Faulkner G.J. L1 retrotransposon heterogeneity in ovarian tumor cell evolution. *Cell Rep.* 2018;23(13):3730-3740. doi 10.1016/j.celrep.2018.05.090
- Percharde M., Lin C.J., Yin Y., Guan J., Peixoto G.A., Bulut-Karslioglu A., Biechele S., Huang B., Shen X., Ramalho-Santos M. A LINE1-nucleolin partnership regulates early development and esc identity. *Cell.* 2018;174(2):391-405.e19. doi 10.1016/j.cell.2018.05.043
- Polynnikova A.K., Zinovyeva O.E., Solokha O.A., Misyuryaeva E.V. Hereditary neuropathy with liability to pressure palsies: a case report. *Neurology, Neuropsychiatry, Psychosomatics.* 2021;13(4):116-122. doi 10.14412/2074-2711-2021-4-116-122 (in Russian)
- Ravel-Godreuil C., Znaidi R., Bonnifet T., Joshi R.L., Fuchs J. Transposable elements as new players in neurodegenerative diseases. *FEBS Lett.* 2021;595(22):2733-2755. doi 10.1002/1873-3468.14205
- Roy N., Haq I., Ngo J.C., Bennett D.A., Teich A.F., De Jager P.L., Olah M., Sher F. Elevated expression of the retrotransposon LINE-1 drives Alzheimer's disease-associated microglial dysfunction. *Acta Neuropathol.* 2024;148(1):75. doi 10.1007/s00401-024-02835-6
- Sato S., Gillette M., de Santiago P.R., Kuhn E., Burgess M., Doucette K., Feng Y., ... Birrer M.J., Skates S.J., Burns K.H., Carr S.A., Drapkin R. LINE-1 ORF1p as a candidate biomarker in high grade serous ovarian carcinoma. *Sci Rep.* 2023;13(1):1537. doi 10.1038/s41598-023-28840-5
- Sen S.K., Han K., Wang J., Lee J., Wang H., Callinan P.A., Dyer M., Cordaux R., Liang P., Batzer M.A. Human genomic deletions mediated by recombination between *Alu* elements. *Am J Hum Genet.* 2006;79(1):41-53. doi 10.1086/504600
- Slotkin R.K., Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007;8(4):272-285. doi 10.1038/nrg2072
- Smith Z.D., Chan M.M., Mikkelsen T.S., Gu H., Gnirke A., Regev A., Meissner A. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature.* 2012;484(7394):339-344. doi 10.1038/nature10960
- Taniguchi-Ikeda M., Kobayashi K., Kanagawa M., Yu C.C., Mori K., Oda T., Kuga A., Kurahashi H., Akman H.O., DiMauro S., Kaji R., Yokota T., Takeda S., Toda T. Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. *Nature.* 2011;478(7367):127-131. doi 10.1038/nature10456
- Tubio J.M.C., Li Y., Ju Y.S., Martincorena I., Cooke S.L., Tojo M., Gundem G., ... McDermott U., Campbell P.J., ICGC Breast Cancer Group; ICGC Bone Cancer Group; ICGC Prostate Cancer Group. Mobile DNA in cancer: Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science.* 2014;345(6196):1251343. doi 10.1126/science.1251343
- van den Hurk J.A.J.M., Meij I.C., Selem M.C., Kano H., Nikopoulos K., Hoefsloot L.H., Sistermans E.A., de Wijs I.J., Mukhopadhyay A., Plomp A.S., de Jong P.T., Kazazian H.H., Cremers F.P. L1 retrotransposition can occur early in human embryonic development. *Hum Mol Genet.* 2007;16(13):1587-1592. doi 10.1093/hmg/ddm108
- Vasiliev S.A., Tolmacheva E.N., Kashevarova A.A., Sazhenova E.A., Lebedev I.N. Methylation status of line-1 retrotransposon in chromosomal mosaicism during the early stages of human embryonic development. *Mol Biol.* 2015;49(1):144-152. doi 10.1134/S0026893314060193
- Vasilyev S.A., Tolmacheva E.N., Vasilyeva O.Y., Markov A.V., Zhigalina D.I., Zatula L.A., Lee V.A., Serdyukova E.S., Sazhenova E.A., Nikitina T.V., Kashevarova A.A., Lebedev I.N. LINE-1 retrotransposon methylation in chorionic villi of first trimester miscarriages with aneuploidy. *J Assist Reprod Genet.* 2021;38(1):139-149. doi 10.1007/s10815-020-02003-1
- Warren J., Im M., Ballesteros A., Ha C., Moore T., Lambert F., Lucas S., Hinz B., Dveksler G. Activation of latent transforming growth factor-β1, a conserved function for pregnancy-specific beta 1-glycoproteins. *Mol Hum Reprod.* 2018;24:602-612. doi 10.1093/molehr/gay044
- Wyrwoll M.J., Gaasbeek C.M., Golubickaite I., Stakaitis R., Oud M.S., Nagirnaja L., Dion C., ... Conrad D.F., Almstrup K., Veltman J.A., Tüttelmann F., van der Heijden G.W. The piRNA-pathway factor FKBp6 is essential for spermatogenesis but dispensable for control

- of meiotic LINE-1 expression in humans. *Am J Hum Genet.* 2022; 109(10):1850-1866. doi 10.1016/j.ajhg.2022.09.002
- Yano N., Chong P.F., Kojima K.K., Miyoshi T., Luqman-Fatah A., Kimura Y., Kora K., ... Hagiwara M., Kondo T., Kira R., Takita J., Yoshida T. Long-read sequencing identifies an SVA_D retrotransposon insertion deep within the intron of *ATP7A* as a novel cause of occipital horn syndrome. *J Med Genet.* 2024;61(10):950-958. doi 10.1136/jmg-2024-110056
- Zhang B., Xing X., Li J., Lowdon R.F., Zhou Y., Lin N., Zhang B., Sundaram V., Chiappinelli K.B., Hagemann I.S., Mutch D.G., Goodfellow P.J., Wang T. Comparative DNA methylome analysis of endometrial carcinoma reveals complex and distinct deregulation of cancer promoters and enhancers. *BMC Genomics.* 2014;15(1):868. doi 10.1186/1471-2164-15-868
- Zhang W., Huang C., Yao H., Yang S., Jiapaer Z., Song J., Wang X. Retrotransposon: an insight into neurological disorders from perspectives of neurodevelopment and aging. *Transl Neurodegener.* 2025;14(1):14. doi 10.1186/s40035-025-00471-y
- Zheng K., Wang P.J. Blockade of pachytene piRNA biogenesis reveals a novel requirement for maintaining post-meiotic germline genome integrity. *PLoS Genet.* 2012;8(11):e1003038. doi 10.1371/journal.pgen.1003038

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

Received November 25, 2025. Revised January 22, 2026. Accepted February 19, 2026.

doi 10.18699/vjgb-26-70

Evolution of the species *Ovis aries*

T.N. Khamiruev 

Research Institute of Veterinary Medicine of Eastern Siberia – Branch of the Siberian Federal Scientific Centre of AgroBioTechnologies
of the Russian Academy of Sciences, Chita, Russia

 tnik0979@mail.ru

Abstract. The sheep was one of the first domesticated animals in Neolithic Southwest Eurasia. The presented study suggests that the domestication of sheep occurred on the Anatolian plateau, to the northwest of the commonly accepted boundaries of the Fertile Crescent, rather than within it. However, many aspects of the domestication process (the specific place, time, and history after domestication) have remained not fully understood. The review examines in detail the complex origin, traces the primary role of the Asian mouflon (*Ovis gmelini*) as an ancestor, and also discusses controversial aspects of the contribution of other wild *Ovis* species. It reveals the history of the introduction and migration of sheep in the world, and summarizes the current scientific understanding of the phylogenetic relationships between populations of wild mouflons and domestic sheep (*Ovis aries*). A multifaceted process of domestication is considered, and the proposed evolutionary mechanisms are discussed, such as the domestication syndrome and hypotheses about thyroid hormones, as well as the human-mediated selection of key phenotypic traits. The article analyzes the results obtained using various genetic markers, including mitochondrial DNA haplogroups. Phylogenetic analysis using mitochondrial DNA has been successfully applied to identify the phylogeographic patterns and divergence times, from the early Neolithic to the Middle Ages, of sheep migration from the domestication center to Asia, Europe, and Africa. Domesticated sheep, having survived and endured extreme climate changes that occurred in the last post-glacial period, became the ancestors of modern local sheep breeds. Starting from the seventh millennium BC, domesticated sheep were brought to the Caucasus, Central Asia, and Europe. The spread of sheep in Asia began from the Middle East to the Mongolian Plateau and the Indian subcontinent, then to the north and southwest of China. In Russia, the territory of which covers a significant part of Eurasia, a unique breed diversity of sheep has been developed, with haplogroup B typical for breeds of the European-type origin (western geographic regions), and haplogroup A of the Asian-type sheep (eastern geographic regions).

Key words: evolution; genetics; mouflon; domestication; domestic sheep; migration; haplogroup

For citation: Khamiruev T.N. Evolution of the species *Ovis aries*. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2026;30(4):696-707. doi 10.18699/vjgb-26-70

Funding. This study was supported by the Ministry of Science and Higher Education of the Russian Federation, project No. FNUU-2025-0003.

Эволюция вида *Ovis aries*

Т.Н. Хаамируев 

Научно-исследовательский институт ветеринарии Восточной Сибири – филиал Сибирского федерального научного центра агробитехнологий
Российской академии наук, Чита, Россия

 tnik0979@mail.ru

Аннотация. Овца – одно из первых одомашненных животных в неолитической Юго-Западной Евразии. Представленное исследование позволяет предположить, что одомашнивание овец произошло на Анатолийском плато, к северо-западу от общепринятых границ Плодородного полумесяца, а не внутри них. Однако многие аспекты процесса доместикации (конкретное место, время и история после доместикации) остаются до конца невыясненными. В обзоре подробно рассматривается филогенез, прослеживается основная роль азиатского муфлона (*Ovis gmelini*) как предка, а также обсуждаются спорные аспекты вклада других диких видов *Ovis*, раскрывается история интродукции и миграции овец в мире, обобщается современное научное понимание филогенетических связей между популяциями диких муфлонов и домашних овец (*Ovis aries*). Рассматриваются многогранный процесс приручения, обсуждаются предлагаемые эволюционные механизмы, такие как синдром одомашнивания, и гипотезы о гормонах щитовидной железы, а также опосредованный человеком отбор ключевых фенотипических признаков. В статье анализируются результаты, полученные с помощью различных генетических маркеров, включая гаплогруппы митохондриальной ДНК. Филогенетический анализ с использованием митохондриальной ДНК был успешно применен для выявления филогеографических закономерностей и времени дивергенции с раннего неолита и до средневековья, миграции овец из центра одомашнивания в Азию, Европу и Африку. Одомашненные овцы, пережившие и выдержавшие экстремальные изменения климата, происходившие в последний послеледниковый период, стали родоначальниками современных местных пород овец. Начиная с седьмого тысячелетия до нашей эры одомашненные овцы были завезены на Кавказ, в Центральную Азию и Европу.

Распространение овец в Азии началось с Ближнего Востока на Монгольское плато и Индийский субконтинент, затем на север и юго-запад Китая. В России, территория которой занимает значительную часть Евразии, выведено уникальное породное разнообразие овец, при этом гаплогруппа В типична для пород овец европейского типа происхождения (западные географические регионы), а гаплогруппа А – для овец азиатского типа (восточные географические регионы).

Ключевые слова: эволюция; генетика; муфлон; domestикация; домашняя овца; миграция; гаплогруппа

Introduction

The genus *Ovis* has existed for about 8.31 million years and includes eight living species: domestic sheep (*O. aries*), argali (*O. ammon*), Asiatic mouflon (*O. orientalis*), European mouflon (*O. musimon*), urial (*O. vignei*), bighorn sheep (*O. canadensis*), thinnhorn sheep (*O. dalli*) and snow sheep (*O. nivicola*) (Rezaei et al., 2010). At the same time, the origin and “wild” status of the European mouflon are the subject of significant discussions (Hiendleder et al., 2002; Chessa et al., 2009). P. Mereu and colleagues (2025) believe that it should be classified as a subspecies of the Asian mouflon. It has been historically believed that European mouflons, especially those living in Corsica and Sardinia, are remnants of an ancient population of European wild sheep (Poplin, 1979). However, archaeological evidence indicates that they were brought to Cyprus by humans in the early Neolithic (about 10,000 years ago), and then to Corsica and Sardinia (Vigne et al., 2011). These populations were far from domesticated in the modern sense, and were kept in fenced areas only for protection from predators and to provide a constant, easily accessible source of meat and hides with minimal interaction between humans and animals (Barbato et al., 2022; Portanier et al., 2022; Mereu et al., 2024).

The domestic sheep *O. aries* Linnaeus is among the first domesticated species of livestock. Domestication occurred about 12,000 years ago in the Middle East, in the Zagros Mountains region (Her et al., 2022; Mereu et al., 2024), at the same time, many aspects of the domestication process, such as the exact place and time, remain unclear (Kaptan et al., 2024). Subsequently, as a result of human migration, sheep spread throughout Eurasia and under the influence of natural and artificial selection formed various populations (Wang et al., 2025).

The origin of domestic sheep included the initial domestication and early divergence, followed by extensive post-Neolithic diversification. At the early stage of domestication, human involvement was minimal and was mainly focused on protecting and providing food for the sheep, rather than on intensive breeding aimed at developing specific traits (Portanier et al., 2022; Mereu et al., 2025). It is extremely important to understand this primitive phase of sheep domestication in order to distinguish the genetic changes that existed before domestication from those directly associated with selection carried out by humans. This historical context shows that domestication is not a single event, but a continuous process characterized by gradual accumulation of genetic, phenotypic, and physiological changes (Jackson et al., 2020; Mereu et al., 2024).

F. Teletchea and P. Fontaine (2014) proposed a classification with five levels of domestication: the first level corresponds to initial trials of acclimating wild animals in captivity. Then, when part of the life cycle is under control, the second level comes, where the most important thing is breeding in captivity. The third level occurs when the entire life cycle is under control, but the instincts of wild ancestors are still present. At the fourth level, domesticated animals differ significantly from their wild counterparts, and the fifth level leads to the creation of certain breeds. It is also worthy of note that different population groups within a single species can indeed exhibit different levels of domestication, even within the same geographic area (Teletchea, Fontaine, 2014; Zeder, 2015).

Despite intensive work using genomic data of modern sheep (Lv et al., 2022; Matyukov et al., 2023; Deniskova et al., 2025; Dossybayev et al., 2025), neither the exact location nor the wild ancestors involved in the first domestication process, nor the post-Neolithic demographic history of domestic sheep or mouflon populations have been clarified so far (Kaptan et al., 2024).

Domestic sheep origin

One of the open questions is the origin of the gene pool of domestic sheep. Earlier studies pointed to several species of Asian sheep (*O. ammon*, Linnaeus, 1758; *O. gmelini*, Blyth, 1841; *O. vignei*, Blyth, 1841) as the probable ancestor of domestic sheep (Pedrosa et al., 2005; Sanna et al., 2015). However, recent genetic studies confirm the view that the Asian mouflon is the only ancestor of the domestic sheep (Chen et al., 2021); at the same time, according to D. Kaptan et al. (2024), this fact has not been fully established. More recent genomic studies indicate that the *VPS13B* gene shows signals of introgression from urials or mouflons, while argali contributed alleles to the lines of domestic sheep in Southeast Asia, introgressing the *MSRB3* gene, which has been found to be specifically associated with variations in ear morphology (Cheng et al., 2023).

The *O. gmelina* species includes five subspecies: the Armenian mouflon (*O. gmelini gmelini*, Blyth, 1841), Isfahan mouflon (*O. gmelini isphahanica*, Nasonov, 1910), Laristan mouflon (*O. gmelini laristanica*, Nasonov, 1910), Cyprus mouflon (*O. gmelini ophion*, Blyth, 1841) and Anatolian mouflon (*O. gmelini anatolica*, Valenciennes, 1856). Due to the complex intraspecific taxonomy of the Asian mouflon, the genetic contribution of various subspecies to the domestication of sheep remains unclear (Demirci et al., 2013; Li R. et al., 2021; Daly et al., 2025). According to D. Sanna et al. (2015), these subspecies represent local wild populations since the beginning of the Holocene, with the exception of the Cypriot

mouflon, which was brought from mainland Anatolia around the 12th millennium BC.

S. Hiendleder et al. (2002) suggested that sheep were domesticated from two different subspecies in Turkey and western Iran (Middle East), corresponding to mtDNA lineages A and B of domestic sheep, respectively. Significant nucleotide diversity of mitochondrial lineages in this region confirms the geographical role in sheep domestication and indicates a more complex picture of domestication than previously assumed (Meadows et al., 2007; Machová et al., 2022).

Recent studies, including genetic and geographic data of Asian mouflons across their range, as well as modern and ancient sheep from Anatolia, Southwest Asia, Europe, and Africa, suggest that early sheep domestication probably originated in Anatolia and is associated with the Anatolian subspecies of the mouflon (*O. gmelini anatolica*) (Her et al., 2022; Atağ et al., 2024; Sandoval-Castellanos et al., 2024).

It is assumed that Anatolian and Cypriot mouflons in the past were subject to domestication according to the classification F. Teletchea and P. Fontaine (2014), corresponding to the first level (Sanna et al., 2015; Barbato et al., 2017). However, this is only a hypothesis for now, and the genetic connections between these mouflons, ancient and modern domestic sheep remain unclear.

Another largely unresolved question is the history of domestic sheep breeds. Modern sheep breeds are divided into two main geographical groups: Europe and Asia–Africa (Naval-Sanchez et al., 2018; Li X. et al., 2020). This eastern and western genetic division can be traced back to 7,000–6,000 BCE, indicating their early diversification. Such a division is also observed in modern groups of mitochondrial haplotypes of sheep: European sheep mainly carry haplotype B (>50 % of the world's sheep population), while Asian sheep predominantly have haplotype A (34 % of the world's sheep population) (Machová et al., 2022; Mereu et al., 2024).

Based on the analysis of polymorphism in the control region of sheep mtDNA, seven haplogroups were identified, two of which (F and G) have disappeared, while the other five are present in modern breeds (Wood, Phua, 1996; Chen et al., 2021). In addition to the mentioned haplogroups A and B, a third recognized phylogenetic branch, haplogroup C (9 % of the world's sheep population), was discovered, which was found in local Portuguese sheep, as well as in individuals from the Caucasus, the Middle East, and Asia (Guo et al., 2005; Pedrosa et al., 2005; Pereira et al., 2006; Mereu et al., 2024). Then the fourth maternal lineage, haplogroup D, which is closest to *O. gmelini anatolica*, was identified (Demirci et al., 2013; Sanna et al., 2015).

Haplogroup E, identified through analysis of D-loop, *CytB* sequences, and the entire mitogenome, is the key mitochondrial lineage of the Iranian Asian mouflon (Mereu et al., 2025; Wang et al., 2025). The most recently discovered D and E haplogroups (<0.5 % of the world's sheep population) are the rarest, having been identified only in animals from Turkey and the Caucasus (Tapio et al., 2006), Europe (Gáspárdy et al., 2021), Tibet (Liu et al., 2016), Iran (Rafia, Tarang, 2016).

All haplogroups were presumably formed between 5 and 35 thousand years ago (Rezaei et al., 2010).

Scientific literature reports on the eighth “wild” mitochondrial lineage X, which is found in Cypriot, Anatolian, and Iranian mouflons and apparently is basal in relation to domestic haplogroups C and E (Guerrini et al., 2021; Mereu et al., 2025).

The results of ancient DNA studies have shown that Anatolian Neolithic sheep are more similar to modern European breeds, while the Kyrgyz Neolithic sheep are more similar to modern Asian breeds, suggesting an early formation of this division (Yurtman et al., 2021).

These patterns imply either the presence of several centers of domestication or a large heterogeneous precursor population that went through several independent “bottlenecks”. However, in the study by E. Yurtman et al. (2021), all modern breeds showed greater genetic similarity to each other than to Neolithic sheep, indicating significant breed mixing in the post-Neolithic period among continental sheep populations, including possible introgression from wild sheep into domestic flocks, as well as the spread and breeding of sheep with desired traits across the continents (Deng et al., 2020; Cheng et al., 2023).

Introgression from wild relatives into the domestic sheep population allowed the latter to adapt to various environmental conditions and spread throughout the world, undergoing genetic improvements in various production systems. General signals of allele introgression have been established in the genes related to olfaction (*ADCY3* and *TRPV1*), and in the *PADI* gene family, including *PADI2*, which is associated with innate immunity.

Further analysis of whole-genome sequences showed that introgressed alleles in a specific region of *PADI2* (chr2: 248302667–248306614) correlate with resistance to pneumonia. This allowed sheep to adapt to various climatic and environmental conditions after domestication (Ciani et al., 2020; Cao et al., 2021). As a result, many unique breeds have emerged, and the extensive variations observed in both local and improved breeds highlight the genomic diversity, adaptive traits, and important productive characteristics inherent in domestic sheep (Li X. et al., 2020; Cao et al., 2021; Alipanah et al., 2025). The differentiation of local sheep populations into breeds became more pronounced from the 18th century due to the use of systematic breeding with clearly defined goals (Ciani et al., 2020).

Phenotypic changes in domestic sheep

In the transition from wild ancestors, the mouflons, to modern domestic sheep, a number of morphological, physiological, and behavioral traits were formed, which led to the current phenotypic differences between wild and domesticated species. These traits, including brain and tooth size, the size and shape of ears and tail, spotted coloration, hormonal changes, and the length of the reproductive season, characterize many species that have been domesticated in various ways and for different reasons. Moreover, some of these traits apparently appeared

without deliberate selection: body size shrinkage, a change in the size and shape of the tail (Zeder, 2015).

Fat-tailed sheep originated from thin-tailed sheep as a result of long-term natural selection in extremely harsh geographical and climatic conditions (Caiye et al., 2023). The presence of a fat tail allows sheep to better adapt to harsh environmental conditions (Kalds et al., 2022a). The connection between domestication and a set of phenotypic traits was first noted by C.R. Darwin (1875) and has still not been fully elucidated.

S.J. Crockford et al. (2002) suggested that the phenotypic traits of domesticated animals may be associated with the changes in the “signaling actions” of thyroid hormones. In turn, A.S. Wilkins et al. (2014) proposed the hypothesis of the “domestication syndrome”, according to which the phenotypic traits of domestic animal species may be associated with the changes in the formation, differentiation, or migratory patterns of neural crest cells (Wilkins et al., 2020). Both hypotheses were later evaluated based on the genomic data, and although the “domestication syndrome” hypothesis received significant support, they are not necessarily mutually exclusive, as they could have influenced the domestication process differently across time and space (Karlsson et al., 2016; Fitak et al., 2020).

Thus, selection for the trait of “complaisance” likely indirectly influenced a wide range of genes and signaling pathways involved in behavior, morphology, and physiology. However, considering the hypotheses outlined above, it should not be overlooked that for certain traits, intentional selection carried out by humans should be regarded as the most plausible hypothesis for the development of the domesticated animal’s phenotype (Mereu et al., 2024).

Wild ancestors usually had horns in both sexes, coarse wool, and specific color patterns, which gradually evolved predominantly into hornless, woolly breeds of sheep with white wool color (Garel et al., 2022). These transformations reveal important aspects of the selective pressure that acted during domestication.

Wild sheep usually have horns, which matters for intra-specific competition (Hu et al., 2016). Unlike them, domestic sheep are more often hornless (Simon et al., 2022). The *RXFP2* gene plays an important role in the presence or absence of horns (Cheng et al., 2023). Hornlessness is associated with a 1.8 kb insertion in the 3'-untranslated region of the *RXFP2* gene, but this variant does not fully explain the variability of horn condition in different breeds (Simon et al., 2022).

The key transformation was the loss of the ability to shed wool annually in wild mouflons and the continuous growth of wool in domestic sheep (Jackson et al., 2020). The variability of wool color, often white in domestic sheep, is a complex trait influenced by genes such as *Extension* (*E* gene), *ASIP* (*agouti*), and *POMC*, which regulate melanin production (Trigo et al., 2021; Zhang et al., 2023).

The short, thin tail of the mouflon suggests that the various tail phenotypes observed in domestic sheep appeared later in the process of domestication (Kalds et al., 2022b). The development of a fat tail, for example, is considered an adaptive response to climate change, with ancient breeders

selecting sheep with fat tails for their increased adaptability to desert conditions and as a valuable source of fat (Ahbara et al., 2022; McManus et al., 2025). Genes such as bone morphogenetic protein 2 (*BMP-2*) and platelet-derived growth factor D (*PDGF-D*) are potential genetic markers for tail shape and size (Luo et al., 2023; Jin et al., 2026). The T-box T (*TBXT*) transcription factor gene is involved in regulating tail length in mammals, including sheep (Kalds et al., 2022b). CT/CT genotype of the *TBXT* gene: c. [333G>C; 334G>T] was found in fat-tailed sheep breeds and was absent in long- and short-tailed breeds (Han et al., 2019).

The emergence of these phenotypic changes is not accidental, but represents a direct adaptation to environmental conditions, highlighting the strong selective pressure exerted by humans in the process of domestication. Analysis of these phenotypic shifts in combination with the identification of the underlying genes provides valuable insights into the genetic architecture of domestication and how human needs and environmental factors have shaped the evolution of the species *O. aries*.

Later introgressions from wild ancestors to domestic sheep contain loci with potential domestication genes *PAPPA2*, *NR6A1*, *SH3GL3*, *RF3X* and *CAMK4*, which are associated with morphological, immune, reproductive, or productive traits (wool, meat, milk), *NEURL1* – with a nervous reaction, *PRUNE2* – with neurogenesis, *USH2A* – with hearing, *AGTPBP1*, *CRTAC1* and *RPGRIP1L* – with the sense of smell, *PAG11* and *PAG3* – with placental viability (Chen et al., 2021), *FAT3* – with adaptation to dry environmental conditions (Yang J. et al., 2016), *PCDH15* – with the immune response (Atlija et al., 2016), *PDGFD* – with the tail shape (Li X. et al., 2020).

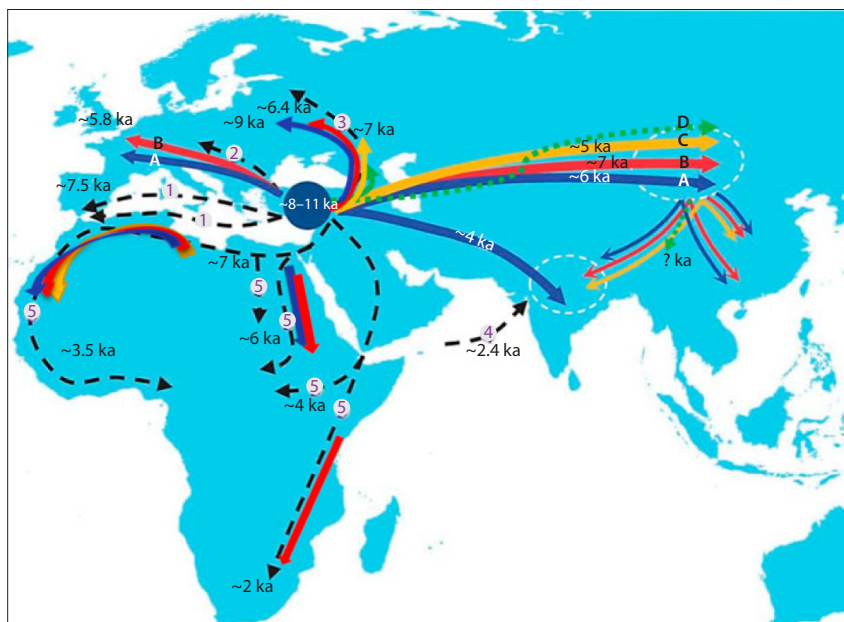
Migration of domesticated sheep

Starting from the 7th millennium BC, domestic sheep were brought from Southwest Asia to the northeast, to the Caucasus (Chataigner et al., 2014), to Central Asia (Taylor et al., 2021) and to Europe (Arbuckle, Atici, 2013; Lv et al., 2015).

According to F.-H. Lv et al. (2015) and K. Machová et al. (2022), the spread of sheep across continents from the domestication center occurred according to the Figure.

There is a hypothesis about the existence of two routes for the spread of sheep from Anatolia to Europe: a continental route and a sea route through the Mediterranean Sea (Racimo et al., 2020; Brigand et al., 2022). The sheep of Northern France during the Neolithic era arrived in this territory along with the Linear Pottery culture, which, as is known, spread westward from the middle Danube to the Rhine and Seine basins, as well as northward and eastward to the Elbe and Vistula basins (Hachem, 2018; Auxiette, Hachem, 2021).

Iberian sheep of the Neolithic era were possibly brought by the Mediterranean Sea route (Kaptan et al., 2024). Later, the settlement of Northern Italy and Southern France took place. Then, along the Danube route, sheep spread to Central and Northern Europe (Larsson et al., 2024). Sheep appeared in the Alps just over 5,000 years ago, on the Iberian Peninsula



The main routes of sheep spread from the center of domestication (ka – thousand years BC) (Lv et al., 2015; Machová et al., 2022).

A, B, C, D – paths of the main mitochondrial lineages (Lv et al., 2015; Liu et al., 2016).

1 – Mediterranean route (Ryder, 1984; Zeder, 2017); 2 – Danube route (Ryder, 1984; Zeder, 2017); 3 – route to Northern Europe (Tapio et al., 2006); 4 – routes of the ancient sea to the Indian subcontinent (Singh et al. 2013); 5 – routes to Africa (Gornas et al., 2011; Álvarez et al., 2013; Muigai, Hanotte, 2013; Resende et al., 2016).

(Spain, Portugal), 7,700–7,300 years ago (Chen et al., 2021). However, according to another hypothesis, there was a settlement of Northern Europe through the domestication center in the Caucasus (Tapio et al., 2006; Niemi et al., 2013).

Analysis of the spread of the lines in Eastern Eurasia showed the presence of two migratory waves of sheep that occurred 4.5–6.8 thousand years ago (lines A and B, about 6.4–6.8 thousand years ago; C, about 4.5 thousand years ago) (Lv et al. 2015).

According to the study by D. Cai et al. (2011), line A was the most numerous in Bronze Age Ancient China (95.5 %), and its population increased from west to east. It is assumed that sheep from the Near Eastern center of domestication migrated through the Caucasus and Central Asia, and settled in Northern and Southwestern China (lines A, B, and C), and then on the Indian subcontinent (lines B and C) (Lv et al., 2015).

The spread of sheep in Asia could have occurred through several routes, as it did in Europe (Singh et al., 2013). A combined phylogenetic analysis using ancient and modern sheep mitogenomes showed that Uzbekistan and the Mongolian Plateau (Altai) could have been the migration centers for the spread of sheep in East Asia. Through Uzbekistan and northwestern China, sheep spread to the middle and lower reaches of the Yellow River (around 4,000 BC), through the Mongolian Plateau (Altai) to the central part of Inner Mongolia (around 4,429–2,500 BC) (Yang L. et al., 2024). Earlier studies confirm that sheep spread to China from the Mongolian Plateau approximately 5,000–5,700 years ago. Then, about 2000–2600 years ago, sheep reached the Qinghai-Tibet and

Yunnan-Guizhou Plateaus, following the migration routes of the Di-Qian people from the north to the southwest (Zhao X.Y. et al., 2017).

Sheep probably reached North Africa by two routes about 7,000 years ago. The first was a colonization settlement across the Mediterranean Basin, the second was via the Sinai Peninsula across the Red Sea (Zeder, 2017). There were several routes of settlement on the African continent: to the south to the Middle Nile valley, to the west to the central Sahara, and to the north to Libya. Another possibility is the spread of sheep from the Mediterranean along the northern coasts of Africa. In addition, direct trade connections between East Africa and the Arabian Peninsula are also suggested (Muigai, Hanotte, 2013). As in Europe, mitochondrial haplogroup B dominates in Africa, which is confirmed by the studies in different regions of the continent – South Africa (Horsburgh, Rhines, 2010), Sudan (Gornas et al., 2011), Kenya (Resende et al., 2016), West Africa and the Canary Islands (Álvarez et al., 2013).

The spread of sheep in Australia occurred in modern history in the 18th–19th centuries (Ciani et al., 2015). The first sheep brought by the Spanish to Central America were either fine-wool (West African fine-wool sheep) or coarse-wool (from the Iberian Peninsula), which were later crossed with Merinos and gave rise to the Creole type of sheep (Alonso et al., 2017).

Fine-wool sheep were brought to America from the Canary Islands by C. Columbus and the first colonizers and at the beginning of the seventeenth century along with slaves from other regions of West Africa, and their contribution to the

gene pool of modern American fine-wool sheep is the most significant (Spangler et al., 2017).

The first sheep in Australia were brought from India, South Africa (fat-tailed) and Spain (Merinos) after 1788, as well as from the British Isles (Saxon Merinos, Southdowns and Romneys) after 1840 (Ryder, 1984). Thus, it can be assumed that the same haplotypes exist as in the populations from which the Australian breeds originated. This assumption was confirmed by a study conducted on 18 breeds kept in Australia, which revealed a 55 % prevalence of haplogroup B and a 45 % prevalence of haplogroup A (Meadows et al., 2005).

Based on the analysis of ancient samples, it has been established that modern breeds show greater genetic similarity to each other than to Neolithic sheep, which implies significant post-Neolithic mixing of sheep populations (Kaptan et al., 2024).

Originally, sheep breeding was mainly focused on meat production, while specialization in wool and milk arose in Asia (7–6 thousand years BC) and then in Europe (Becker et al., 2016; Deng et al., 2020). Specialization in wool production apparently originated in Southwest Asia and only then spread to Europe, as evidenced by research (Chessa et al., 2009) and the DNA analysis of European Bronze Age sheep (Sabatini et al., 2019). Archaeological findings also indicate the emergence of a new breed in Central Europe during the Late Stone Age. Comparison with earlier data confirmed an increase in the body size of sheep raised in the Bohemia and Moravia region (the territory of modern-day Czech Republic), as well as since the beginning of the Bronze Age in Hungary (Kysely, 2016). Another example is the spread of Merino sheep from the Iberian Peninsula throughout Europe from the second half of the 15th century (Landi et al., 2019).

Early domestication highlights the fundamental importance of sheep in human history, as since then they have represented a crucial resource on a global scale, providing populations around the world with meat, wool, pelts, sheepskins, milk, and other essential materials (Kaptan et al., 2024; Mereu et al., 2024).

Early domestication was probably a less controlled process, based on the use of available wild populations rather than on intensive selection of a single homogeneous ancestral group. Therefore, the genetic diversity observed in modern sheep partially reflects ancient wild diversity, and is not exclusively the product of post-domestication mutation or selection. At the same time, the early radiation (about 810 thousand years ago) determined the variability of domestic sheep and mouflons (Mereu et al., 2025).

Thus, the process of sheep domestication should not be considered as a simplified model of domestication in the form of a single event from a homogeneous ancestral population, but as a more complex action involving the borrowing of diverse wild genetic backgrounds.

Phylogenetic analysis of modern sheep breeds based on mtDNA

Subsequent breeding of domesticated sheep, along with natural and artificial selection, led to the emergence of at least

1,400 breeds, and the prolonged selective activity of humans played an important role in the development of the necessary productive and biological traits in sheep (Dymova et al., 2019). Comparative analysis of the genetic diversity of ancient and modern sheep contributes to a better understanding of the processes of the origin of these animals, domestication, and their migration, and also allows assessing the role of humans at different stages of the formation of modern biological diversity of sheep breeds (Lv et al., 2015).

In the study by X. Li et al. (2020), 36 local and 6 cultured breeds of sheep (*O. aries*) from Europe, Asia, Africa, and the Middle East were found to have a lower level of genomic diversity compared to their wild ancestors (*O. orientalis*). This indicates that a significant part of genomic variability was lost during and after domestication, while genomic diversity in local breeds of sheep has been preserved to a greater extent than in commercial breeds.

The formation of sheep breeds in Europe, according to E. Ciani et al. (2020), indicates that Greek sheep breeds acted as a barrier between Asian and European breeds. Apparently, Greece remained a “blind spot” on the evolutionary map of sheep migration to Western Europe, since the historical patterns of gene flow have not yet been assessed (Michailidou et al., 2025).

In Italian sheep breeds, haplogroup B predominated in the ancient samples (90 %), while haplogroup A was found in 10 % of them, which corresponds to the current distribution in modern sheep in Italy. This indicates that the current A/B ratio was formed back in the Middle Ages and is not the result of subsequent events, such as selective breeding (Gabbianelli et al., 2015). Eight of the most common local breeds of sheep in Bulgaria showed a high prevalence of European haplogroup B (95.2 %), while the remaining individuals were assigned to haplogroup A (4.8 %). The median-joining network showed that almost all haplotypes belonging to haplogroup B form a star-like network, indicating weak genetic differentiation and high gene flow among Bulgarian indigenous breeds (Kalaydzhiiev et al., 2023).

The presence of haplogroups B and A was confirmed in two Hungarian breeds, Tsigai and Cikta, while haplogroup C was found only in the Cikta breed. At the same time, haplogroup B predominates, which is present at a level of more than 80 % in both breeds, confirming the presence of a common ancestral root (Gáspárdy et al., 2022). A homogeneous maternal origin of nine breeds of sheep from the Eastern Adriatic, characteristic of European breeds, was established in the studies of M. Ferenčakovic et al. (2013); analysis of mitochondrial DNA revealed a high frequency of type B haplotypes in them.

Among the 27 analyzed Indian sheep breeds, three mtDNA lineages were identified, namely A, B, and C. Lineage A predominated among Indian sheep, whereas lineages B and C were observed at a low frequency. At the same time, the Mandya and Sonadi breeds differed significantly from other Indian breeds by a high proportion of line B, which presumably reached the Indian subcontinent not from the Mongolian Plateau, but via the Arabian Sea route (Kamalakkannan et al., 2021). The

authors of the cited work, based on the research conducted, contrary to existing hypotheses, believe that the domestication of sheep of haplogroup A occurred on the Indian subcontinent.

Phylogenetic analysis of the D-loop mtDNA sequences of 963 individuals from 16 indigenous breeds distributed across seven geographic regions of China showed that three haplogroups, A, B, and C, were identified in all breeds, except for the mountainous region of Southwest China, where only haplogroups A and B were present (Zhao E. et al., 2013).

Seven local Indonesian sheep breeds on the island of Java are closely related to each other and grouped into two haplogroups A and B. At the same time, most of them belong to haplogroup B, except for the Garut and Priangan sheep breeds (Ibrahim et al., 2023). Sheep of the main breed of Vietnam, Phan Rang, belong to haplogroups A (28 %) and B (72 %), which confirms the hypothesis of their dual origin with the influence of both Asian and European lines (Luong et al., 2024). Phylogenetic analysis of six breeds of sheep in Kazakhstan showed that 44 % of the animals belonged to haplogroup A, 39 % to haplogroup B, and 17 % to haplogroup C, which is consistent with the distribution of haplogroups in ancient domestic sheep. Currently, line A predominates in Kazakh sheep breeds. This indicates that most Kazakh sheep breeds are genetically closer to the Asian genotype than to the European one (Tarlykov et al., 2021).

Sheep in Africa belong to different mitochondrial lineages, which gives them an advantage in genetic diversity of breeds and populations, and has practical significance for use in breeding programs to increase productivity and adaptive traits. Meta-analysis of 399 breeds allowed the identification of three main haplogroups (A, B, and C), with haplogroup B being the dominant one (Wanjala et al., 2021; Salim et al., 2023).

Currently, 17 fine-wool breeds (57.9 % of the total sheep population) are raised in Russia, as well as 17 semi-fine-wool breeds (4.1 %), 2 semi-coarse-wool breeds (1.1 %), 16 coarse-wool breeds (30.3 %), and unidentified breeds (6.6 %) (Brigida et al., 2025). Semi-fine-wool sheep breeds mainly have a meat-wool production direction, while semi-coarse-wool and coarse-wool breeds have a meat-fat direction. The abundance of sheep breeds in the country is associated with the diversity of natural-climatic and geographical zones.

Genetic analysis of mtDNA of ancient sheep from Russia indicates that during the Early Bronze Age, at least two lineages (A and B) existed in the south of Western Siberia (Dymova et al., 2019). Sequencing of the mitochondrial DNA D-loop from 17 sheep bone remains (aged ~4,000–1,000 years), found in archaeological complexes in southern Altai (Russia), showed the presence of mitochondrial lineages A, B, C, D, and E. The relatively high diversity of sheep haplotypes, including the presence of two basal haplotypes, indicates that the Altai region could have been a route for the migration of domesticated sheep. Determining the affiliation of ancient samples to phylogenetic lines showed that 70 % belong to line B, 25 % to A, and 2.5 % each to C and D (Dymova et al., 2017).

Phylogenetic analysis of 25 Russian local sheep breeds revealed four haplogroups, including A, B, C, and D, which is

explained by the wide range of the animals studied; the most common were haplogroups B and A, characteristic of sheep of European and Asian origin, with the B lineage predominating in the western part of Russia (Koshkina et al., 2021).

Analysis of the median-joining network and the Bayesian tree of the Russian breeds showed that the most common haplogroup is B (64.8 %), haplogroup A accounted for 28.9 %, haplogroup C, for 5.5 %, and only 0.8 % was assigned to haplogroup D. It is noted that haplogroups A and B were present in all sheep breeds, haplogroup C was represented by haplotypes of sheep breeds from the North Caucasus and Far Eastern Federal Districts, while haplogroup D was detected in one animal from the Southern Federal District (Koshkina et al., 2023). M. Tapio et al. (2006) discovered haplogroup D in one animal of the Karachay breed in the North Caucasus, indicating the presence of this mitochondrial line in the territory of Russia.

The study of the nucleotide sequence of the D-loop of mtDNA in fine-wool sheep breeds (Salsk, Stavropol, and Soviet Merino) showed the division of the studied sheep populations into two clusters: the first cluster is represented by haplogroup B, characteristic of European domestic sheep, as well as sheep from New Zealand and Australia; the second cluster is haplogroup A (about 30 %), dominant in Asian sheep. The obtained research results made it possible to establish the similarities and differences of the Russian Merino sheep breeds with the European and Australian Merino selection (Shirokova et al., 2018). In the determination of the Romanov sheep (Bakoev et al., 2018), the Kuibyshev breed (Koshkina et al., 2022a), it was established that they belong to haplogroup B.

Analysis of the mitochondrial genomes of three breeds with different production directions (Grozny – fine-wool, Gorno-Altai – semi-fine-wool, and Buubey – coarse-wool) from various ecological and feed regions of the country showed that all individuals of the three breeds belong to the commonly accepted haplogroups A ($n = 46.7$ %), B ($n = 46.7$ %), and C ($n = 6.6$ %), characteristic of sheep of European and Asian origin (Koshkina et al., 2022b).

Conclusion

Thus, based on the analysis of scientific and informational sources in the subject area, it has been established that with domestication, aspects such as the specific place, time, and further development after domestication remain unclear to this day.

Mitochondrial DNA is an excellent tool of evolutionary biology. It is inherited from generation to generation almost without changes, since it does not participate in recombination, thus the time of occurrence of each mutation can be estimated based on the average mutation rate. This ensures that female haplogroups allow us to determine events that occurred in the distant past (processes of domestication and migration). Seven haplogroups have been identified in domesticated sheep, two of which (F and G) have disappeared, while the other five (A, B, C, D, E) are present in modern breeds.

The variability of mtDNA to some extent correlates with the distance from the domestication center. The greatest genomic diversity is found closer to the place of origin, as has been established in humans (Li J.Z. et al., 2008). This review confirms the existence of a single center of domestication in the Middle East, from which sheep spread to Europe through the Mediterranean Sea and the Danube Valley, as evidenced by studies of small ruminant lentiviruses (Molae et al., 2020). It is possible that there was another way of spreading sheep into Europe, which passed through the Caucasus, Russia, and Northern Europe (Tapio et al., 2006). First, line B probably reached Finland, then, in the early Middle Ages, it was followed by line A (Niemi et al., 2013). The origin of lines C and D in Central Europe remains unclear, but there is a suggestion that they could have arrived there with prehistoric humans or much later, for example, during the Ottoman expansion (Gáspárdy et al., 2021). The Mongolian Plateau acted as a migration center, from which the lines spread from the domestication center into Asia (Ganbold et al., 2019). In India, a high nucleotide diversity of lines A and B has been established, in Northern China, line C has been found (Lv et al., 2015).

During the process of migration, domesticated sheep adapted to various ecological and feeding conditions of their habitats on all continents of the world, and as a result, modern domestic sheep have developed a number of morphological, physiological, and behavioral traits, which have led to the current phenotypic differences between wild and domestic species.

The results of ancient DNA studies have shown that Anatolian Neolithic sheep have more similarity with modern European breeds, while Kyrgyz sheep of the Neolithic and Bronze Age demonstrate more similarity with modern Asian breeds.

Currently, domestic sheep are divided into two main geographical groups, Europe and Asia: European sheep mainly carry haplotype B, while Asian sheep predominantly carry haplotype A. The diversity of the haplogroups represented in both domesticated and modern breeds of sheep indirectly indicates their migration across the territory of Eurasia, including the Russian Federation, in two directions – to Europe (westward) and to Asia (eastward).

References

- Ahbara A.M., Musa H.H., Robert C., Abebe A., Al-Jumaili A.S., Kebede A., Latairish S., Agoub M.O., Clark E., Hanotte O., Mwacharo J.M. Natural adaptation and human selection of northeast African sheep genomes. *Genomics*. 2022;114(5):110448. doi 10.1016/j.ygeno.2022.110448
- Alipanah M., Gharari F., Mazloom S.M. Comparison of selection signatures among five meat sheep breeds. *Small Ruminant Res*. 2025; 249:107524. doi 10.1016/j.smallrumres.2025.107524
- Alonso R.A., Ulloa-Arvizu R., Gayosso-Vázquez A. Mitochondrial DNA sequence analysis of the Mexican Creole sheep (*Ovis aries*) reveals a narrow Iberian maternal origin. *Mitochondrial DNA A DNA Mapp Seq Anal*. 2017;28(6):793-800. doi 10.1080/24701394.2016.1192613
- Álvarez I., Capote J., Traoré A., Fonseca N., Pérez K., Cuervo M., Fernández I., Goyache F. Mitochondrial analysis sheds light on the origin of hair sheep. *Anim Genet*. 2013;44(3):344-347. doi 10.1111/j.1365-2052.2012.02398.x
- Arbuckle B.S., Atici L. Initial diversity in sheep and goat management in Neolithic South-Western Asia. *Levant*. 2013;45(2):219-235. doi 10.1179/0075891413Z.000000000026
- Atağ G., Kaptan D., Yüncü E., Başak Vural K., Mereu P., Pirastru M., Barbato M., ... Togan İ., Günther T., Özer F., Hadjisterkotis E., Somel M. Population genomic history of the endangered Anatolian and Cyprian mouflons in relation to worldwide wild, feral, and domestic sheep lineages. *Genome Biol Evol*. 2024;16(5):evae090. doi 10.1093/gbe/evae090
- Atlija M., Arranz J.J., Martínez-Valladares M., Gutiérrez-Gil B. Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. *Genet Sel Evol*. 2016;48:4. doi 10.1186/s12711-016-0182-4
- Auxiette G., Hachem L. Farm, Hunt, Feast, Celebrate. Animals and Society in Neolithic Bronze and Iron Age Northern France. Leiden: Sidestone Press, 2021
- Bakoev N.F., Getmantseva L.V., Bakoev S.Yu., Kostyunina O.V. Analysis of the nucleotide sequence of the mtDNA D-loop in Romanov sheep. *Ovcy, Kozy, Sherstyanoe Delo = Sheep, Goats, and Wool Industry*. 2018;4:18-20 (in Russian)
- Barbato M., Hailer F., Orozco-terWengel P., Kijas J., Mereu P., Cabras P., Mazza R., Pirastru M., Bruford M.W. Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Sci Rep*. 2017;7(1):7623. doi 10.1038/s41598-017-07382-7
- Barbato M., Masseti M., Pirastru M., Columbano N., Scali M., Vignani R., Mereu P. Islands as time capsules for genetic diversity conservation: the case of the Giglio island mouflon. *Diversity*. 2022; 14(8):609. doi 10.3390/d14080609
- Becker C., Benecke N., Grabundžija A., Küchelmann H.-C., Pollock S., Schier W., Schoch C., Schrakamp I., Schütt B., Schumacher M. The textile revolution. Research into the origin and spread of wool production between the Near East and Central Europe. *eTopoi J Ancient Stud*. 2016;6:102-151
- Brigand R., Dubouloz J., Weller O. Colonization dynamics of LBK farmers in Europe under geostatistics test. *Documenta Praehistorica*. 2022;49:2-45. doi 10.4312/dp.49.12
- Brigida A.V., Pridanova I.E., Mashtaler D.V. Prerequisites for the development of dairy sheep breeding in the Russian Federation (Review). *Veterinariya i Kormlenie*. 2026;7:13-16. doi 10.30917/TT-VK-1814-9588-2025-7-3 (in Russian)
- Cai D., Tang Z., Yu H., Han L., Ren X., Zhao X., Zhu H., Zhou H. Early history of Chinese domestic sheep indicated by ancient DNA analysis of Bronze Age individuals. *J Archaeol Sci*. 2011;38(4): 896-902. doi 10.1016/j.jas.2010.11.019
- Caiye Z., Song S., Li M., Huang X., Luo Y., Fang S. Genome-wide DNA methylation analysis reveals different methylation patterns in Chinese indigenous sheep with different type of tail. *Front Vet Sci*. 2023;10:1125262. doi 10.3389/fvets.2023.1125262
- Cao Y.H., Xu S.S., Shen M., Chen Z.H., Gao L., Lv F.H., Xie X.L., ... Kantanen J., Bruford M.W., Lenstra J.A., Kijas J., Li M.H. Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. *Mol Biol Evol*. 2021;38(3): 838-855. doi 10.1093/molbev/msaa236
- Chataigner C., Badalyan R., Arimura M. The Neolithic of the Caucasus. Oxford Handbooks Online, 2014. doi 10.1093/oxfordhb/978019935413.013.13
- Chen Z.H., Xu Y.X., Xie X.L., Wang D.F., Aguilar-Gómez D., Liu G.J., Li X., ... Periasamy K., Coltman D.W., Lenstra J.A., Nielsen R., Li M.H. Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. *Commun Biol*. 2021;4(1):1307. doi 10.1038/s42003-021-02817-4
- Cheng H., Zhang Z., Wen J., Lenstra J.A., Heller R., Cai Y., Guo Y., ... Zhang L., Billah M., Wang X., Liu M., Jiang Y. Long divergent haplotypes introgressed from wild sheep are associated with distinct morphological and adaptive characteristics in domestic sheep. *PLoS Genet*. 2023;19(2):e1010615. doi 10.1371/journal.pgen.1010615

- Chessa B., Pereira F., Arnaud F., Amorim A., Goyache F., Mainland I., Kao R.R., ... Ajmone-Marsan P., Bruford M.W., Kantanen J., Spencer T.E., Palmirani M. Revealing the history of sheep domestication using retrovirus integrations. *Science*. 2009;324(5926):532-536. doi 10.1126/science.1170587
- Ciani E., Lasagna E., D'Andrea M., Alloggio I., Marroni F., Ceccobelli S., Delgado Bermejo J.V., Sarti F.M., Kijas J., Lenstra J.A., Pilla F.; International Sheep Genomics Consortium. Merino and Merino-derived sheep breeds: a genome-wide intercontinental study. *Genet Sel Evol*. 2015;47(1):64. doi 10.1186/s12711-015-0139-z
- Ciani E., Mastrangelo S., Da Silva A., Marroni F., Ferenčaković M., Ajmone-Marsan P., Baird H., ... Simčič M., Stepanek O., Curik I., Cubric-Curik V., Lenstra J.A. On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. *Genet Sel Evol*. 2020;52(1):25. doi 10.1186/s12711-020-00545-7
- Crockford S.J. Animal domestication and heterochronic speciation: the role of thyroid hormone. In: Minugh-Purvis N., McNamara K. (Eds) *Human Evolution through Developmental Change*. Baltimore: Johns Hopkins University Press, 2002;122-153
- Daly K.G., Mullin V.E., Hare A.J., Halpin Á., Mattiangeli V., Teasdale M.D., Rossi C., ... Zeder M., Atıcı L., Mashkour M., Peters J., Bradley D.G. Ancient genomics and the origin, dispersal, and development of domestic sheep. *Science*. 2025;387(6733):492-497. doi 10.1126/science.adn2094
- Darwin C.R. *The Variation of Animals and Plants under Domestication*. London: John Murray, 1875
- Demirci S., Koban Baştanlar E., Dağtaş N.D., Pişkin E., Engin A., Ozer F., Yüncü E., Doğan S.A., Togan I. Mitochondrial DNA diversity of modern, ancient and wild sheep (*Ovis gmelinii anatolica*) from Turkey: new insights on the evolutionary history of sheep. *PLoS One*. 2013;8(12):e81952. doi 10.1371/journal.pone.0081952
- Deng J., Xie X.L., Wang D.F., Zhao C., Lv F.H., Li X., Yang J., ... Lenstra J.A., Han J.L., Zhang H.P., Li L., Li M.H. Paternal origins and migratory episodes of domestic sheep. *Curr Biol*. 2020;30(20):4085-4095.e6. doi 10.1016/j.cub.2020.07.077
- Deniskova T.E. Identification of QTL localized in the runs of homozygosity in sheep breeds raised in Russia. *Agrarnaya Nauka = Agrarian Science*. 2025;10:102-109. doi 10.32634/0869-8155-2025-399-10-102-109 (in Russian)
- Dossybayev K., Ualiyeva D., Kapassuly T., Amandykova M., Kozhahmet A., Bekmanov B., Amzeyev R., Naruya S. Genetic diversity and population structure of fat-tailed coarse-wooled sheep breeds *Ovis aries* from Kazakhstan. *Vet Sci*. 2025;12(10):988. doi 10.3390/vetsci12100988
- Dymova M.A., Zadorozhny A.V., Mishukova O.V., Khrapov E.A., Druzhkova A.S., Trifonov V.A., Kichigin I.G., Tishkin A.A., Grushin S.P., Filipenko M.L. Mitochondrial DNA analysis of ancient sheep from Altai. *Anim Genet*. 2017;48(5):615-618. doi 10.1111/age.12569
- Dymova M.A., Tishkin A.A., Mishukova O.V., Khrapov E.A., Grushin S.P., Filipenko M.L. Haplotypic diversity of the ancient sheep (*Ovis aries*), based on material from Western Siberian Bronze Age archaeological sites. *Zoologicheskii Zhurnal*. 2019;98(7):836-842. doi 10.1134/S0044513419070043 (in Russian)
- Ferenčaković M., Curik I., Pérez-Pardal L., Royo L.J., Cubric-Curik V., Fernández I., Álvarez I., Kostelic A., Sprem N., Krapinec K., Goyache F. Mitochondrial DNA and Y-chromosome diversity in East Adriatic sheep. *Anim Genet*. 2013;44(2):184-192. doi 10.1111/j.1365-2052.2012.02393.x
- Fitak R.R., Mohandesan E., Corander J., Yadamsuren A., Chulumbat B., Abdelhadi O., Raziq A., Nagy P., Walzer C., Faye B., Burger P.A. Genomic signatures of domestication in Old World camels. *Commun Biol*. 2020;3(1):316. doi 10.1038/s42003-020-1039-5
- Gabbianelli F., Gargani M., Pariset L., Mariotti M., Alhaique F., De Minicis E., Barelli L., Ciammetti E., Redi F., Valentini A. Mitochondrial DNA analysis of medieval sheep (*Ovis aries*) in central Italy reveals the predominance of haplogroup B already in the Middle Ages. *Anim Genet*. 2015;46(3):329-332. doi 10.1111/age.12289
- Ganbold O., Lee S.-H., Seo D., Paek W.K., Manjula P., Munkhbayer M., Lee J.H. Genetic diversity and the origin of Mongolian native sheep. *Livest Sci*. 2019;220:17-25. doi 10.1016/j.livsci.2018.12.007
- Garel M., Marchand P., Bourgoïn G., Santiago-Moreno J., Portanier E., Piegert H., Hadjisterkotis E., Cugnasse J.-M. Mouflon *Ovis gmelini* Blyth, 1841. In: Hackländer K., Zachos F.E. (Eds) *Handbook of the Mammals of Europe*. Springer, 2022;1-35. doi 10.1007/978-3-319-65038-8_34-1
- Gáspárdy A., Berger B., Zabavnik-Piano J., Kovács E., Annus K., Zenke P., Sáfár L., Maróti-Agóts Á. Comparison of mtDNA control region among descendant breeds of the extinct Zaupe sheep revealed haplogroup C and D in Central Europe. *Vet Med Sci*. 2021;7(6):2330-2338. doi 10.1002/vms3.585
- Gáspárdy A., Zenke P., Kovács E., Annus K., Posta J., Sáfár L., Maróti-Agóts Á. Evaluation of maternal genetic background of two Hungarian autochthonous sheep breeds coming from different geographical directions. *Animals (Basel)*. 2022;12(3):218. doi 10.3390/ani12030218
- Gornas N., Weimann C., El Hussien A., Erhardt G. Genetic characterization of local Sudanese sheep breeds using DNA markers. *Small Ruminant Res*. 2011;95:27-33. doi 10.1016/j.smallrumres.2010.08.009
- Guerrini M., Panayides P., Niktoresh Mofrad N., Kassinis N., Ioannou I., Forcina G., Hadjigerou P., Barbanera F. Spatial genetic structure and *Ovis* haplogroup as a tool for an adaptive conservation management of the endangered Cyprus mouflon. *Zoology*. 2021;148:125959. doi 10.1016/j.zool.2021.125959
- Guo J., Du L.X., Ma Y.H., Guan W.J., Li H.B., Zhao Q.J., Li X., Rao S.Q. A novel maternal lineage revealed in sheep (*Ovis aries*). *Anim Genet*. 2005;36(4):331-336. doi 10.1111/j.1365-2052.2005.01310.x
- Hachem L. Animals in LBK society: identity and gender markers. *J Archaeol Sci Rep*. 2018;20:910-921. doi 10.1016/j.jasrep.2017.09.020
- Han J., Yang M., Guo T., Niu C., Liu J., Yue Y., Yuan C., Yang B. Two linked *TBXT* (brachyury) gene polymorphisms are associated with the tailless phenotype in fat-rumped sheep. *Anim Genet*. 2019;50(6):772-777. doi 10.1111/age.12852
- Her C., Rezaei H.R., Hughes S., Naderi S., Duffraisie M., Mashkour M., Naghash H.R., ... Tresset A., Vigne J.D., Taberlet P., Hänni C., Pompanon F. Broad maternal geographic origin of domestic sheep in Anatolia and the Zagros. *Anim Genet*. 2022;53(3):452-459. doi 10.1111/age.13191
- Hiendleder S., Kaup B., Wassmuth R., Janke A. Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proc Biol Sci*. 2002;269(1494):893-904. doi 10.1098/rspb.2002.1975
- Horsburgh K.A., Rhines A. Genetic characterization of an archaeological sheep assemblage from South Africa's Western Cape. *J Archaeol Sci*. 2010;37(11):2906-2910. doi 10.1016/j.jas.2010.06.035
- Hu Z., Wu C., Shi Y., Guo H., Zhao X., Yin Z., Yang L., ... Zhou B., Lu D., Wu T., Lin D., Shen H. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat Genet*. 2011;43(8):792-796. doi 10.1038/ng.875
- Ibrahim A., Baliarti E., Budisatria I.G.S., Artama W.T., Widayanti R., Maharani D., Tavares L., Margawati E.T. Genetic diversity and relationship among Indonesian local sheep breeds on Java Island based on mitochondrial cytochrome b gene sequences. *J Genet Eng Biotechnol*. 2023;21(1):34. doi 10.1186/s43141-023-00491-z
- Jackson N., Maddocks I.G., Watts J.E., Scobie D., Mason R.S., Gordon-Thomson C., Stockwell S., Moore G.P.M. Evolution of the sheep coat: the impact of domestication on its structure and development. *Genet Res (Camb)*. 2020;102:e4. doi 10.1017/S0016672320000063
- Jin M., Liu G., Liu E., Wang L., Jiang Y., Zheng Z., Lu J., ... Cao J., Yuan Z., Du L., Wang H., Wei C. Genomic insights into the population history of fat-tailed sheep and identification of two mutations

- that contribute to fat tail adipogenesis. *J Adv Res.* 2026;80:725-741. doi 10.1016/j.jare.2025.05.011
- Kalaydzhiyev G., Palova N., Dundarova H., Lozanova L., Mehandjiyski I., Radoslavov G., Hristov P. Mitochondrial diversity and phylogenetic relationship of eight native Bulgarian sheep breeds. *Animals (Basel)*. 2023;13(23):3655. doi 10.3390/ani13233655
- Kaldis P., Huang S., Chen Y., Wang X. Ovine *HOXB13*: expanding the gene repertoire of sheep tail patterning and implications in genetic improvement. *Commun Biol.* 2022a;5(1):1196. doi 10.1038/s42003-022-04199-7
- Kaldis P., Zhou S., Gao Y., Cai B., Huang S., Chen Y., Wang X. Genetics of the phenotypic evolution in sheep: a molecular look at diversity-driving genes. *Genet Sel Evol.* 2022b;54(1):61. doi 10.1186/s12711-022-00753-3
- Kamalakkannan R., Kumar S., Bhavana K., Prabhu V.R., Machado C.B., Singha H.S., Sureshgopi D., Vijay V., Nagarajan M. Evidence for independent domestication of sheep mtDNA lineage A in India and introduction of lineage B through Arabian sea route. *Sci Rep.* 2021;11(1):19733. doi 10.1038/s41598-021-97761-y
- Kaptan D., Atağ G., Vural K.B., Morell Miranda P., Akbaba A., Yüncü E., Buluktaev A., ... Geigl E.M., Togan İ.Z., Günther T., Somel M., Özer F. The population history of domestic sheep revealed by paleogenomes. *Mol Biol Evol.* 2024;41(10):msae158. doi 10.1093/molbev/msae158
- Karlsson A.C., Fallahshahroudi A., Johnsen H., Hagenblad J., Wright D., Andersson L., Jensen P. A domestication related mutation in the thyroid stimulating hormone receptor gene (*TSHR*) modulates photoperiodic response and reproduction in chickens. *Gen Comp Endocrinol.* 2016;228:69-78. doi 10.1016/j.ygcen.2016.02.010
- Koshkina O.A., Deniskova T.E., Dotsev A.V., Kunz E., Upadhyay M., Krebs S., Solovieva A.D., Medugorac I., Zinovieva N.A. A study of maternal variability of Russian local sheep breeds based on the analysis of cytochrome B gene polymorphism. *Sel'skokozyajstvennaya Biologiya = Agricultural Biology.* 2021;56(6):1134-1147. doi 10.15389/agrobiology.2021.6.1134eng
- Koshkina O.A., Deniskova T.E., Zinovieva N.A. Phylogenetic analysis of complete mitochondrial genome sequences of Kuibyshev sheep. *Ovcy, Kozy, Sherstyanoe Delo = Sheep, Goats, and Wool Industry.* 2022a;4:8-11. doi 10.26897/2074-0840-2022-4-8-11 (in Russian)
- Koshkina O.A., Deniskova T.E., Zinovieva N.A. Development and implementation of test system for investigation of complete mitochondrial genomes of domestic sheep. *Dostizheniya Nauki i Tekhniki APK = Achievements of Science and Technology in Agro-industrial Complex.* 2022b;36(10):60-65. doi 10.53859/02352451_2022_36_10_60 (in Russian)
- Koshkina O., Deniskova T., Dotsev A., Kunz E., Selionova M., Medugorac I., Zinovieva N. Phylogenetic analysis of russian native sheep breeds based on mtDNA sequences. *Genes (Basel).* 2023;14(9):1701. doi 10.3390/genes14091701
- Kysely R. The size of domestic cattle, sheep, goats and pigs in the Czech Lengyel and Eneolithic periods: temporal variations and their causes. *Archaeofauna.* 2016;25:33-78. doi 10.15366/archaeofauna.2016.25.003
- Landi V., Lasagna E., Ceccobelli S., Martinez A., Santos-Silva F., Vega-Pla J.L., Panella F., ... Dunner S., Da Gama L.T., Barba C., Delgado J.V., Sarti F.M. An historical and biogeographical assessment of European Merino sheep breeds by microsatellite markers. *Small Ruminant Res.* 2019;177:76-81. doi 10.1016/j.smallrumres.2019.06.018
- Larsson M.N.A., Morell Miranda P., Pan L., Başak Vural K., Kaptan D., Rodrigues Soares A.E., Kivikero H., Kantanen J., Somel M., Özer F., Johansson A.M., Storå J., Günther T. Ancient sheep genomes reveal four millennia of North European short-tailed sheep in the Baltic sea region. *Genome Biol Evol.* 2024;16(6):evae114. doi 10.1093/gbe/evae114
- Li J.Z., Absher D.M., Tang H., Southwick A.M., Casto A.M., Ramachandran S., Cann H.M., Barsh G.S., Feldman M., Cavalli-Sforza L.L., Myers R.M. Worldwide human relationships inferred from genome-wide patterns of variation. *Science.* 2008;319(5866):1100-1104. doi 10.1126/science.1153717
- Li R., Yang P., Li M., Fang W., Yue X., Nanaei H.A., Gan S., Du D., Cai Y., Dai X., Yang Q., Cao C., Deng W., He S., Li W., Ma R., Liu M., Jiang Y. A Hu sheep genome with the first ovine Y chromosome reveal introgression history after sheep domestication. *Sci China Life Sci.* 2021;64(7):1116-1130. doi 10.1007/s11427-020-1807-0
- Li X., Yang J., Shen M., Xie X.L., Liu G.J., Xu Y.X., Lv F.H., ... Kijas J.W., Bruford M.W., Periasamy K., Wang X.H., Li M.H. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat Commun.* 2020;11(1):2815. doi 10.1038/s41467-020-16485-1
- Liu J., Ding X., Zeng Y., Yue Y., Guo X., Guo T., Chu M., ... Sun X., Niu C., Yang B., Guo J., Yuan C. Genetic diversity and phylogenetic evolution of Tibetan sheep based on mtDNA D-loop sequences. *PLoS One.* 2016;11(7):e0159308. doi 10.1371/journal.pone.0159308
- Luo Y., Zhang M., Guo Z., Wijayanti D., Xu H., Jiang F., Lan X. Insertion/Deletion (InDel) variants within the sheep fat-deposition-related *PDGFD* gene strongly affect morphological traits. *Animals (Basel).* 2023;13(9):1485. doi 10.3390/ani13091485
- Luong N.N., Ha H.T.T., Huy N.X., Loi B.V., Van N.H., Quang H.T., Loc N.H. Characterizing the Phan Rang sheep: a first look at the Y chromosome, mitochondrial DNA, and morphometrics. *Animals (Basel).* 2024;14(14):2020. doi 10.3390/ani14142020
- Lv F.-H., Peng W.-F., Yang J., Zhao Y.-X., Li W.-R., Liu M.-J., Ma Y.-H., ... Iso-Touru T., Bruford M.W., Kantanen J., Han J.L., Li M.H. Mitogenomic meta-analysis identifies two phases of migration in the history of eastern Eurasian sheep. *Mol Biol Evol.* 2015;32(10):2515-2533. doi 10.1093/molbev/msv139
- Lv F.-H., Cao Y.H., Liu G.J., Luo L.Y., Lu R., Liu M.J., Li M.H., ... Kantanen J., Coltman D.W., Bruford M.W., Lenstra J.A., Li M.H. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, and agronomically important loci. *Mol Biol Evol.* 2022;39(2):msab353. doi 10.1093/molbev/msab353
- Machová K., Málková A., Vostrý L. Sheep post-domestication expansion in the context of mitochondrial and Y chromosome haplogroups and haplotypes. *Genes (Basel).* 2022;13(4):613. doi 10.3390/genes13040613
- Matyukov V.S., Zharikov Ya.A., Kaneva L.A. Analysis of the allele fund of semi-fine wool sheep of the Pechora population using STR markers. *Russ J Genet.* 2023;59(7):732-738. doi 10.1134/s1022795423060108
- McManus C., Pimentel D., Junqueira V.S., Pimentel F., Peripolli V. The importance of sheep genetics and diversity: a bibliometric review. *Trop Anim Health Prod.* 2025;57(5):261. doi 10.1007/s11250-025-04503-7
- Meadows J.R., Li K., Kantanen J., Tapio M., Sipos W., Pardeshi V., Gupta V., Calvo J.H., Whan V., Norris B., Kijas J.W. Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. *J Hered.* 2005;96(5):494-501. doi 10.1093/jhered/esi100
- Meadows J.R., Cemal I., Karaca O., Gootwine E., Kijas J.W. Five ovine mitochondrial lineages identified from sheep breeds of the near East. *Genetics.* 2007;175(3):1371-1379. doi 10.1534/genetics.106.068353
- Mereu P., Pirastru M., Sanna D., Bassu G., Naitana S., Leoni G.G. Phenotype transition from wild mouflon to domestic sheep. *Genet Sel Evol.* 2024;56:1. doi 10.1186/s12711-023-00871-6
- Mereu P., Pirastru M., Morell Miranda P., Atağ G., Basak Vural K., Wilkens B., Rodrigues Soares A.E., ... Hadjisterkotis E., Somel M., Özer F., Günther T., Leoni G.G. Revised phylogeny of mouflon based on expanded sampling of mitogenomes. *PLoS One.* 2025;20(5):e0323354. doi 10.1371/journal.pone.0323354
- Michailidou S., Kyritsi M., Pavlou E., Tsourekis A., Argiriou A. Genetic diversity, population structure, and historical gene flow patterns of

- nine indigenous Greek sheep breeds. *Biology*. 2025;14(7):845. doi 10.3390/biology14070845
- Molae V., Bazzucchi M., De Mia G.M., Otarod V., Abdollahi D., Rosati S., Lühken G. Phylogenetic analysis of small ruminant lentiviruses in Germany and Iran suggests their expansion with domestic sheep. *Sci Rep*. 2020;10(1):2243. doi 10.1038/s41598-020-58990-9
- Muigai A.W.T., Hanotte O. The origin of African sheep: archaeological and genetic perspectives. *Afr Archaeol Rev*. 2013;30:39-50. doi 10.1007/s10437-013-9129-0
- Naval-Sanchez M., Nguyen Q., McWilliam S., Porto-Neto L.R., Telam R., Vuocolo T., Reverter A., ... Muzny D.M., Jhangiani S.N., Cockett N., Daetwyler H., Kijas J. Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. *Nat Commun*. 2018;9(1):859. doi 10.1038/s41467-017-02809-1
- Niemi M., Bläuer A., Iso-Touru T., Nyström V., Harjula J., Taavitsainen J.P., Stora J., Lidén K., Kantanen J. Mitochondrial DNA and Y-chromosomal diversity in ancient populations of domestic sheep (*Ovis aries*) in Finland: comparison with contemporary sheep breeds. *Genet Sel Evol*. 2013;45(1):2. doi 10.1186/1297-9686-45-2
- Pedrosa S., Uzun M., Arranz J.-J., Gutiérrez-Gil B., San Primitivo F., Bayón Y. Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. *Proc Biol Sci*. 2005;272(1577):2211-2217. doi 10.1098/rspb.2005.3204
- Pereira F., Davis S.J., Pereira L., McEvoy B., Bradley D.G., Amorim A. Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. *Mol Biol Evol*. 2006;23(7):1420-1426. doi 10.1093/molbev/msl007
- Poplin F. Origine du Mouflon de Corse dans une nouvelle perspective paléontologique: par marronnage. *Ann Genet Sel Anim*. 1979;11(2):133-143. doi 10.1186/1297-9686-11-2-133 (in French)
- Portanier E., Chevret P., Gélén P., Benedetti P., Sanchis F., Barbanera F., Kaerle C., Queney G., Bourgoin G., Devillard S., Garel M. New insights into the past and recent evolutionary history of the Corsican mouflon (*Ovis gmelini musimon*) to inform its conservation. *Conserv Genet*. 2022;23:91-107. doi 10.1007/s10592-021-01399-2
- Racimo F., Woodbridge J., Fyfe R.M., Sikora M., Sjögren K.-G., Kristiansen K., Vander Linden M. The spatiotemporal spread of human migrations during the European Holocene. *Proc Natl Acad Sci USA*. 2020;117(16):8989-9000. doi 10.1073/pnas.1920051117
- Rafia P., Tarang A. Sequence variations of mitochondrial DNA displacement-loop in Iranian indigenous sheep breeds. *Iran J Appl Anim Sci*. 2016;6:363-368
- Resende A., Gonçalves J., Muigai A.W., Pereira F. Mitochondrial DNA variation of domestic sheep (*Ovis aries*) in Kenya. *Anim Genet*. 2016;47(3):377-381. doi 10.1111/age.12412
- Rezaei H.R., Naderi S., Chintauan-Marquier I.C., Jordan S., Taberlet P., Virk A.T., Naghash H.R., Rioux D., Kaboli M., Luikart G., Pompanon F. Evolution and taxonomy of the wild species of the genus *Ovis* (Mammalia, Artiodactyla, Bovidae). *Mol Phylogenet Evol*. 2010;54(2):315-326. doi 10.1016/j.ympev.2009.10.037
- Ryder M.L. Sheep. In: Manson I.L. (Ed.) *Evolution of Domesticated Animals*. London; New York: Longman Group Limited, 1984;63-84
- Sabatini S., Bergerbrant S., Brandt L.O., Margaryan A., Allentoft M.E. Approaching sheep herds origins and the emergence of the wool economy in continental Europe during the Bronze Age. *Archaeol Anthropol Sci*. 2019;11:4909-4925. doi 10.1007/s12520-019-00856-x
- Salim B., Alasmari S., Mohamed N.S., Ahmed M.A., Nakao R., Hanotte O. Genetic variation and demographic history of Sudan desert sheep reveal two diversified lineages. *BMC Genomics*. 2023;24(1):118. doi 10.1186/s12864-023-09231-6
- Sandoval-Castellanos E., Hare A.J., Lin A.T., Dimopoulos E.A., Daly K.G., Geiger S., Mullin V.E., ... Burger J., Frantz L., Medugorac I., Bradley D.G., Peters J. Ancient mitogenomes from Pre-Pottery Neolithic Central Anatolia and the effects of a Late Neolithic bottleneck in sheep (*Ovis aries*). *Sci Adv*. 2024;10(15):eadj0954. doi 10.1126/sciadv.adj0954
- Sanna D., Barbato M., Hadjisterkotis E., Cossu P., Decandia L., Trova S., Pirastru M., Leoni G.G., Naitana S., Francalacci P., Masala B., Manca L., Mereu P. The first mitogenome of the Cyprus mouflon (*Ovis gmelini ophion*): new insights into the phylogeny of the genus *Ovis*. *PLoS One*. 2015;10(12):e0144257. doi 10.1371/journal.pone.0144257
- Simon R., Drögemüller C., Lühken G. The complex and diverse genetic architecture of the absence of horns (polledness) in domestic ruminants, including goats and sheep. *Genes (Basel)*. 2022;13(5):832. doi 10.3390/genes13050832
- Shirokova N.V., Bakoev N.F., Kolosova M.A., Getmantseva L.V., Kolosov Yu.A. Study of the nucleotide sequence of the mtDNA D-loop in Merino sheep of the Russian Federation. *Mezhdunarodnyi Nauchno-Issledovatel'skij Zhurnal = International Research Journal*. 2018;11(77):21-25. doi 10.23670/IRJ.2018.77.11.039 (in Russian)
- Singh S., Kumar S. Jr., Kolte A.P., Kumar S. Extensive variation and sub-structuring in lineage A mtDNA in Indian sheep: genetic evidence for domestication of sheep in India. *PLoS One*. 2013;8(11):e77858. doi 10.1371/journal.pone.0077858
- Spangler G.L., Rosen B.D., Ilori M.B., Hanotte O., Kim E.S., Sonstegard T.S., Burke J.M., Morgan J.L.M., Notter D.R., Van Tassell C.P. Whole genome structural analysis of Caribbean hair sheep reveals quantitative link to West African ancestry. *PLoS One*. 2017;12(6):e0179021. doi 10.1371/journal.pone.0179021
- Tapio M., Marzanov N., Ozerov M., Cinkulov M., Gonzarenko G., Kiselyova T., Murawski M., Viinalass H., Kantanen J. Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. *Mol Biol Evol*. 2006;23(9):1776-1783. doi 10.1093/molbev/msl043
- Tarlykov P., Atavlijeva S., Auganova D., Akhmetollayev I., Loshakova T., Varfolomeev V., Ramankulov Y. Mitochondrial DNA analysis of ancient sheep from Kazakhstan: evidence for early sheep introduction. *Heliyon*. 2021;7(9):e08011. doi 10.1016/j.heliyon.2021.e08011
- Taylor W.T., Pruvost M., Posth C., Rendu W., Krajcarz M.T., Abdykanova A., Brancaloni G., ... Douka K., Krivoschapkin A., Jeong C., Warinner C., Shnaider S. Evidence for early dispersal of domestic sheep into Central Asia. *Nat Hum Behav*. 2021;5(9):1169-1179. doi 10.1038/s41562-021-01083-y
- Teletchea F., Fontaine P. Levels of domestication in fish: implications for the sustainable future of aquaculture. *Fish Fish*. 2014;15(2):181-195. doi 10.1111/faf.12006
- Trigo B.B., Utsunomiya A.T.H., Fortunato A.A.A.D., Milanesi M., Torrecilha R.B.P., Lamb H., Nguyen L., ... Bickhart D., Smith T.P.L., Sonstegard T.S., Garcia J.F., Utsunomiya Y.T. Variants at the ASIP locus contribute to coat color darkening in Nellore cattle. *Genet Sel Evol*. 2021;53(1):40. doi 10.1186/s12711-021-00633-2
- Vigne J.-D., Carrère I., Briois F., Guilaine J. The early process of mammal domestication in the Near East. New evidence from the Pre-Neolithic and Pre-Pottery Neolithic in Cyprus. *Curr Anthropol*. 2011;52(S4):255-271. doi 10.1086/659306
- Wang D.-F., Orozco-terWengel P., Salehian-Dehkordi H., Esmailizadeh A., Lv F.-H. Genomic analyses of Asiatic mouflon in Iran provide insights into the domestication and evolution of sheep. *Genet Sel Evol*. 2025;57:31. doi 10.1186/s12711-025-00978-y
- Wanjala G., Bagi Z., Kusza S. Meta-analysis of mitochondrial DNA control region diversity to shed light on phylogenetic relationship and demographic history of African sheep (*Ovis aries*) breeds. *Biology (Basel)*. 2021;10(8):762. doi 10.3390/biology10080762
- Wilkins A.S. A striking example of developmental bias in an evolutionary process: the “domestication syndrome”. *Evol Dev*. 2020;22(1-2):143-153. doi 10.1111/ede.12319
- Wilkins A.S., Wrangham R.W., Fitch W.T. The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*. 2014;197(3):795-808. doi 10.1534/genetics.114.165423

- Wood N.J., Phua S.H. Variation in the control region sequence of the sheep mitochondrial genome. *Anim Genet.* 1996;27(1):25-33. doi 10.1111/j.1365-2052.1996.tb01173.x
- Yang J., Li W.R., Lv F.H., He S.G., Tian S.L., Peng W.F., Sun Y.W., ... Lu H.F., Kantanen J., Han J.L., Li M.H., Liu M.J. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol Biol Evol.* 2016;33(10):2576-2592. doi 10.1093/molbev/msw129
- Yang L., Zhang X., Hu Y., Zhu P., Li H., Peng Z., Xiang H., Zhou X., Zhao X. Ancient mitochondrial genome depicts sheep maternal dispersal and migration in Eastern Asia. *J Genet Genomics.* 2024; 51(1):87-95. doi 10.1016/j.jgg.2023.06.002
- Yurtman E., Özer O., Yüncü E., Dağtaş N.D., Koptekin D., Çakan Y.G., Özkan M., ... Özbal R., Götherström A., Somel M., Togan İ., Özer F. Archaeogenetic analysis of Neolithic sheep from Anatolia suggests a complex demographic history since domestication. *Commun Biol.* 2021;4(1):1279. doi 10.1038/s42003-021-02794-8
- Zeder M.A. Core questions in domestication research. *Proc Natl Acad Sci USA.* 2015;112(11):3191-3198. doi 10.1073/pnas.1501711112
- Zeder M.A. Out of the Fertile Crescent: the dispersal of domestic livestock through Europe and Africa. In: *Human Dispersal and Species Movement.* Cambridge University Press, 2017;261-303. doi 10.1017/9781316686942.012
- Zhang L., Liang L., Zhang X., Li W.R., Li H., He S. Understanding key genetic make-up of different coat colour in Bayinbuluke sheep through a comparative transcriptome profiling analysis. *Small Ruminant Res.* 2023;226:107028. doi 10.1016/j.smallrumres.2023.107028
- Zhao E., Yu Q., Zhang N., Kong D., Zhao Y. Mitochondrial DNA diversity and the origin of Chinese indigenous sheep. *Trop Anim Health Prod.* 2013;45(8):1715-1722. doi 10.1007/s11250-013-0420-5
- Zhao Y.X., Yang J., Lv F.-H., Hu X.J., Xie X.L., Zhang M., Li W.R., ... Hehua E., Kantanen J., Arjen Lenstra J., Han J.L., Li M.H. Genomic reconstruction of the history of native sheep reveals the peopling patterns of nomads and the expansion of early pastoralism in East Asia. *Mol Biol Evol.* 2017;34(9):2380-2395. doi 10.1093/molbev/msx181

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

Received February 9, 2026. Revised March 12, 2026. Accepted March 12, 2026.

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 СМЭ 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, ВИНТИ, 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/>

При перепечатке материалов ссылка обязательна.

✉ 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 30.06.2026. Формат 60 × 84 1/8. Уч.-изд. л. 23.6.