2025 • 29 • 6

Online ISSN 2500-3259

Сетевое издание

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

VAVILOV JOURNAL OF GENETICS AND BREEDING

Основан в 1997 г.
Периодичность 8 выпусков в год
doi 10.18699/vjgb-25-80

2025 • 29 • 6

Online ISSN 2500-3259

Online edition

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

VAVILOV JOURNAL OF GENETICS AND BREEDING

Founded in 1997
Publication frequency: 8 issues a year
doi 10.18699/vjgb-25-80

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

СОДЕРЖАНИЕ • 2025 • 29 • 6

Молекулярная и клеточная биология

737 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Новая комбинация 5′- и 3′-нетранслируемых областей способствует повышению экспрессии мРНК *in vitro* и *in vivo*. *Д.Н. Антропов, О.В. Марков, А.С. Доме, П.А. Пучков, Е.В. Шмендель, Д.В. Гладких, В.М. Голышев, А.М. Матвеева, М.А. Маслов, Г.А. Степанов (на англ. языке)*

744 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Интерактом белка HOXB13 в клетках рака простаты: биохимические и функциональные взаимодействия между транскрипционными факторами HOXB13 и TBX3. *М.М. Ерохин, Н.Я. Козельчук, Р.Х. Зиганшин, В.В. Татарский, Д.А. Четверина*

Генетика растений

753 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Кариологическая дифференциация среди сортов мягкой пшеницы (*Triticum aestivum* L.), контрастных по селекционному статусу и типу развития. *А.Ф. Мутерко, Е.Д. Бадаева, Е.В. Зуев, Е.А. Салина*

769 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Картирование локусов и генов, определяющих время колошения и созревания яровой мягкой пшеницы в условиях длинного дня, и оценка их влияния на урожайность. *А.А. Киселёва, А.И. Стасюк, И.Н. Леонова, Е.А. Салина*

Селекция растений на иммунитет и качество

779 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Пребридинговые исследования почти изогенных линий яровой мягкой пшеницы, отличающихся по наличию/отсутствию хромосомного замещения 3R(3D) от сорта тритикале Satu. *С.Н. Сибикеев, И.Г. Адонина, А.Е. Дружин, З.Е. Фитилева, О.А. Баранова*

789 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Вариабельность минерального состава зерна твердой пшеницы (*Triticum durum* Desf.) в различных экологических условиях. *И.Н. Леонова, П.Н. Мальчиков, Н.А. Виниченко, В.В. Пискарев, М.Г. Мясникова, В.А. Апарина, Т.В. Чахеева*

Генетика животных

798 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
К генетической структуре и происхождению малого суслика *Spermophilus pygmaeus* (Pallas, 1778) на Северном Кавказе. *Ф.А. Темботова, А.Х. Амшокова, М.С. Гудова*

812 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Метаболические эффекты трегалозы у мышей линии C57BL/6 с ожирением, вызванным диетой с высоким содержанием углеводов и жиров. *А.Б. Пупышев, Н.М. Бажан, А.Ю. Казанцева, Т.В. Яковлева, В.М. Беличенко, Н.В. Гончарова, Т.А. Короленко, М.А. Тихонова*

Медицинская генетика

819 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Ассоциация варианта rs823144 гена *RAB29* с активностью лизосомных гидролаз в клетках крови и риском болезни Паркинсона. *К.С. Башарова, А.И. Безрукова, К.А. Сенкевич, Г.В. Байдакова, А.В. Рыбаков, И.В. Милюхина, А.А. Тимофеева, Е.Ю. Захарова, С.Н. Пчелина, Т.С. Усенко*

828 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Вклад генов транскрипционных факторов семейства FOXO (*FOXO1*, *FOXO3*) в развитие хронической обструктивной болезни легких. *В.А. Маркелов, Л.З. Ахмадишина, Т.Р. Насибуллин, Ю.Г. Азнабаева, О.В. Кочетова, Н.Н. Хуснутдинова, С.М. Измайлова, Н.Ш. Загидуллин, Г.Ф. Корытина*

Генетика поведения

838 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Использование полигенных показателей для оценки предрасположенности к манифестации антисоциального поведения. *А.В. Казанцева, Д.В. Яковлева, Ю.Д. Давыдова, Э.К. Хуснутдинова (на англ. языке)*

847 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Эндогенный окситоцин и межсамцовые взаимодействия после введения окситоцина у серых крыс, селекционируемых по поведению. *С.Г. Шихевич, Р.В. Кожемякина, Р.Г. Гулевич, Ю.Э. Гербек*

Филогенетика и таксономия

856 ОБЗОР
Последовательности ДНК агробактериального происхождения в филогенетических исследованиях растений. *Т.В. Матвеева, П.М. Журбенко, Г.В. Хафизова, А.Д. Шапошников, Р.Р. Жидкин, А.В. Родионов*

868 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Новый молекулярный маркер для филогенетических исследований стрекоз (Insecta, Odonata), включающий части консервативных генов гистонов H3 и H4 и спейсер между ними, применимый и к другим организмам. *А.В. Мглинец, В.С. Булгакова, О.Э. Костерин (на англ. языке)*

883 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Полные последовательности пластомов видов *Lonicera* L.: значение для филогении и сравнительный анализ. *Ш.С. Альмерекова, М.М. Ермагамбетова, Д.Е. Ерболатов, М.Ю. Ишмуратова, Е.К. Туруспеков (на англ. языке)*

Актуальные технологии

896 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Увеличение доли дальнего красного света сокращает вегетационный период тритикале в условиях спидбридинга. *А.О. Блинков, В.М. Нагамова, Я.В. Минькова, Н.Ю. Свистунова, С. Радзенице, А.А. Кочешкова, Н.Н. Слепцов, А.В. Фрейманс, В.В. Панченко, А.Г. Черноок, Г.И. Карлов, М.Г. Дивашук*

CONTENTS • 2025 • 29 • 6

Molecular and cell biology

737 ORIGINAL ARTICLE
A new combination of 5′- and 3′-untranslated regions increases the expression of mRNAs *in vitro* and *in vivo*. *D.N. Antropov, O.V. Markov, A.S. Dome, P.A. Puchkov, E.V. Shmendel, D.V. Gladkikh, V.M. Golyshev, A.M. Matveeva, M.A. Maslov, G.A. Stepanov*

744 ORIGINAL ARTICLE
HOXB13 interactome in prostate cancer cells: biochemical and functional interactions between the transcription factors HOXB13 and TBX3. *M.M. Erokhin, N.Y. Kozelchuk, R.H. Ziganshin, V.V. Tatarskiy, D.A. Chetverina*

Plant genetics

753 ORIGINAL ARTICLE
Karyological differentiation among bread wheat cultivars (*Triticum aestivum* L.) with distinct breeding statuses and growth habits. *A.F. Muterko, E.D. Badaeva, E.V. Zuev, E.A. Salina*

769 ORIGINAL ARTICLE
Mapping loci and genes controlling heading and maturity time in common wheat under long-day conditions and assessing their effects on yield-related traits. *A.A. Kiseleva, A.I. Stasyuk, I.N. Leonova, E.A. Salina*

Plant breeding for immunity and quality

779 ORIGINAL ARTICLE
Prebreeding studies of near-isogenic spring bread wheat lines, differing by presence or absence of the 3R(3D) chromosomal substitution from the triticale cultivar Satu. *S.N. Sibikeev, I.G. Adonina, A.E. Druzhin, Z.E. Fitileva, O.A. Baranova*

789 ORIGINAL ARTICLE
Variability of the mineral composition of durum wheat grain (*Triticum durum* Desf.) under different environmental conditions. *I.N. Leonova, P.N. Malchikov, N.A. Vinichenko, V.V. Piskarev, M.G. Myasnikova, V.A. Aparina, T.V. Chaheeva*

Animal genetics

798 ORIGINAL ARTICLE
On the genetic structure and origin of the little ground squirrel *Spermophilus pygmaeus* (Pallas, 1778) in the North Caucasus. *F.A. Tembotova, A.Kh. Amshokova, M.S. Gudova*

812 ORIGINAL ARTICLE
Metabolic effects of trehalose in mice of the C57BL/6 strain with obesity induced by a high carbohydrate-fat diet. *A.B. Pupyshev, N.M. Bazhan, A.Yu. Kazantseva, T.V. Yakovleva, V.M. Belichenko, N.V. Goncharova, T.A. Korolenko, M.A. Tikhonova*

Medical genetics

819 ORIGINAL ARTICLE
Association of the rs823144 variant of the *RAB29* gene with the activity of lysosomal hydrolases in blood cells and risk of Parkinson's disease. *K.S. Basharova, A.I. Bezrukova, K.A. Senkevich, G.V. Baydakova, A.V. Rybakov, I.V. Miliukhina, A.A. Timofeeva, E.Yu. Zakharova, S.N. Pchelina, T.S. Usenko*

828 ORIGINAL ARTICLE
The contribution of FOXO family transcription factor genes (*FOXO1*, *FOXO3*) to chronic obstructive pulmonary disease. *V.A. Markelov, L.Z. Akhmadishina, T.R. Nasibullin, Y.G. Aznabaeva, O.V. Kochetova, N.N. Khusnutdinova, S.M. Izmailova, N.Sh. Zagidullin, G.F. Korytina*

Behavioral genetics

838 ORIGINAL ARTICLE
Using polygenic scores to assess liability to antisocial behavior. *A.V. Kazantseva, D.V. Yakovleva, Yu.D. Davydova, E.K. Khusnutdinova*

847 ORIGINAL ARTICLE
Endogenous oxytocin and intermale interactions after oxytocin administrations in Norway rats selected for behavior. *S.G. Shikevich, R.V. Kozhemyakina, R.G. Gulevich, Yu.E. Herbeck*

Phylogenetics and taxonomy

856 REVIEW
Agrobacterium-derived DNA sequences in phylogenetic studies of plants. *T.V. Matveeva, P.M. Zhurbenko, G.V. Khafizova, A.D. Shaposhnikov, R.R. Zhidkin, A.V. Rodionov*

868 ORIGINAL ARTICLE
A new molecular marker including parts of conservative histone H3 and H4 genes and the spacer between them for phylogenetic studies in dragonflies (Insecta, Odonata), extendable to other organisms. *A.V. Mglinets, V.S. Bulgakova, O.E. Kosterin*

883 ORIGINAL ARTICLE
Complete plastome sequences of *Lonicera* L. species: implications for phylogeny and comparative analysis. *S.S. Almerekova, M.M. Yermagambetova, D.Y. Yerbolatov, M.Y. Ishmuratova, Y.K. Turuspekov*

Mainstream technologies

896 ORIGINAL ARTICLE
A higher far-red intensity promotes the transition to flowering in triticale grown under speed breeding conditions. *A.O. Blinkov, V.M. Nagamova, Y.V. Minkova, N.Yu. Svistunova, S. Radzeniece, A.A. Kocheshkova, N.N. Sleptsov, A.V. Freymans, V.V. Panchenko, A.G. Chernook, G.I. Karlov, M.G. Divashuk*

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):737-743

doi 10.18699/vjgb-25-81

A new combination of 5′- and 3′-untranslated regions increases the expression of mRNAs *in vitro* and *in vivo*

D.N. Antropov[1], O.V. Markov[1], A.S. Dome[1], P.A. Puchkov[2], E.V. Shmendel[2], D.V. Gladkikh[1], V.M. Golyshev[1], A.M. Matveeva[1], M.A. Maslov[2], G.A. Stepanov [1]

[1] Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Lomonosov Institute of Fine Chemical Technologies, MIREA – Russian Technological University, Moscow, Russia
stepanovga@niboch.nsc.ru

Abstract. mRNA vaccine technologies have been actively developing since the beginning of the 21st century and have received a major boost from new findings about the functioning of the immune system and the development of efficient vehicles for nucleic acid delivery. The mRNA vaccine demonstrates superior properties compared to the DNA vaccine, primarily due to accelerated mRNA vaccine development, enhanced flexibility, and the absence of integration into the genome. Artificial mRNAs have biotechnological and medical applications, including the development of antiviral and anticancer mRNA therapeutics. The effective expression of therapeutic mRNA depends upon the appropriate selection of structural elements. Along with the addition of the 5′-cap, appropriate polyadenylation, and sequence codon optimization, 5′- and 3′-untranslated regions (UTRs) play an important role in the translation efficiency of therapeutic mRNAs. In this study, new plasmids containing a novel combination of UTR pairs, namely 5′-UTR-4 and 3′-UTR AES-mtRNR1, were constructed to obtain artificial mRNAs encoding green fluorescent protein (GFP) and firefly luciferase (FLuc) with new structural elements and properties. The novel combination of the UTRs, which is described in this article for the first time, in addition to sufficient polyadenylation and pseudouridinilation of mRNA, was demonstrated to strongly increase the translation of codon-optimized sequences of reporter mRNAs. We generated lipoplexes containing the aforementioned reporter mRNAs and liposomes composed of cationic lipid 2X3 (1,26-bis(cholest-5-en-3beta-yloxycarbonylamino)-7,11,16,20-tetraazahexacosane tetrahydrochloride) and helper lipid DOPE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine). For *in vivo* experiments, the liposomes were decorated with 2 % of 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[amino(polyethylene glycol)-2000] (DSPE-PEG_{2000}). The translation efficiency of mRNAs was found to be superior for the novel UTR combination compared with HBB gene UTRs, both *in vitro* and *in vivo*. When mRNA is administered intramuscularly, the proposed combination of UTRs provides lasting expression for more than 4 days. The results demonstrated that the novel UTR pair combination could be useful in the development of artificial mRNAs with enhanced translation efficiency, potentially reducing the dose for mRNA-based therapeutics.

Key words: synthetic mRNA; RNA delivery; nucleotide modifications; untranslated region; lipid nanoparticle

For citation: Antropov D.N., Markov O.V., Dome A.S., Puchkov P.A., Shmendel E.V., Gladkikh D.V., Golyshev V.M., Matveeva A.M., Maslov M.A., Stepanov G.A. A new combination of 5′- and 3′-untranslated regions increases the expression of mRNAs *in vitro* and *in vivo*. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):737-743. doi 10.18699/vjgb-25-81

Funding. This research was funded by the Russian Science Foundation (grant number 22-75-10153 for physicochemical LNP characterization, mRNA construction and synthesis, and *in vitro* transfection and *in vivo* experiments; 23-73-10168 for lipid synthesis and liposome preparation).

Новая комбинация 5′- и 3′-нетранслируемых областей способствует повышению экспрессии мРНК *in vitro* и *in vivo*

Д.Н. Антропов[1], О.В. Марков[1], А.С. Доме[1], П.А. Пучков[2], Е.В. Шмендель[2], Д.В. Гладких[1], В.М. Голышев[1], А.М. Матвеева[1], М.А. Маслов[2], Г.А. Степанов [1]

[1] Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Институт тонких химических технологий им. М.В. Ломоносова, МИРЭА – Российский технологический университет, Москва, Россия
stepanovga@niboch.nsc.ru

Аннотация. Технология мРНК-вакцин начала активно развиваться в начале XXI в. и получила хороший стимул за счет расширения знаний о функционировании иммунной системы человека и успехов в синтезе вариантов молекул-доставщиков. Иммунизация с помощью мРНК-вакцин является более эффективной, чем иммунизация с помощью ДНК, благодаря более быстрой разработке, гибкости технологии и отсутствию интеграции в геном. В наши дни искусственные мРНК используют в различных биотехнологических и медицин-

ских целях, включая разработку противовирусных и противораковых мРНК-вакцин. Для их эффективной экспрессии необходимо правильно подобрать структурные элементы мРНК. Помимо добавления в структуру мРНК 5′-кэпа, достаточного уровня полиаденилирования и оптимизации последовательности кодонов, 5′- и 3′-нетранслируемые области (НТО) играют важную роль в трансляционной эффективности терапевтических мРНК. В настоящем исследовании для получения искусственных мРНК были сконструированы плазмидные конструкции, содержащие в своем составе новую комбинацию нетранслируемых областей – 5′-UTR-4 и 3′-UTR AES-mtRNR1. Для новой комбинации НТО, впервые описанной в данной работе, было показано значительное увеличение уровня трансляции кодон-оптимизированных последовательностей репортерных мРНК, кодирующих GFP (зеленый флюоресцентный белок) и FLuc (люцифераза светлячка), содержащих в своем составе псевдоуридин и поли(А)-последовательность. В ходе работы были сформированы комплексы вышеупомянутых репортерных мРНК с липосомами, состоящими из катионного липида 2X3 (1,26-бис(холест-5-ен-3β-илоксикарбониламино)-7,11,16,20-тетраазогексакозан тетрагидрохлорид) и липида-хелпера DOPE (1,2-диолеил-sn-глицеро-3-фосфоэтаноламин). Для экспериментов *in vivo* в состав липосомальной композиции добавляли 2 % 1,2-дистеароил-sn-глицеро-3-фосфоэтаноламин-N-[амино(полиэтиленгликоль)-2000] (DSPE-PEG_{2000}). Новая комбинация НТО продемонстрировала более высокую эффективность трансляции мРНК в сравнении с β-глобиновыми НТО как *in vitro*, так и *in vivo*. При внутримышечном введении мРНК предложенная комбинация НТО обеспечивает длительную экспрессию более четырех суток. Результаты исследования показали высокую эффективность новой комбинации НТО для повышения уровня трансляции искусственных мРНК, что может быть использовано для снижения терапевтической дозы мРНК в составе вакцин.
Ключевые слова: синтетическая мРНК; доставка РНК; модификации нуклеотидов; нетранслируемая область мРНК; липидные наночастицы

Introduction

mRNA vaccine technologies are emerging every year to defend humans from viral pathogens and even cancers. The COVID-19 pandemic proved the necessity of the fast development of vaccines targeted against a certain species of viruses. The release of different mRNA vaccines, such as BNT162b2, mRNA-1273, and others, has enabled effective vaccination of the population. It is clear that for effective and prolonged expression of antibodies, specific structure of mRNA is crucial (Fig. 1).

The necessity of UTRs in mRNA for various mRNA models is widely recognized (Chatterjee, Pal, 2009; Morais et al., 2021; Chen et al., 2022; Kirshina et al., 2023), along with such components as the Kozak sequence, which is required for translation (Kozak, 1989). The incorporation of 5′-UTR sequences enhances translation not only by protecting the coding sequences from nucleases but also by recruiting ribosomal machinery to a translation start site (Chatterjee, Pal, 2009). The most common method for enhancing translation is the addition of IRES (internal ribosome entry site) elements with complex secondary structure to the 5′ end of mRNA, which promotes the recruitment of translational complexes at various stages of translation (Andreev et al., 2009). Sample et al. (2019) have successfully identified a high-ribosome-load sequence of a 5′-UTR (designated as "5′-UTR-4") using machine learning methods. The application of computationally selected sequences enabled the researchers to enhance translation intensity more than 100-fold, thus proving the efficacy of machine learning techniques.

3′-UTRs provide the mRNA molecule with a defense against nucleases, including the prevention of poly(A) tail degradation, which can influence the half-life of mRNA and the yield of the encoded protein. Derived from the human HBB gene, β-globin UTRs have been commonly used and demonstrated remarkable efficacy in both reporter and viral mRNA applications (Zhuang et al., 2020). The incorporation of a chimeric 3′-UTR called AES-mtRNR1, which comprises a part of 16S ribosomal rRNA of archaea (AES: amino-terminal enhancer of split) and a part of a mitochondrial rRNA (mtRNR1), allowed for 3-fold higher translation intensity *in vitro* in comparison to the standard β-globin 3′-UTR (Orlandini von Niessen et al., 2019). Their result was confirmed by *in vivo* application to a luciferase model (Orlandini von Niessen et al., 2019), indicating the universality of such a tandem UTR sequence. Thus, the previous investigations directed at the selection of the most effective UTR structures allow to design combinations and test their translational effectiveness.

In this study, we introduced a novel combination of 5′-UTR-4 and 3′-UTR AES-mtRNR1 (hereafter, "synthetic UTRs") into a single mRNA for the first time. We constructed reporter mRNAs, encoding GFP or FLuc, containing the novel combination of UTRs for delivery into human cells *in vitro* and live animal tissues *in vivo*. Cationic liposomes composed of the cationic lipid 2X3, the helper lipid DOPE, and DSPE-PEG2000 (for *in vivo* experiments) were used to deliver mRNAs. The expression of the reporter mRNAs with synthetic UTRs was demonstrated to be 5–6-times higher than that of β-globin UTRs *in vitro* and *in vivo*.

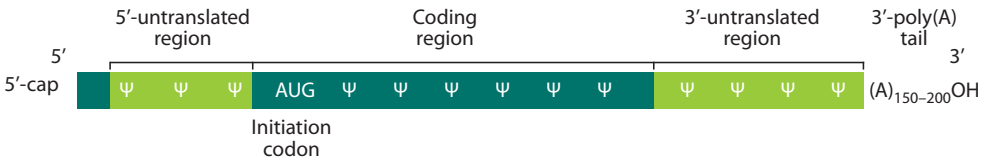


Fig. 1. The mRNA design optimal for effective translation. Inserted pseudouridines are indicated by the "Ψ" sign.

Materials and methods

Plasmid design, construction, and synthesis. Sequence of the hyperactive T7 promoter was taken from (Conrad et al., 2020). Sequences of β-globin 5′- and 3′-UTRs and those of UTRs 5′-UTR-4 and AES-mtRNR1 (3′-UTR) (see Supplementary Table S1)[1] were obtained from NCBI and (Andreev et al., 2009; Leppek et al., 2022), respectively. A multiple cloning site (MCS) between 5′- and 3′-UTRs was designed using several most popular unique restriction sites. Downstream of 3′-UTRs, an XbaI restriction site was inserted to generate a linearized DNA template for *in vitro* transcription.

Fragments containing MCS and UTRs were synthesized by the Laboratory of Synthetic Biology at the Institute of Chemical Biology and Fundamental Medicine (ICBFM) SB RAS and cloned into the pCMV6-Entry vector (OriGene, USA; see Supplementary Fig. S1) by means of restriction sites Psp124BI and XmaI (SibEnzyme, Russia). Two parallel cloning reactions resulted in plasmid vectors: pCMV6_T7_bglob_AGG (containing 5′- and 3′-UTRs of the human HBB gene; see Supplementary Fig. S2) and pCMV6_T7_synth_AGG (containing 5′-UTR-4 and AES-mtRNR1; see Supplementary Fig. S3).

ORFs of GFP and FLuc were PCR-amplified from plasmids phMGFP (Promega, USA) and pCDH-EF1a-Luc2-IRES-mKate2 (Yuzhakova et al., 2022), respectively, with primers containing restriction sites, the Kozak sequence, and start and stop codons and were cloned by the restriction–ligation method into plasmids pCMV6_T7_bglobUTR_AGG and pCMV6_T7_synthUTR_AGG.

***In vitro* transcription and mRNA purification.** GFP and FLuc with β-globin UTRs and the synthetic UTRs were obtained by *in vitro* transcription using T7 polymerase (Biolabmix, Russia). The Anti-Reverse Cap Analog (ARCA) (Biolabmix, Russia) and pseudouridine triphosphate (Biolabmix, Russia) were added during the transcription to modify mRNA structure. After the RNA synthesis, the DNA template was removed with DNase I (Thermo Fisher Scientific, USA). Poly(A) polymerase (New England Biolabs, USA) was used to polyadenylate 3′ termini of the synthesized mRNA by the standard protocol. The RNA products were purified via precipitation with 2.5 M LiCl followed by storage of the precipitate at −20 °C for 30 min and subsequent centrifugation at 16,000 × *g* for 15 min at 4 °C. The pellet was washed with 70 % ethanol and dried for 10 min at room temperature with subsequent dilution in diethylpyrocarbonate (DEPC)-treated H_2O.

Preparation of cationic liposomes. A solution of 1,26-bis(cholest-5-en-3β-yloxycarbonylamino)-7,11,16,20-tetraazahexacosane tetrahydrochloride (2X3; see Supplementary Fig. S4) in a $CHCl_3$–CH_3OH mixture (1:1, v/v) was added to a solution of 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) in $CHCl_3$ at a molar ratio of 1:3 and gently stirred. To obtain PEG-containing cationic liposomes, a solution of DSPE-PEG_{2000} (Lipoid, Germany) (2 % mol.) in $CHCl_3$ was added to the 2X3-DOPE solution at a molar ratio of 1:3. Organic solvents were removed in vacuo, and the obtained lipid film was dried for 4 h at 0.1 Torr to remove residual organic solvents. Then, it was hydrated using deionized water at 4 °C overnight. The liposomal dispersion was sonicated for 15 min at 70–75 °C in a bath-type sonicator (Bandelin Sonorex Digitec DT 52H, Berlin, Germany), filtered (0.45 μm Chromafil® CA-45/25; Macherey–Nagel, Düren, Germany), flushed with argon, and stored at 4 °C.

Size and zeta-potential measurement. Lipoplexes were pre-formed via mixing of equal (25 μL) volumes of the RNA and liposome solutions at appropriate concentrations in saline (154 mM sodium chloride). Lipoplex formation was carried out for 20 min at 25 °C. Next, 10-μL aliquots of lipoplexes were diluted in 1 mL of DEPC-treated water. To measure physicochemical parameters, 1 mL of a lipoplex or liposome suspension was placed into a DTS1070 folded capillary cuvette (Malvern Instruments, Malvern, UK). The size and polydispersity index (PDI) of lipoplexes were measured in three biological replicates by dynamic light scattering (DLS) on a Malvern Zetasizer Nano instrument (Malvern Instruments, Malvern, UK) at a 173° scattering angle and 25 °C. The measurements were performed in Malvern's Zetasizer v7.11 software (Malvern Instruments). A viscosity of 0.8872 centiposes (cP), a refractive index (RI) of 1.330 for the dispersant, and an RI of 1.020 and absorption of 1.335 for the material in suspension were chosen as settings in the software. An equilibration duration of 30 s was selected before the total measurement. ζ-Potential was measured at 25 °C in three biological replicates. Before the measurement, the equilibration duration was set to 120 s. Each measurement was paused for 30 s before the next one.

Atomic force microscopy (AFM) imaging. AFM images were captured in ambient air. Sample preparation for AFM was as follows: (1) dilution of samples to desired concentration, (2) deposition of 6 μL of a sample onto a freshly prepared mica slide (1 × 1 cm) for adsorption for 60 s, (3) rinsing with 100–1,000 μL of MilliQ water, and (4) drying the specimen with a gentle argon stream. Images were acquired on a Multimode 8 (Bruker) atomic force microscope in "ScanAsyst in Air" mode using ScanAsyst-Air probes (Bruker) or in tapping mode with a diamondlike carbon NSG-10 series AFM cantilever (NT-MDT, Zelenograd, Russia) having a tip curvature radius of 1–3 nm. Images were processed, prepared, and analyzed in the Gwyddion software.

Cell lines. The HEK293T/17 cell line was purchased from ATCC (cat. # CRL-11268). Cells were cultured at 37 °C in the DMEM/F12 (1:1) medium supplemented with 10 % of fetal bovine serum (FBS), 1× sodium pyruvate, 1× GlutaMax, 1× antibiotic/antimycotic, and 1× non-essential amino acids (all solutions from Gibco, USA) in a humidified atmosphere with 5 % of CO_2.

Cell transfection *in vitro*. The transfection was performed on HEK293T/17 cells. For the assay, cells were seeded in 24-well plates at 1.4×10^5 cells/well and cultured to 60–70 % confluence in the medium described above. To avoid the degradation of RNA in the lipoplexes, the FBS-containing culture medium was replaced with the 450 μL/well of FBS-free culture medium (the cells were washed with PBS in-between). For the formation of lipoplexes, both RNA (500 ng per well) and liposomes 2X3-DOPE (1:3) were diluted with PBS to a volume of 25 μL per sample with their subsequent mixing. The mixture was incubated for 20 min for lipoplex formation. The lipoplexes were added to the FBS-free cell medium, and the

[1] Supplementary Table S1 and Figs S1–S7 are available at: https://vavilovj-icg.ru/download/pict-2025-29/appx24.pdf

transfection lasted for 5 h. To stop the transfection, the FBS-free medium was replaced with the FBS-containing medium (with intermediate washing with PBS).

Flow cytometry. The transfection of GFP mRNA was carried out as described above in 24-well plates. At 24 h post-transfection, the cells were detached with TrypLE (Gibco, USA), centrifuged for 5 min at 500 × *g*, washed with PBS once, and resuspended in 1 mL of PBS containing 0.5 % of FBS. To assess the level of GFP expression, 10,000 events per sample were acquired on a BD FACSCanto II flow cytometer (BD Biosciences, USA). Transfection efficiency was measured by flow cytometry with the help of two parameters: transfection percentage and mean fluorescence intensity (MFI). The transfection percentage was calculated as the percentage of GFP-positive singlets. The MFI was computed as the mean for a gated cell population. The results were analyzed in the FlowJo software and are presented as the mean and standard deviation (SD) from three replicates.

The time course of luminescence detection *in vitro*. The transfection of Fluc2 mRNA was carried out as described above in 24-well plates. 24h post-transfection medium was removed and 200 μL of cold Luciferase Assay Buffer (25 mM Tris-HCl pH 7.8, 1 % Triton-X100, 5 mM EDTA, 15 mM $MgCl_2$, 75 mM NaCl, 2 mM DTT, 2 mM ATP) was added. The plates were incubated at +4 °C for 20 min. After lysis, the suspension from each well was centrifuged in a separate 1.5 mL tubes at +4 °C, 12,000 *g*, 5 min, then 190 μL of each supernatant was transferred into a new 24-well plate. The luminescence level (represented in relative luminescence units, RLU) was measured with ClarioStar Plus (BMG Labtech, Germany) after injecting 10 μL of 3 mg/mL D-luciferin substrate solution (D-luciferin Potassium Salt, GoldBio, USA) per well. The data were analyzed in BMG Labtech CLARIOstar MARS Software.

The time course of luminescence detection *in vivo*. For *in vivo* experiments, female 4–6-week-old BALB/c mice were obtained from the vivarium of the ICBFM SB RAS (Novosibirsk, Russia). The animal experiments were conducted in accordance with the recommendations for the proper use and care of laboratory mice (ECC Directive 2010/63/EU). All experimental protocols were approved by the Animal Research Ethics Committee at the Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia) (protocol No. 173 of 7 May 2024).

The experiments with mice were conducted in three biological repeats. BALB/c mice were intramuscularly (i.m.) injected with lipoplexes of FLuc mRNA with liposomes in PBS (N/P = 6/1, 10 μg of mRNA, 150 μL per animal). Luciferase expression *in vivo* was assessed 4, 8, 24, 48, 72, and 96 h after administration of lipoplexes to mice. The animals were injected intraperitoneally (i.p.) with 150 μL (3.6 mg per mouse) of freshly prepared D-luciferin potassium salt (Gold Biotechnology, CA, USA) in PBS. After 15 min, the animals were anesthetized with isoflurane, and bioluminescence was visualized using IVIS Lumina X5 (Perkin Elmer, Waltham, MA, USA) with an exposure time of 3 min. The intensity of the luminescent signals was estimated by densitometry using Living Image software v.4.7.4 (Perkin Elmer, Waltham, MA, USA).

Statistical analysis. All data plotted with error bars are expressed as the mean with standard deviation, unless otherwise indicated. GFP signal data were analyzed using one way ANOVA, FLuc signal – using a two-tailed unpaired *t*-test. Significance was evaluated at $p < 0.05$.

Results and discussion

Construction and synthesis of the reporter mRNAs (GFP and FLuc mRNAs) with different types of UTRs

To correctly evaluate the effectiveness of the mRNA delivery in different conditions, the mRNA structure containing 5′-cap, UTRs and poly(A)-tail was suggested. The artificial pseudouridinilated and capped mRNAs were synthesized from linearized plasmids coding for a respective RNA with subsequent T7-mediated transcription and purification (Fedorovskiy et al., 2024).

The abundance of the nucleotide modifications and the combination of modifications in mRNAs with different UTRs were typical (100 % substitution of uridine by pseudouridine); therefore, they did not affect the expression level when comparing mRNAs with different UTRs. For all mRNAs, polyadenylation was carried out under identical conditions, which also could not alter the expression level. Thus, the mRNA synthesis was varied only in terms of the UTR combination.

The purity and integrity of the synthesized *in vitro* polyadenylated mRNAs for the subsequent assays were tested by electrophoresis in a 1.5 % agarose gel (Supplementary Fig. S5).

The physicochemical characterization of the liposomes and their complexes with reporter mRNAs

Upon mRNA characterization, we tested the characteristics of their complexes with the lipid carriers (lipoplexes). The 2X3-DOPE composition at a 1:3 ratio (Fedorovskiy et al., 2024) was used as the carrier in this work as one of the most efficient liposomes tested in previous studies (Markov et al., 2015; Gladkikh et al., 2021), particularly due to the positive impact of the high content of helper lipid DOPE on efficient lipoplex formation and delivery (Vysochinskaya et al., 2022). For the subsequent *in vivo* assays, the polyethyleneglycol (PEG)-decorated lipoconjugate was added to the liposome composition for more prolonged circulation in the blood stream and better clearance. The N/P ratios of 4/1 for the *in vitro* and 6/1 for the *in vivo* delivery were used. Physicochemical properties of the lipoplexes containing FLuc mRNA were examined, including hydrodynamic diameters and ζ-potentials of the liposomes and lipoplexes as described in (Fedorovskiy et al., 2024) (see the Table).

It was shown that the formed lipoplexes were characterized by a small diameter of <200 nm and a positive surface charge of +25…+45 mV, which facilitates their permeabilization through the cell membrane. Additionally, the diameters of the lipoplexes were evaluated by AFM (see Supplementary Fig. S6). The lipoplexes were shown to form homogenous nanoparticles sized 100–200 nm, which confirms the dynamic light scattering measurements. The results demonstrated that the characteristics of the formed lipoplexes were appropriate for *in vitro* and *in vivo* delivery (Vysochinskaya et al., 2022; Fedorovskiy et al., 2024).

Diameters and ζ-potentials of 2X3-DOPE 1:3 and 2X3-DOPE 1:3 + 2 % DSPE-PEG_{2000} liposomes and their lipoplexes with FLuc mRNA (synthetic UTRs)

Lipid nanoparticles	Diameter, nm	ζ-potential, mV
2X3-DOPE 1:3	84.2 ± 1.1	28.5 ± 1.9
2X3-DOPE 1:3 + FLuc mRNA	82.6 ± 4.3	43.8 ± 3.8
2X3-DOPE 1:3 + 2 % DSPE-PEG_{2000}	104.7 ± 0.9	58.7 ± 0.6
2X3-DOPE 1:3 + 2 % DSPE-PEG_{2000} + FLuc mRNA	193.4 ± 3.0	25.6 ± 0.5

Note. The data are presented as the mean ± SD of three replicates.

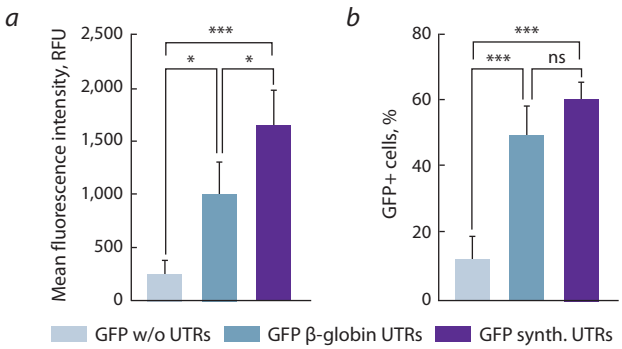


Fig. 2. The comparative analysis of fluorescent signal intensity in GFP mRNA transfected HEK293T/17 cells.

a, The mean fluorescence intensity (RFU) in transfected cells; *b*, the proportion of GFP-positive cells in transfected cells. The data are shown as the mean ± SD of three biological replicates. Data were statistically analyzed using ordinary one-way ANOVA. * $p = 0.0191$; 0.0474 for comparison of cells transfected with GFP with β-globin UTRs or GFP mRNA without UTRs; cells transfected with GFP with β-globin UTRs or GFP with synthetic UTRs respectively (mean GFP+ RFU); *** $p = 0.0009$; 0.0005; 0.0001 for comparison of cells transfected with GFP with synthetic UTRs or GFP mRNA without UTRs (mean GFP+ RFU); GFP with β-globin UTRs and GFP without UTRs; GFP with synthetic UTRs and GFP without UTRs respectively (GFP-positive cells, %), respectively; ns – for comparison of cells transfected with GFP with β-globin UTRs or GFP with synthetic UTRs (GFP-positive cells, %).

The *in vitro* comparison of reporter mRNA expression levels

To identify the most effective mRNA structure upon lipoplex delivery, two mRNA models, namely, GFP and FLuc, were used to test the reporter protein expression *in vitro* on HEK293T/17 cells. Initially, we examined expression efficiency of GFP mRNAs constructed with either β-globin or synthetic UTRs. To confirm the necessity of 5′- and 3′-UTRs in mRNA structure, we estimated the expression level for mRNAs without UTRs in the same experiment (Fig. 2).

The crucial role of UTRs for the prominent expression of synthetic mRNA was confirmed by the increase by 3.9–4.7 times in the number of GFP-positive cells and a 1.5–2.0-fold increase in the mean fluorescence intensity after the addition of UTRs to the mRNA structures. The fluorescence assay revealed a 1.5-fold increase in the fluorescence level of the cells transfected with mRNAs containing the novel UTR combination as compared to mRNAs carrying the commonly used β-globin UTRs. Moreover, this finding supports the idea of more effective expression due to the higher ribosome load at the 5′ end of mRNA (Orlandini von Niessen et al., 2019) and shows effective interaction of nucleotide motifs from these UTRs.

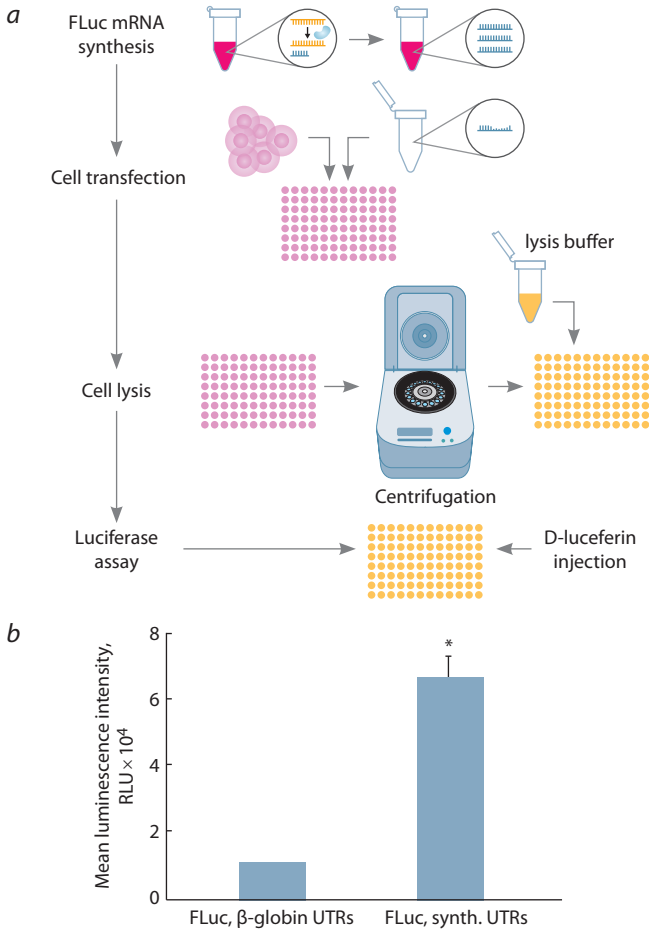


Fig. 3. The luciferase assay and luminescence analysis of FLuc mRNA containing β-globin or synthetic UTRs.

a, The schematic illustration of the luciferase assay *in vitro*. *b*, The average luminescent signals of the transfected cells. The data are presented as the mean ± SD of three biological replicates. Data were statistically analyzed using two-tailed Student's *t*-test. * $p = 0.00009$ as compared with FLuc mRNA containing β-globin UTRs.

To further confirm the efficacy of the novel UTR combination in the translation of reporter mRNAs within cells, an alternative mRNA encoding FLuc was used. The results demonstrated that mRNA flanked by the synthetic UTRs exhibited a luminescent signal intensity that was 6–7 times greater than that observed in mRNA containing β-globin UTRs (Fig. 3*b*). The more sufficient growth of the specific signal in the luminescent assay could be explained by the more significant sensitivity of the luminescence detection

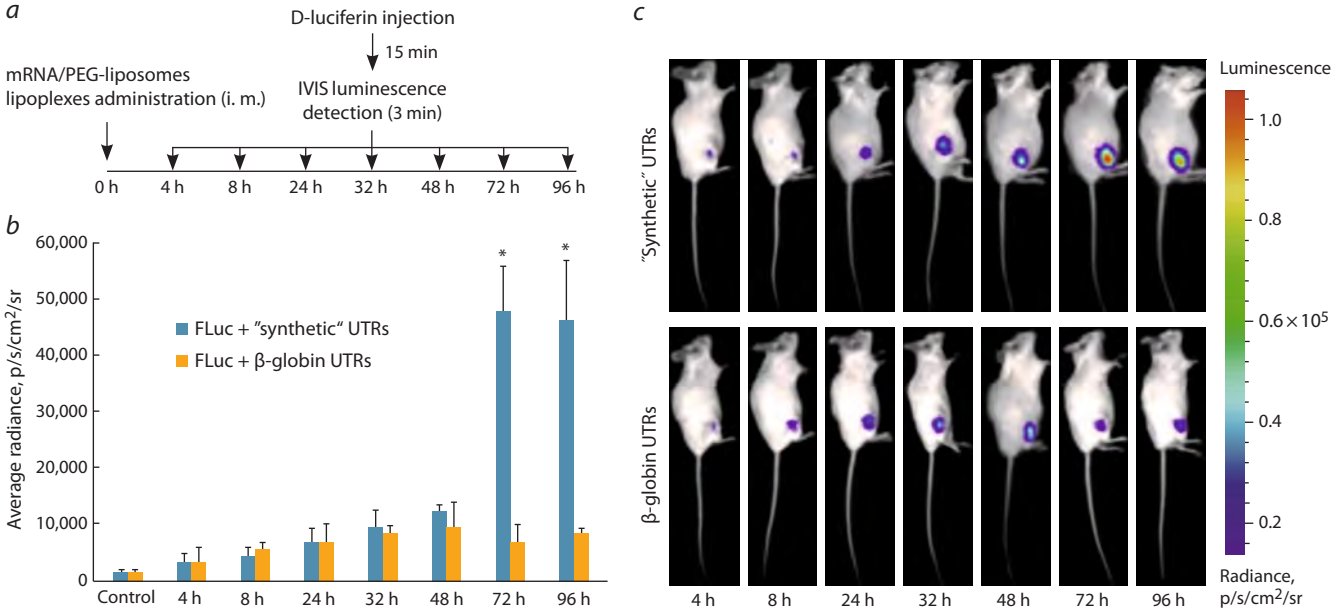


Fig. 4. *In vivo* luminescence analysis of BALB/c mice injected with FLuc mRNA.

a, The scheme of the experiment with key time intervals; *b*, the mean luminescence intensities of regions of interest (ROIs) in mice. The blue and orange bars represent the mice immunized with the FLuc mRNA containing synthetic or β-globin UTRs, respectively. The results are presented as the mean ± SD of three replicates. *c*, Representative IVIS images of Fluc luminescence in mice after the injection of lipoplexes. Data were statistically analyzed using two-tailed Student's *t*-test. * $p = 0.0001$ compared with FLuc mRNA containing β-globin UTRs.

(Troy et al., 2004). Indeed, the luminescence detection tends to be 100-fold more sensitive than the commonly investigated fluorescent detection; thus, the result allows a more precise evaluation of the comparative effectiveness of the investigated combination of UTRs. On the other hand, the higher signal for the luminescence assay upon the substitution of the β-globin UTRs with synthetic UTRs could be caused by the higher expression level of the *FLuc* gene rather than GFP. Mrksich et al. showed that the longer hydrophobic tail of the cationic lipid facilitates the higher translation level of the more prolonged mRNA. As 2X3 used in our assays is quite compact and less branched even compared with C12-200 (Mrksich et al., 2024), it promotes the higher expression of FLuc mRNA with more prolonged UTRs.

The *in vivo* comparison of reporter mRNA expression levels

To evaluate the influence of different UTRs on the efficiency of mRNA translation *in vivo*, lipoplexes of PEGylated liposomes with FLuc mRNAs containing either β-globin or synthetic UTRs were intramuscularly injected into BALB/c mice (Fig. 4).

The *in vivo* results demonstrated that mRNA containing a novel combination of 5′-UTR-4 and 3′-UTR AES-mtRNR1 exhibited dramatically elevated luminescence signal at late time points (≥72 h post injection) that was six times higher compared to β-globin-UTR-containing mRNA. *In vivo* findings revealed a dramatic increase in the luminescent signal observed at 72 h post injection, whereas for the β-globin UTRs, the specific signal tended to decay at 48 h post injection. It is worth noting that in our study, the signal peak shifted from several hours post injection (as shown e. g. (Panova et al., 2023)) to 48 h post injection and later. This shift could be caused not only by the carrier molecules used in this work and their properties but also by the specific translation pattern resulting from the novel UTR combination. According to (Ruis de los Mozos et al., 2013), 5′ and 3′ components of mRNA tend to interact with each other, providing stabilization of the mRNA. Moreover, the extension of 3′-UTR length could have a positive effect on the half-life of the mRNAs through interactions with RNA-binding proteins. The specific luciferase signal was detected even 174 h post injection of artificial mRNA with the novel UTR combination (Supplementary Fig. S7), which may indicate a longer half-life of the mRNA. These results really merge with the *in vitro* assays, indicating the advantages of the novel UTR combination. Long-term presence of mRNA in mammalian tissues and long-term expression of the target gene have been previously described, but for other delivery systems, which explains the difference between our data and other studies (Hassett et al., 2024). The expression enhancement accomplished in our research may facilitate the development of antiviral or anticancer mRNA vaccines possessing higher immunogenicity than the existing analogues.

Conclusion

Overall, the results of this study indicate that the novel combination of synthetic 5′-UTR-4 and 3′-UTR AES-mtRNR1 UTRs introduced into reporter mRNAs demonstrated enhanced mRNA translation in comparison with mRNA containing β-globin UTRs in both *in vitro* and *in vivo* experiments. The optimization of the mRNA structure should improve the development of effective antiviral and anticancer mRNA modalities, which can compete with other types of vaccines and therapeutics.

References

Andreev D.E., Dmitriev S.E., Terenin I.M., Prassolov V.S., Merrick W.C., Shatsky I.N. Differential contribution of the m^7G-cap to the 5′ end-dependent translation initiation of mammalian mRNAs. *Nucleic Acids Res*. 2009;37(18):6135-6147. doi 10.1093/nar/gkp665

Chatterjee S., Pal J.K. Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. *Biol Cell*. 2009;101(5):251-262. doi 10.1042/BC20080104

Chen F., Cocaign-Bousquet M., Girbal L., Nouaille S. 5′UTR sequences influence protein levels in *Escherichia coli* by regulating translation initiation and mRNA stability. *Front Microbiol*. 2022;13:1088941. doi 10.3389/fmicb.2022.1088941

Conrad T., Plumbom I., Alcobendas M., Vidal R., Sauer S. Maximizing transcription of nucleic acids with efficient T7 promoters. *Commun Biol*. 2020;3(1):439. doi 10.1038/s42003-020-01167-x

Fedorovskiy A.G., Antropov D.N., Dome A.S., Puchkov P.A., Makarova D.M., Konopleva M.V., Matveeva A.M., Panova E.V., Shmendel E.V., Maslov M.A., Dmitriev S.E., Stepanov G.A., Markov O.V. Novel efficient lipid-based delivery systems enable a delayed uptake and sustained expression of mRNA in human cells and mouse tissues. *Pharmaceutics*. 2024;16(5):684. doi 10.3390/pharmaceutics16050684

Gladkikh D.V., Sen'kova A.V., Chernikov I.V., Kabilova T.O., Popova N.A., Nikolin V.P., Shmendel E.V., Maslov M.A., Vlassov V.V., Zenkova M.A., Chernolovskaya E.L. Folate-equipped cationic liposomes deliver anti-MDR1-siRNA to the tumor and increase the efficiency of chemotherapy. *Pharmaceutics*. 2021;13(8):1252. doi 10.3390/pharmaceutics13081252

Hassett K.J., Rajlic I.L., Bahl K., White R., Cowens K., Jacquinet E., Burke K.E. mRNA vaccine trafficking and resulting protein expression after intramuscular administration. *Mol Ther Nucleic Acids*. 2024;35(1):102083. doi 10.1016/j.omtn.2023.102083

Kirshina A.S., Vasileva O.O., Kunyk D.A., Seregina K.K., Muslimov A.R., Ivanov R.A., Reshetnikov V.V. Effects of combinations of untranslated-region sequences on translation of mRNA. *Biomolecules*. 2023;13(11):1677. doi 10.3390/biom13111677

Kozak M. The scanning model for translation: an update. *J Cell Biol*. 1989;108(2):229-241. doi 10.1083/jcb.108.2.229

Leppek K., Byeon G.W., Kladwang W., Wayment-Steele H.K., Kerr C.H., Xu A.F., Kim D.S., … Participants E., Dormitzer P.R., Solorzano A., Barna M., Das R. Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. *Nat Commun*. 2022;13(1):1536. doi 10.1038/s41467-022-28776-w

Markov O.V., Mironova N.L., Shmendel E.V., Serikov R.N., Morozova N.G., Maslov M.A., Vlassov V.V., Zenkova M.A. Multicomponent mannose-containing liposomes efficiently deliver RNA in murine immature dendritic cells and provide productive anti-tumour response in murine melanoma model. *J Control Release*. 2015;213:45-56. doi 10.1016/j.jconrel.2015.06.028

Morais P., Adachi H., Yu Y.T. The critical contribution of pseudouridine to mRNA COVID-19 vaccines. *Front Cell Dev Biol*. 2021;9:789427. doi 10.3389/fcell.2021.789427

Mrksich K., Padilla M.S., Joseph R.A., Han E.L., Kim D., Palanki R., Xu J., Mitchell M.J. Influence of ionizable lipid tail length on lipid nanoparticle delivery of mRNA of varying length. *J Biomed Mater Res A*. 2024;112(9):1494-1505. doi 10.1002/jbm.a.37705

Orlandini von Niessen A.G., Poleganov M.A., Rechner C., Plaschke A., Kranz M.L., Fesser M., Diken M., Lower M., Vallazza B., Beissert T., Bukur V., Kuhn A.N., Tureci O., Sahin U. Improving mRNA-based therapeutic gene delivery by expression-augmenting 3′ UTRs identified by cellular library screening. *Mol Ther*. 2019;27(4):824-836. doi 10.1016/j.ymthe.2018.12.011

Panova E.A., Kleymenov D.A., Shcheblyakov D.V., Bykonia E.N., Mazunina E.P., Dzharullaeva A.S., Zolotar A.N., … Dmitriev S.E., Gushchin V.A., Naroditsky B.S., Logunov D.Y., Gintsburg A.L. Single-domain antibody delivery using an mRNA platform protects against lethal doses of botulinum neurotoxin A. *Front Immunol*. 2023;14:1098302. doi 10.3389/fimmu.2023.1098302

Ruiz de los Mozos I., Vergara-Irigaray M., Segura V., Villanueva M., Bitarte N., Saramago M., Domingues S., Arraiano C.M., Fechter P., Romby P., Valle J., Solano C., Lasa I., Toledo-Arana A. Base pairing interaction between 5′- and 3′-UTRs controls *icaR* mRNA translation in *Staphylococcus aureus*. *PLoS Genet*. 2013;9(12):e1004001. doi 10.1371/journal.pgen.1004001

Sample P.J., Wang B., Reid D.W., Presnyak V., McFadyen I.J., Morris D.R., Seelig J. Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. *Nat Biotechnol*. 2019;37(7):803-809. doi 10.1038/s41587-019-0164-5

Troy T., Jekic-McMullen D., Sambucetti L., Rice B. Quantitative comparison of the sensitivity of detection of fluorescent and bioluminescent reporters in animal models. *Mol Imaging*. 2004;3(1):9-23. doi 10.1162/15353500200403196

Vysochinskaya V., Shishlyannikov S., Zabrodskaya Y., Shmendel E., Klotchenko S., Dobrovolskaya O., Gavrilova N., Makarova D., Plotnikova M., Elpaeva E., Gorshkov A., Moshkoff D., Maslov M., Vasin A. Influence of lipid composition of cationic liposomes 2X3-DOPE on mRNA delivery into eukaryotic cells. *Pharmaceutics*. 2022;15(1):8. doi 10.3390/pharmaceutics15010008

Yuzhakova D., Kiseleva E., Shirmanova M., Shcheslavskiy V., Sachkova D., Snopova L., Bederina E., Lukina M., Dudenkova V., Yusubalieva G., Belovezhets T., Matvienko D., Baklaushev V. Highly invasive fluorescent/bioluminescent patient-derived orthotopic model of glioblastoma in mice. *Front Oncol*. 2022;12:897839. doi 10.3389/fonc.2022.897839

Zhuang X., Qi Y., Wang M., Yu N., Nan F., Zhang H., Tian M., Li C., Lu H., Jin N. mRNA vaccines encoding the HA protein of influenza A H1N1 virus delivered by cationic lipid nanoparticles induce protective immune responses in mice. *Vaccines* (*Basel*). 2020;8(1):123. doi 10.3390/vaccines8010123

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

Received Febrary 11, 2025. Revised April 20, 2025. Accepted May 12, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):744-752

doi 10.18699/vjgb-25-82

HOXB13 interactome in prostate cancer cells: biochemical and functional interactions between the transcription factors HOXB13 and TBX3

M.M. Erokhin [1], N.Y. Kozelchuk[1], R.H. Ziganshin[2], V.V. Tatarskiy [1], D.A. Chetverina [1]

[1] Institute of Gene Biology of the Russian Academy of Sciences, Moscow, Russia
[2] Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Moscow, Russia
yermaxbio@yandex.ru; daria.chetverina@gmail.com

Abstract. Transcription factors represent one of the major groups of proteins, whose suppression leads to tumor growth arrest. Different types of cancer express a specific set of transcription factors that create and maintain unique patterns of gene expression. In prostate cancer cells, one of the key transcriptional regulators is the HOXB13 (Homeobox B13) protein. HOXB13 is known to be an important regulator of embryonic development and terminal cell differentiation. HOXB13 regulates the transcription of many genes in normal and transformed prostate cells and is also capable of acting as a pioneer factor that opens chromatin in the regulatory regions of genes. However, little is known about the protein partners and functions of HOXB13 in prostate cells. In the present study, we searched for protein partners of HOXB13 by immunoaffinity purification followed by high-throughput mass spectrometric analysis (IP/LC-MS) using the PC-3 prostate cancer cell line as a model. The main partners of HOXB13 were found to be transcription factors with different types of DNA-binding domains, including the TBX3, TBX2, ZFHX4, ZFHX3, RUNX1, NFAT5 proteins. Using the DepMap resource, we have shown that one of the identified partners, the TBX3 protein is as critical for the growth and proliferation of prostate cancer cell lines *in vitro* as HOXB13. Analysis of individual prostate cancer cell lines revealed that knockout of both genes, *HOXB13* and *TBX3*, leads to the death of the same lines: VCaP, LNCaP (clone FGC), PC-3 and 22Rv1. Thus, HOXB13 and TBX3 can be considered together as potential targets for the development of specific inhibitors that suppress prostate cancer cell growth.

Key words: prostate cancer; transcription factors; regulation of transcription; HOXB13; TBX3; TBX2; ZFHX4; ZFHX3; RUNX1; NFAT5

For citation: Erokhin M.M., Kozelchuk N.Y., Ziganshin R.H., Tatarskiy V.V., Chetverina D.A. HOXB13 interactome in prostate cancer cells: biochemical and functional interactions between the transcription factors HOXB13 and TBX3. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):744-752. doi 10.18699/vjgb-25-82

Funding. The study was supported by Russian Science Foundation grant No. 20-74-10099.

Интерактом белка HOXB13 в клетках рака простаты: биохимические и функциональные взаимодействия между транскрипционными факторами HOXB13 и TBX3

М.М. Ерохин [1], Н.Я. Козельчук[1], Р.Х. Зиганшин[2], В.В. Татарский [1], Д.А. Четверина [1]

[1] Институт биологии гена Российской академии наук, Москва, Россия
[2] Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук, Москва, Россия
yermaxbio@yandex.ru; daria.chetverina@gmail.com

Аннотация. Транскрипционные факторы относятся к одной из главных групп белков, подавление активности которых приводит к остановке роста опухолей. В различных типах рака экспрессируется определенный набор транскрипционных факторов, которые создают и поддерживают специфические паттерны экспрессии генов. В клетках рака простаты ключевым транскрипционным регулятором является белок HOXB13 (Homeobox B13). Известно, что HOXB13 – важный регулятор эмбрионального развития и терминальной клеточной дифференцировки. Он регулирует транскрипцию многих генов в нормальных и трансформированных клетках простаты, а также способен действовать как пионерный фактор, который открывает хроматин в регуляторных областях генов. Однако данных о белковых партнерах и функциях HOXB13 в клетках рака простаты очень мало. В настоящей работе мы провели поиск белковых партнеров HOXB13 методом иммуноаффинной очистки с последующим высокопроизводительным масс-спектрометрическим анализом (IP/LC-MS), используя в качестве модели клеточную линию рака проста-

ты PC-3. Было обнаружено, что основными партнерами HOXB13 являются транскрипционные факторы с разными типами ДНК-связывающих доменов, в том числе белки TBX3, TBX2, ZFHX4, ZFHX3, RUNX1, NFAT5. С помощью ресурса DepMap мы показали, что один из установленных партнеров, белок TBX3, как и HOXB13, критически важен для роста и пролиферации клеточных линий рака простаты *in vitro*. Анализ отдельных клеточных линий рака простаты выявил, что нокаут обоих генов, *HOXB13* и *TBX3*, приводит к гибели одних и тех же линий: VCaP, LNCaP (clone FGC), PC-3 и 22Rv1. Таким образом, HOXB13 и TBX3 могут совместно рассматриваться как потенциальные мишени для создания специфических ингибиторов, подавляющих рост клеток рака простаты.

Ключевые слова: рак простаты; транскрипционные факторы; регуляция транскрипции; HOXB13; TBX3; TBX2; ZFHX4; ZFHX3; RUNX1; NFAT5

Introduction

Prostate cancer is the most commonly diagnosed cancer among men and is one of the leading causes of male cancer mortality (Siegel et al., 2023). Currently, the most common way to target prostate cancer cells chemically is to block the androgen receptor, AR. However, in most cases, tumor cells become resistant to this type of therapy over time, resulting in the development of "castration-resistant prostate cancer" (CRPC) (Crona, Whang, 2017). In this regard, it is important to identify targets for developing new inhibitors of prostate cancer tumor progression.

The transcription factor HOXB13 is a potential target for prostate cancer therapy. This protein is encoded by one of 39 homeobox genes, which contain a DNA-binding HOX domain (also called a homeobox). These genes control transcriptional cascades in various tissues under normal and pathological conditions (Feng et al., 2021; Hubert, Wellik, 2023). HOXB13 was found to be a pioneer factor, with its binding sites often overlapping with those of FOXA1, GATA2 and other DNA-binding proteins in cell lines originating from prostate tissue (Hankey et al., 2020; Pomerantz et al., 2020). *HOXB13* expression levels are elevated in approximately 85 % of prostate adenocarcinoma cases, and this correlates with resistance to AR-targeted therapy, as well as with metastasis and tumor recurrence during treatment (Zabalza et al., 2015; Yao et al., 2019; Weiner et al., 2021). *HOXB13* mutations in tumor cells have also been shown to be associated with a poor prognosis for prostate cancer patients (Ewing et al., 2012; Cai et al., 2015; Adashek et al., 2020).

Despite the important role of HOXB13 in prostate cancer cell proliferation, its biochemical and functional properties are poorly understood. In the present study, we analyzed the HOXB13 protein interactome in the PC-3 prostate cancer cell line. TBX3 was found to be one of HOXB13 protein partners. Both proteins, HOXB13 and TBX3, are required for the growth and proliferation of the same prostate cancer cell lines. Thus, HOXB13 and TBX3 are potential targets for developing new inhibitors for the treatment of prostate cancer

Materials and methods

Immunoprecipitation. Immunoaffinity purification experiments were performed as previously described (Chetverina et al., 2022). The nuclear extract was obtained from the PC-3 cell line. 10^9 cells were washed twice in ice-cold PBS and resuspended in 10 mL of ice-cold Sucrose buffer (10 mM Tris, pH 7.5; 10 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 250 mM sucrose, EDTA-free protease inhibitor cocktail). Cells were homogenized using a Dounce pestle and incubated on ice for 10 min. The nuclei were then pelleted by centrifugation at 3,000g, +4 °C for 10 min. The pellet was resuspended in 1 mL of IP-500 buffer (10 mM Tris, pH 7.5; 500 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 0.1 % NP-40, 10 % glycerol, EDTA-free protease inhibitor cocktail), homogenized with a Dounce pestle and incubated for 1h at +4 °C on a rotator. Lysates were cleared by centrifugation at 18,000g, +4 °C for 10 min. The nuclear extract was then diluted to a final NaCl concentration of 150 mM using IP-0 buffer (10 mM Tris, pH 7.5; 10 mM $MgCl_2$, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 0.1 % NP-40, 10 % glycerol, EDTA-free protease inhibitor cocktail).

Monoclonal antibodies against HOXB13 (EPR17371, ab201682, Abcam) or the IgG of non-immunized rabbits used as a negative control (Jackson ImmunoResearch #011-000-002) were covalently coupled to the protein A Sepharose beads (Pierce) using DMP (Sigma). The nuclear extract containing 150 mM NaCl was incubated with antibodies and Sepharose for 14 h at +4 °C. After washing procedures, the resulting immunoprecipitates were eluted with buffer containing 2 % SDS, 100 mM Tris pH 8.0, 0.5 mM EDTA. Next, the probes were precipitated by TCA, followed by liquid chromatography/tandem mass spectrometry (LC-MS) procedures.

Mass spectrometry analysis of samples. The obtained samples were analyzed as previously described (Chetverina et al., 2022). Sodium deoxycholate (SDC) reduction and alkylation buffer, pH 8.5 (20 µL), containing 100 mM Tris, 1 % (w/v) SDC, 10 mM TCEP and 20 mM 2-chloroacetamide, was added to a 20-µg of each protein sample. Each sample was sonicated in an ultrasonic water bath for 1 min, heated at 95 °C for 10 min, cooled to a room temperature, and an equal volume of trypsin solution in 100 mM Tris pH 8.5 was added in a 1:50 (w/w) ratio. After overnight digestion at 37 °C, peptides were acidified by 40 µL of 2 % trifluoroacetic acid (TFA), mixed with 80 µL of ethyl acetate and purified using SDB-RPS StageTips. After washing the StageTips with 1 % TFA/ethyl acetate 1:1 mixture (2 times) and 0.2 % TFA (1 time), peptides were eluted into a clean tube by 50 % acetonitrile/5 % ammonia mixture. The collected material was vacuum-dried and stored at –80 °C. Before analyses, the peptides were dissolved in 2 % acetonitrile/0.1 % TFA buffer and sonicated for 1 min.

The raw data and detailed protocol of the liquid chromatography and mass spectrometry experiments are publicly available in PRIDE (http://www.ebi.ac.uk/pride), project number PXD059115. The top 20 nuclear proteins were selected accor-

ding to the following parameter: HOXB13 Spectral count/IgG Spectral count>=2. The TNMplot resource was used for GO analysis (Bartha, Győrffy, 2021).

Analysis of the sensitivity of cancer cell lines and gene expression levels in clinical samples. The analysis was based on the DepMap database (https://depmap.org/portal/). The CRISPR release (DepMap Public 24Q4+Score, Chronos) was used to analyze the data obtained by the CRISPR method. The RNAi release (Achilles+DRIVE+Marcotte, DEMETER2) was used to analyze the data obtained by the RNAi method.

Gene expression in tumor samples and the respective normal tissues was evaluated with the Mann–Whitney test using the TNMplot database (https://tnmplot.com), which contains transcriptome data from The Cancer Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTEx) repositories (Bartha, Győrffy, 2021).

The survival analysis was carried out using the UCSC Xena database (http://xena.ucsc.edu/) (Goldman et al., 2020), using TCGA Prostate Cancer (PRAD) cohort, Illumina HiSeq 2000 RNA (dataset ID – TCGA.PRAD.sampleMap/HiSeqV2) and "Primary tumor" filter.

Results

The HOXB13 protein interactome in the PC-3 prostate cancer cell line

The HOXB13 protein partners were identified using immunoaffinity purification (IP) followed by liquid chromatography/tandem mass spectrometry (IP/LC-MS) analysis. For this, we used the PC-3 prostate cancer line, which shows a high level of *HOXB13* gene expression. A nuclear extract was isolated from the PC-3 cell line and incubated with either antibodies against HOXB13 or IgG from a non-immunized animal (the negative control), both of which had been coupled to Protein A Sepharose. After immunoprecipitation and a series of washes, the proteins were eluted from the Sepharose beads using SDS-containing buffer and analyzed by LC-MS mass spectrometry. Figure 1*A*, *B* shows 20 proteins with the highest enrichment in the IP/LC-MS analysis.

GO analysis of 20 nuclear proteins with the highest signal enrichment in the mass spectrometry analysis revealed that 11 out of 20 were DNA-binding transcription factors (Fig. 1*C*).

The two homologous proteins ZFHX4 and ZFHX3 (zinc finger homeobox 4 and 3, respectively) each contain 17 non-clustered C2H2-type zinc finger motifs and four homeobox-type DNA-binding domains (HOX domains) (Fig. 1*A*). The TRPS1 protein also contains several non-clustered C2H2-type zinc finger motifs. The HMG20A protein contains an HMG (high mobility group) DNA-binding domain. Three factors have a T-box-type DNA-binding domain: the MGA (MAX gene-associated protein) and two homologue proteins TBX3 and TBX2 (T-box transcription factors 3 and 2, respectively). The TCF20 protein has an A.T.hook domain; the RUNX1 protein has a Runt domain; the NFAT5 protein has an RHD domain; and the MRE11 protein has a Mre11 domain (Fig. 1*A*).

Thus, many of the top HOXB13 protein partners are transcription factors that have different types of DNA-binding domains.

DepMap database analysis: proteins encoded by the *HOXB13* and *TBX3* genes are most significant for proliferation of prostate cancer cell lines

Next, to determine the functional significance of the identified HOXB13 partners, we queried the DepMap database to find out which cancer types are most affected by the knockout (CRISPR) or knockdown (RNAi) of genes encoding HOXB13 or its protein partners (see the Table). The DepMap project contains data from the screening of a large panel of cancer cell lines originating from various tissues. The project uses CRISPR or RNAi methods to investigate the dependence of cell proliferation on the suppression of individual genes (Tsherniak et al., 2017; Vazquez, Sellers, 2021). The probability of each cell line being dependent on the queried gene is represented by the "Gene effect" score, where strong negative values indicate that a given gene is particularly important for the growth and survival of the respective cell line.

Analysis of the DepMap data revealed that tumor cell lines of different tissue origins respond differently to the depletion of the tested genes. Notably, disruption of either the *TBX3* or *HOXB13* genes activity leads to the preferential death of prostate cancer cell lines. Furthermore, impaired growth and proliferation of prostate tumor cells was observed for both *HOXB13* and *TBX3* using two depletion methods: knockout (CRISPR) and knockdown (RNAi) (Fig. 2*A*, *B*, see the Table).

To understand which cell lines are most sensitive to *HOXB13* and *TBX3* genes deletions, the cell lines were analyzed individually (Fig. 2*C*). The DepMap resource contains data on the effects of CRISPR knockout in 10 lines originating from prostate tissue. Five of them (VCaP, LNCaP (clone FGC), PC-3, 22Rv1 and DU145) originated from aggressive adenocarcinomas. The P4E6, Shmac 4, Shmac 5 lines are derived from cells of well or moderately differentiated non-metastatic prostate carcinomas additionally immortalized by expression of the *E6 HPV* gene (Lang et al., 2006). WPE1-NA22 was derived from the normal prostate RWPE-1 cells following exposure to a chemical carcinogen (MNU) (Webber et al., 2001). BPH-1 was established from primary prostate epithelial cells by immortalization with SV40 large virus T antigen (Hayward et al., 1995).

Four out of five prostate adenocarcinoma cell lines were highly sensitive to the deletion of both the *HOXB13* and *TBX3* genes: VCaP, LNCaP (clone FGC), PC-3 and 22Rv1. By contrast, the deletion of *HOXB13* and *TBX3* was not significant for growth and proliferation of the DU145, WPE1-NA22, P4E6, Shmac 4, and Shmac 5 cell lines. The BPH-1 cell line was sensitive to the deletion of *TBX3*, but not *HOXB13*. Analysis of *HOXB13* and *TBX3* genes transcription in the studied lines (data from the DepMap resource) revealed that these factors exhibited the highest level of transcription in the VCaP, LNCaP (clone FGC), PC-3 and 22Rv1 lines, which are sensitive to the knockout of both genes (Fig. 2*D*). Thus, HOXB13 and TBX3 are preferentially required for the proliferation of cell lines originating from prostate adenocarcinoma samples.

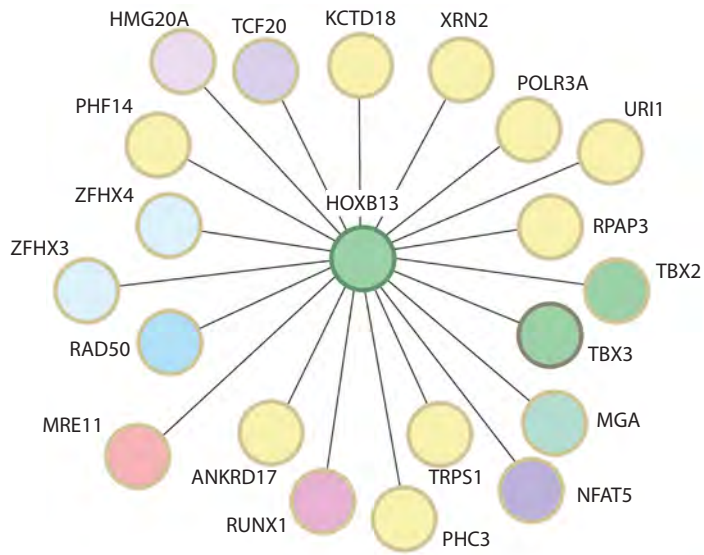
Next, we analyzed the transcription levels of the *HOXB13* and *TBX3* genes in clinical samples using the TNMplot resource. *HOXB13* transcription is normally restricted to prostate tissues (Fig. 3*A*). In clinical tumor samples, high levels of *HOXB13* gene transcripts are observed in prostate and rectal tumor tissues. The *TBX3* gene is transcribed in a

A

DNA-binding protein

	Protein	IP/LC-MS spectral count HOXB13	IgG	HOXB13–IgG	Structural domain/function
	HOXB13	14	2	12	HOX-domain
	Top 20 proteins purified by HOXB13 antibody				
1	KCTD18	108	3	105	BTB/POZ domain
2	XRN2	100	33	67	5′→3′ exoribonuclease activity
3	TCF20	87	36	51	DNA binding, A.T.hook domain, PHD domain
4	ZFHX4	40	1	39	DNA binding, Zinc Fingers C2H2-type, HOX-domains
5	PHF14	36	11	25	PHD domain, Histone-binding protein
6	RAD50	24	3	21	Double-strand break (DSB) repair
7	ZFHX3	21	1	20	DNA binding, Zinc Fingers C2H2-type, HOX-domains
8	POLR3A	18	1	17	DNA-directed RNA polymerase III subunit
9	MGA	26	10	16	DNA binding, T-box domain
10	RPAP3	15	0	15	RNA polymerase II-associated protein
11	TRPS1	23	8	15	DNA binding, Zinc Fingers C2H2-type
12	TBX3	14	1	13	DNA binding, T-box domain
13	HMG20A	18	5	13	DNA binding, HMG domain
14	ANKRD17	17	5	12	Cell cycle and DNA regulation
15	TBX2	10	0	10	DNA binding, T-box domain
16	URI1	11	1	10	Chromatin binding
17	RUNX1	9	0	9	DNA binding, Runt-domain
18	NFAT5	8	0	8	DNA binding, Rel homology domain (RHD)
19	MRE11	10	2	8	Mre11-DNA-binding, Double-strand break (DSB) repair
20	PHC3	12	5	7	Polycomb Group protein

B



C

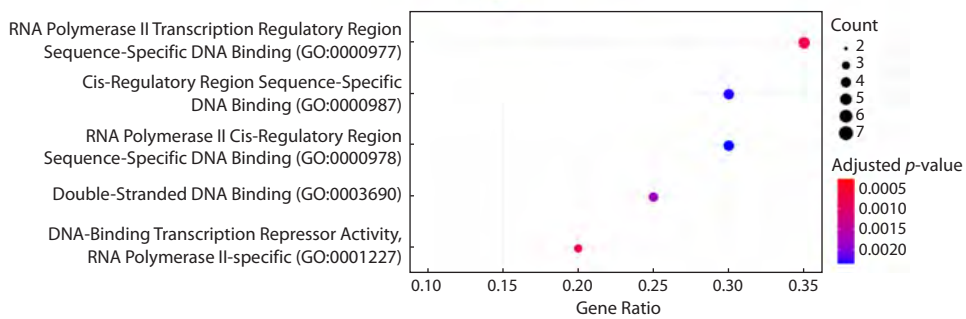
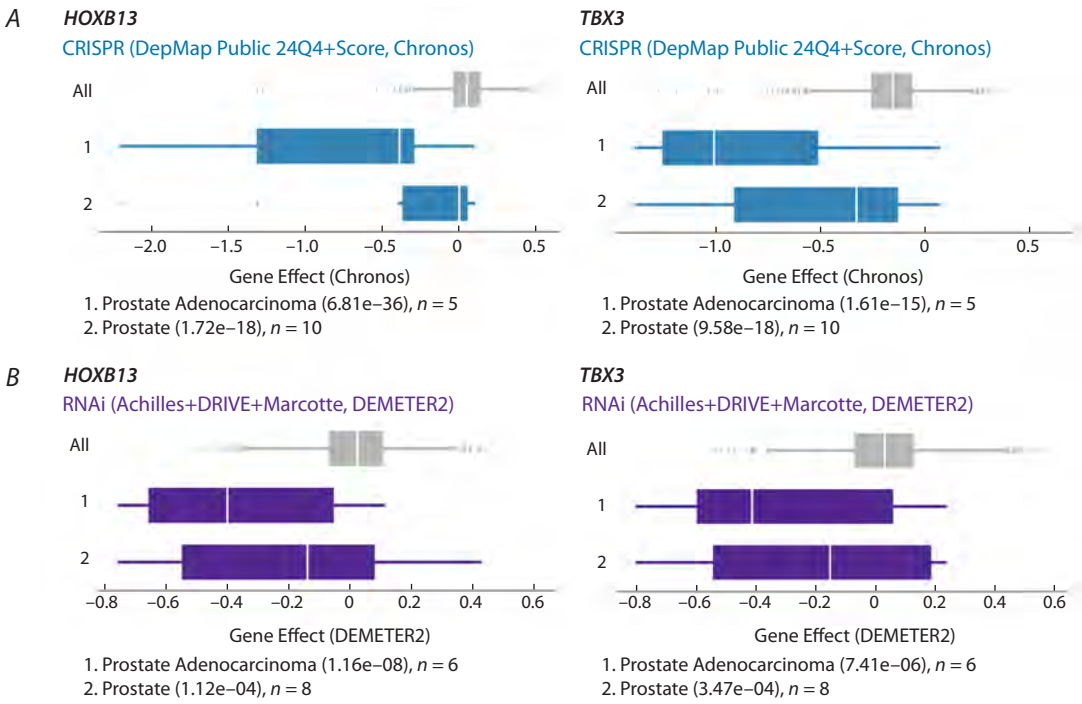


Fig. 1. Transcription factors of the HOXB13 interactome in PC-3 prostate cancer cells.

A – spectral counts of the 20 nuclear proteins displaying the highest enrichment in the IP/LC-MS analysis (top 20). The HOXB13 and IgG columns show the results of the IP/LC-MS analysis performed using HOXB13-specific antibodies or IgG from a non-immunized animal, respectively. The HOXB13–IgG column shows the difference between the HOXB13 and IgG signals. The type of DNA-binding domain is indicated in the rightmost column. *B* – schematic representation of the HOXB13 interactome. *C* – GO analysis of the top 20 nuclear proteins enriched in the IP/LC-MS analysis.



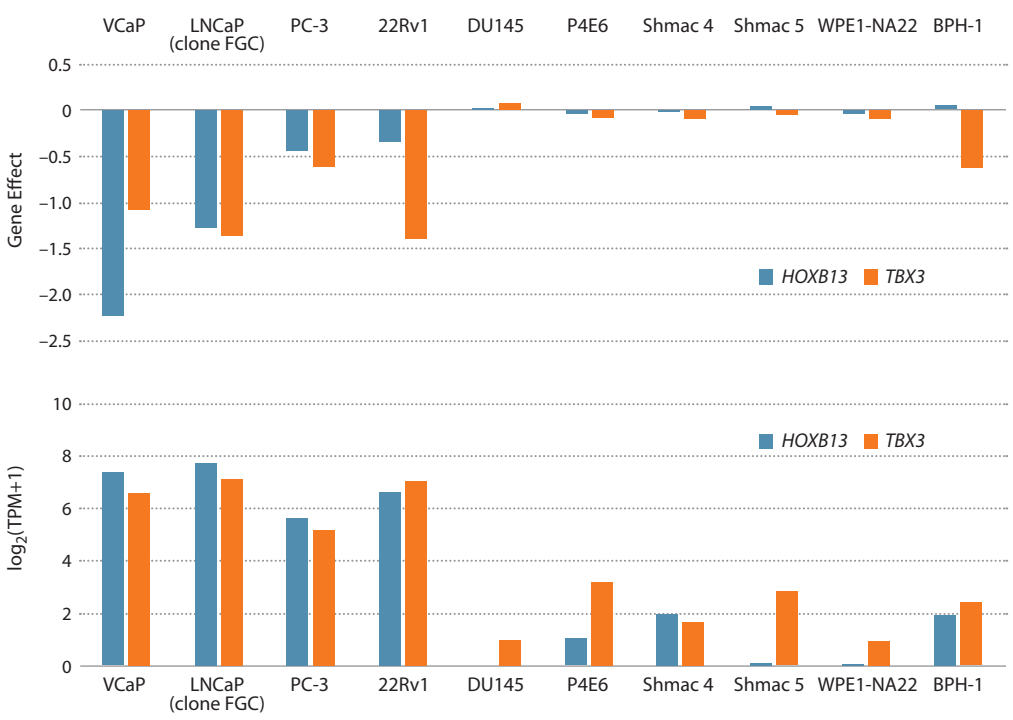


Fig. 2. Prostate cancer cell lines are the most sensitive to knockout and knockdown of the *HOXB13* and *TBX3* genes. The analysis was performed using the DepMap database.

A – DepMap analysis revealed that the knockout (CRISPR) of the *HOXB13* (left) or *TBX3* genes (right) specifically affects the growth and survival of cell lines of prostate origin. The effects of the test gene deletion on all cell lines are greyed out (1,178 cell lines in total). The effects of the test gene deletion on the cell lines of prostate origin are shown in blue. The X-axis shows the Gene Effect scores, which reflect the level of cancer cell proliferation when the tested gene is deleted. The lower the Gene Effect score, the more negative the effect of inhibiting gene activity on cell growth. The median value of the Gene Effect score is indicated by a white space. The analysis of DepMap reveals two groups of lines that are most specifically sensitive to knockout: 1 – the average effect on five lines corresponding to prostate adenocarcinoma; 2 – the average effect on ten prostate-origin cell lines, including those from adenocarcinoma. *B* – knockdown (RNAi, purple) affects the growth and survival of prostate-origin cell lines. The two groups of cell lines that are most specifically sensitive to knockdown are: 1 – the average effect on six prostate adenocarcinoma cell lines; 2 – the average effect on eight cell lines originating from the prostate, including adenocarcinoma cell lines. *C* – the effect of the *HOXB13* or *TBX3* genes knockout using CRISPR on the inhibition of proliferation of ten prostate cell lines, analyzed individually. The Gene Effect scores are indicated on the Y-axis. *D* – the Y-axis shows the transcription levels ($\log_2$(TPM+1)) in ten cell lines originating from prostate tissue. The expression data shown are for cell lines without gene knockout.

The tumor types that are most sensitive to depletion of *HOXB13* or its partners by CRISPR knockout or RNAi knockdown, according to the DepMap database

No.	Gene	CRISPR knockout	RNAi knockdown	No.	Gene	CRISPR knockout	RNAi knockdown
Control	*HOXB13*	Prostate	Prostate	11	*TRPS1*	Breast	Breast
1	*KCTD18*	Ovary	Kidney	12	*TBX3*	Prostate	Prostate
2	*XRN2*	No	No	13	*HMG20A*	No	No
3	*TCF20*	No	No	14	*ANKRD17*	Rhabdomyosarcoma	Breast
4	*ZFHX4*	No	No	15	*TBX2*	Neuroblastoma	Rhabdoid Cancer
5	*PHF14*	No	Mesothelioma	16	*URI1*	No	Head and Neck
6	*RAD50*	Lymphoma	No	17	*RUNX1*	Lymphoma	Lymphoma
7	*ZFHX3*	Rhabdoid Cancer	No	18	*NFAT5*	Stomach	Melanoma
8	*POLR3A*	No	No	19	*MRE11*	No	No
9	*MGA*	Head and Neck	Solid	20	*PHC3*	Ovary	No
10	*RPAP3*	Head and Neck	No				

Note. A "No" designation indicates that none of the tumor types displayed preferential sensitivity to the inactivation of a given gene.

greater number of tissues, with the highest expression levels being observed in the adrenal, prostate and thyroid samples (Fig. 3*B*). The levels of *TBX3* transcription are often lower in tumor tissues than in normal samples, including those from prostate cancer. Figure 3*C* details the changes in the expression levels of the *HOXB13* and *TBX3* genes in normal and tumor prostate tissues. The *HOXB13* gene showed a significant increase in transcription levels (FC = 3.8, *p*-value = 2.28e-69), while *TBX3* transcription levels showed a slight decrease (FC = 0.84, *p*-value = 2.47e-04). However, no significant correlation was found between *HOXB13* and *TBX3* transcription levels and the overall survival of patients diagnosed with prostate adenocarcinoma (Fig. 3*D*).

Discussion

In the present study we used IP/LC-MS to identify the protein partners of HOXB13 in the PC-3 prostate cancer cell line. We show that many of the HOXB13 partners are DNA-binding proteins.

One of the discovered partners of HOXB13 is the TBX3 protein, which has a T-box type DNA-binding domain in its structure. DepMap portal data analysis revealed that knockout (CRISPR) and knockdown (RNAi) of both the *HOXB13* and *TBX3* genes most significantly inhibited the growth and proliferation of cell lines originating from prostate adenocarcinomas. Transcription levels of the *HOXB13* gene, but not those of *TBX3*, are significantly higher in clinical samples obtained from patients with prostate adenocarcinoma compared to normal prostate tissue samples. However, no correlation was found between increased *HOXB13* or *TBX3* gene transcription levels and the overall survival of patients with prostate adenocarcinoma.

It can be assumed that the HOXB13 and TBX3 proteins are closely related in terms of their function and that they participate in the same transcription regulation cascades. Potentially, the combined inhibition of HOXB13 and TBX3 activities may have a stronger inhibitory effect on prostate cancer cell proliferation than inactivation of these proteins individually. For the TBX3 protein, it has previously been shown that it can repress transcription of tumor suppressor genes such as $p14^{ARF}$ (Lingbeek et al., 2002; Yarosh et al., 2008). It is possible that the cooperation between HOXB13 and TBX3 could increase the repression of the transcription of a subset of tumor suppressors. Further testing is required to confirm this. The role of TBX3 has been investigated in liver and breast tumors (Khan et al., 2020), but to our knowledge, only one study has examined the function of this factor in prostate cancer cells (Hwang et al., 2022). Using the LNCaP cell line, J.H. Hwang et al. demonstrated the presence of various DNA-binding proteins, including TBX3, HOXB13, FOXA1, and AR, in the interactome of the transcriptional co-factor CREB5. Knockdown of TBX3 and FOXA1 reduced the viability of the LNCaP cells.

The two HOXB13 protein partners, ZFHX4 and ZFHX3, are homologues and each of them contains four homeobox domains. It is known that homeobox domains can form protein-protein interactions (Ortiz-Lombardia et al., 2017). It is possible that the HOXB13 and ZFHX4/ZFHX3 homeodomains can interact directly. This will need to be tested in the future.

Previous studies using the VCaP prostate cancer cell line have shown that HOXB13 precipitates with the EED protein, which is a component of the PRC2 repressor complex (Cao et al., 2014). The study also revealed effective EED interaction with the PRC1 Polycomb repressor complex. In the current study, we did not detect any PRC2 complex subunits in the HOXB13 immunoprecipitate. However, two other Polycomb factors were identified: the PHC3 and RUNX1 proteins. PHC3, like its homologues PHC1 and PHC2, is the core subunit of the PRC1 subcomplex, which is known as cPRC1 (Schuettengruber et al., 2017). The transcription factor RUNX1 was previously shown to interact with the PRC1 core component BMI1 (a.k.a. PCGF4) (Yu et al., 2012). HOXB13 may potentially be involved in the regulation of transcription together with the Polycomb group of repressors. Further studies are required to investigate this.

To date, considerable attention has been devoted to identifying new targets for cancer therapy, with transcription

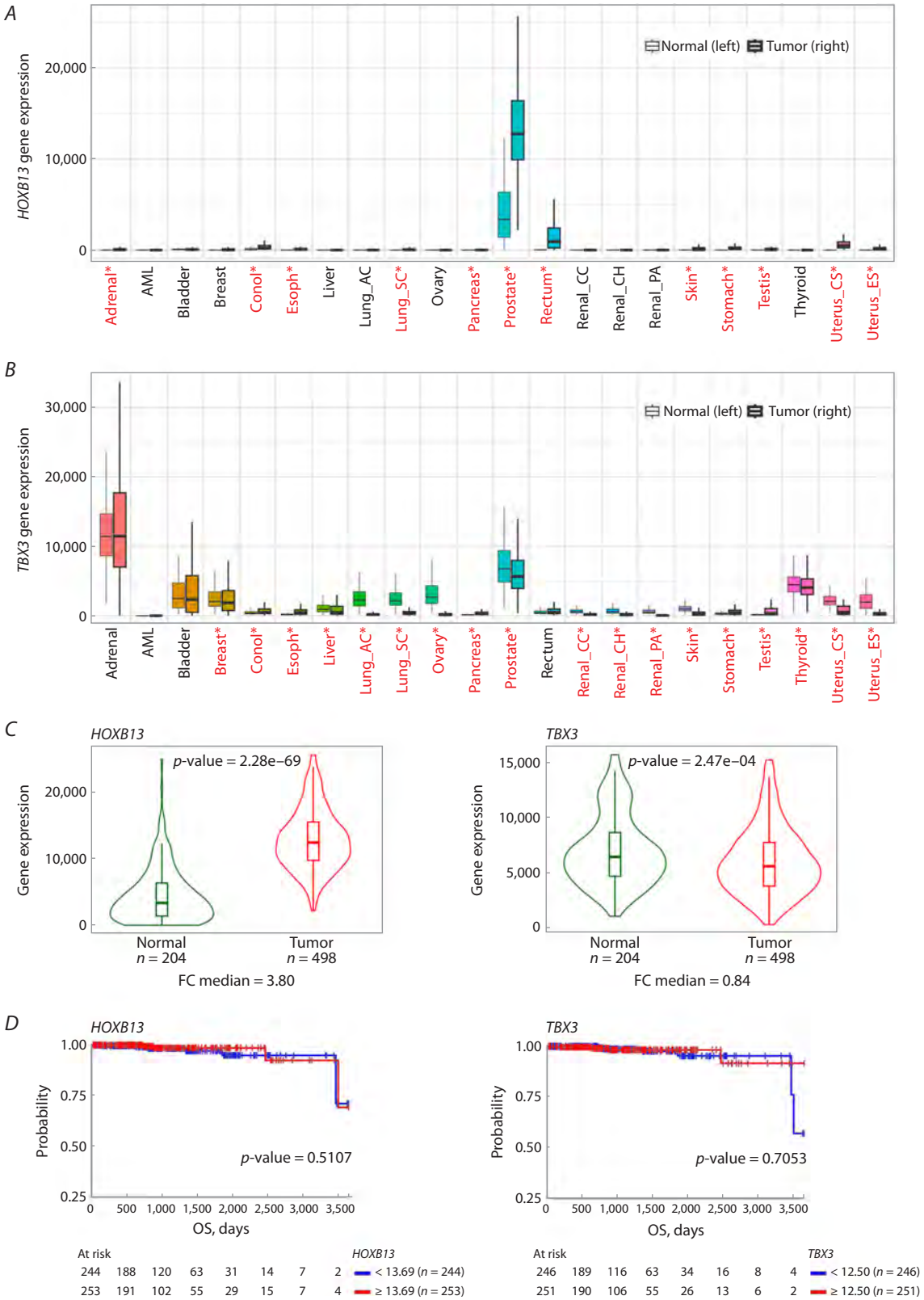


Fig. 3. Expression of *the HOXB13* and *TBX3* genes in clinical samples.

A, *B* – differential *HOXB13* and *TBX3* gene expression in normal (left) and tumor (right) samples of different tissues, analysis performed using TNMplot (Pan-cancer box plot). Cases with *p*-value < 0.05 (Mann–Whitney test) and expression >10 in tumor or normal samples are marked in red. *C* – the differential *HOXB13* and *TBX3* gene expression in normal (green) *vs.* prostate adenocarcinoma tumor (red) samples is presented as violin plots. p-value, Mann–Whitney test; FC – fold change median, analysis performed using TNMplot. *D* – correlations between *HOXB13* and *TBX3* gene transcription levels in prostate adenocarcinoma clinical samples and the overall survival (OS); analysis performed using the UCSC Xena resource. Cohorts with high gene expression are highlighted in red; cohorts with low expression are highlighted in blue.

factors being among the most promising ones (Bouhlel et al., 2015; Hagenbuchner, Ausserlechner, 2016; Lambert et al., 2018). There is growing evidence that suggests a key role for DNA-binding transcription factors in the processes of malignant tumor development (Vishnoi et al., 2020; Zhang et al., 2020). For example, analysis of the DepMap database identifies transcription factors as a crucial class of genes, the expression of which is critical for tumor cell proliferation (Chetverina et al., 2023). To date, methods have been developed to inhibit the activity of DNA-binding transcription factors. This makes transcription factors relevant subjects for study as potential targets for cancer treatment (Bushweller, 2019; Li et al., 2022; Zhuang et al., 2022; Xie et al., 2023).

The results of this study show that a large proportion of the top HOXB13 interactome proteins are DNA-binding factors. The ability of transcription factors to interact with other DNA-binding proteins is probably one of the most common mechanisms for regulating gene transcription in mammals (Jolma et al., 2013, 2015) and other multicellular organisms (Erokhin et al., 2018). Apparently, transcription factors can form ordered macro-complexes through multiple interactions, allowing them to recognize not only single, often degenerate binding sites, but also more extended regions of DNA consisting of a set of motifs for several proteins. The ability of DNA-binding proteins to interact with each other may be crucial for the specific selection of the DNA regions for the binding of regulatory complexes. Modulating the ability of individual DNA-binding proteins to be recruited to chromatin could potentially enable more precise control of gene expression, making transcription factors a promising target for the development of various anti-cancer drugs.

Conclusion

The following main conclusions can be drawn based on the data obtained. 1. The main HOXB13 partners are transcription factors that have different types of DNA-binding domains. 2. Cell lines originating from prostate adenocarcinoma samples are most sensitive to deletion of the *HOXB13* and *TBX3* genes. 3. Suppression of *HOXB13* and *TBX3* gene activities inhibits the growth and proliferation of the same prostate adenocarcinoma cell lines. Further studies are required to understand the effects of co-inhibition of *HOXB13* and *TBX3* *in vitro* and *in vivo*.

References

Adashek J.J., Leonard A., Roszik J., Menta A.K., Genovese G., Subbiah V., Msaouel P. Cancer genetics and therapeutic opportunities in urologic practice. *Cancers* (*Basel*). 2020;12(3):710. doi 10.3390/cancers12030710

Bartha Á., Győrffy B. TNMplot.com: a web tool for the comparison of gene expression in normal, tumor and metastatic tissues. *Int J Mol Sci*. 2021;22(5):2622. doi 10.3390/ijms22052622

Bouhlel M.A., Lambert M., David-Cordonnier M.-H. Targeting transcription factor binding to DNA by competing with DNA binders as an approach for controlling gene expression. *Curr Top Med Chem*. 2015;15(14):1323-1358. doi 10.2174/1568026615666150413154713

Bushweller J.H. Targeting transcription factors in cancer – from undruggable to reality. *Nat Rev Cancer*. 2019;19(11):611-624. doi 10.1038/s41568-019-0196-7

Cai Q., Wang X., Li X., Gong R., Guo X., Tang Y., Yang K., Niu Y., Zhao Y. Germline HOXB13 p.Gly84Glu mutation and cancer susceptibility: a pooled analysis of 25 epidemiological studies with 145,257 participates. *Oncotarget*. 2015;6(39):42312-42321. doi 10.18632/oncotarget.5994

Cao Q., Wang X., Zhao M., Yang R., Malik R., Qiao Y., Poliakov A., … Feng F.Y., Kalantry S., Qin Z.S., Dhanasekaran S.M., Chinnaiyan A.M. The central role of EED in the orchestration of polycomb group complexes. *Nat Commun*. 2014;5:3127. doi 10.1038/ncomms4127

Chetverina D., Vorobyeva N.E., Mazina M.Y., Fab L.V., Lomaev D., Golovnina A., Mogila V., Georgiev P., Ziganshin R.H., Erokhin M. Comparative interactome analysis of the PRE DNA-binding factors: purification of the Combgap-, Zeste-, Psq-, and Adf1-associated proteins. *Cell Mol Life Sci*. 2022;79(7):353. doi 10.1007/s00018-022-04383-2

Chetverina D., Vorobyeva N.E., Gyorffy B., Shtil A.A., Erokhin M. Analyses of genes critical to tumor survival reveal potential 'super-targets': focus on transcription. *Cancers* (*Basel*). 2023;15(11):3042. doi 10.3390/cancers15113042

Crona D.J., Whang Y.E. Androgen receptor-dependent and -independent mechanisms involved in prostate cancer therapy resistance. *Cancers* (*Basel*). 2017;9(6):67. doi 10.3390/cancers9060067

Erokhin M., Georgiev P., Chetverina D. *Drosophila* DNA-binding proteins in polycomb repression. *Epigenomes*. 2018;2(1):1. doi 10.3390/ epigenomes2010001

Ewing C.M., Ray A.M., Lange E.M., Zuhlke K.A., Robbins C.M., Tembe W.D., Wiley K.E., … Montie J.E., Xu J., Carpten J.D., Isaacs W.B., Cooney K.A. Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med*. 2012;366(2):141-149. doi 10.1056/NEJMoa1110000

Feng Y., Zhang T., Wang Y., Xie M., Ji X., Luo X., Huang W., Xia L. Homeobox genes in cancers: from carcinogenesis to recent therapeutic intervention. *Front Oncol*. 2021;11:770428. doi 10.3389/fonc.2021.770428

Goldman M.J., Craft B., Hastie M., Repečka K., McDade F., Kamath A., Banerjee A., Luo Y., Rogers D., Brooks A.N., Zhu J., Haussler D. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 2020;38(6):675-678. doi 10.1038/s41587-020-0546-8

Hagenbuchner J., Ausserlechner M.J. Targeting transcription factors by small compounds – current strategies and future implications. *Biochem Pharmacol*. 2016;107:1-13. doi 10.1016/j.bcp.2015.12.006

Hankey W., Chen Z., Wang Q. Shaping chromatin states in prostate cancer by pioneer transcription factors. *Cancer Res*. 2020;80(12):2427-2436. doi 10.1158/0008-5472.CAN-19-3447

Hayward S.W., Dahiya R., Cunha G.R., Bartek J., Deshpande N., Narayan P. Establishment and characterization of an immortalized but non-transformed human prostate epithelial cell line: BPH-1. *In Vitro Cell Dev Biol Anim*. 1995;31(1):14-24. doi 10.1007/BF02631333

Hubert K.A., Wellik D.M. Hox genes in development and beyond. *Development*. 2023;150(1):dev192476. doi 10.1242/dev.192476

Hwang J.H., Arafeh R., Seo J.-H., Baca S.C., Ludwig M., Arnoff T.E., Sawyer L., … Kregel S., Van Allen E.M., Drake J.M., Freedman M.L., Hahn W.C. CREB5 reprograms FOXA1 nuclear interactions to promote resistance to androgen receptor-targeting therapies. *eLife*. 2022;11:e73223. doi 10.7554/eLife.73223

Jolma A., Yan J., Whitington T., Toivonen J., Nitta K.R., Rastas P., Morgunova E., … Hughes T.R., Lemaire P., Ukkonen E., Kivioja T., Taipale J. DNA-binding specificities of human transcription factors. *Cell*. 2013;152(1-2):327-339. doi 10.1016/j.cell.2012.12.009

Jolma A., Yin Y., Nitta K.R., Dave K., Popov A., Taipale M., Enge M., Kivioja T., Morgunova E., Taipale J. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*. 2015;527(7578):384-388. doi 10.1038/nature15518

Khan S.F., Damerell V., Omar R., Du Toit M., Khan M., Maranyane H.M., Mlaza M., Bleloch J., Bellis C., Sahm B.D.B., Peres J., ArulJothi K.N., Prince S. The roles and regulation of TBX3 in development and disease. *Gene*. 2020;726:144223. doi 10.1016/j.gene.2019.144223

Lambert M., Jambon S., Depauw S., David-Cordonnier M.-H. Targeting transcription factors for cancer treatment. *Molecules*. 2018; 23(6):1479. doi 10.3390/molecules23061479

Lang S.H., Smith J., Hyde C., Macintosh C., Stower M., Maitland N.J. Differentiation of prostate epithelial cell cultures by materigel/stromal cell glandular reconstruction. *In Vitro Cell Dev Biol Anim*. 2006;42(8):273-280. doi 10.1290/0511080.1

Li Y., Song J., Zhou P., Zhou J., Xie S. Targeting undruggable transcription factors with PROTACs: advances and perspectives. *J Med Chem*. 2022;65(15):10183-10194. doi 10.1021/acs.jmedchem.2c00691

Lingbeek M.E., Jacobs J.J.L., van Lohuizen M. The T-box repressors *TBX2* and *TBX3* specifically regulate the tumor suppressor gene $p14^{ARF}$ via a variant T-site in the initiator. *J Biol Chem*. 2002; 277(29):26120-26127. doi 10.1074/jbc.M200403200

Ortiz-Lombardia M., Foos N., Maurel-Zaffran C., Saurin A.J., Graba Y. Hox functional diversity: novel insights from flexible motif folding and plastic protein interaction. *BioEssays*. 2017;39(4):1600246. doi 10.1002/bies.201600246

Pomerantz M.M., Qiu X., Zhu Y., Takeda D.Y., Pan W., Baca S.C., Gusev A., … Lee G.-S.M., Corey E., Long H.W., Zwart W., Freedman M.L. Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat Genet*. 2020;52(8): 790-799. doi 10.1038/s41588-020-0664-8

Schuettengruber B., Bourbon H.-M., Di Croce L., Cavalli G. Genome regulation by Polycomb and Trithorax: 70 years and counting. *Cell*. 2017;171(1):34-57. doi 10.1016/j.cell.2017.08.002

Siegel R.L., Miller K.D., Wagle N.S., Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. 2023;73(1):17-48. doi 10.3322/caac.21763

Tsherniak A., Vazquez F., Montgomery P.G., Weir B.A., Kryukov G., Cowley G.S., Gill S., … Garraway L.A., Root D.E., Golub T.R., Boehm J.S., Hahn W.C. Defining a cancer dependency map. *Cell*. 2017;170(3):564-576.e16. doi 10.1016/j.cell.2017.06.010

Vazquez F., Sellers W.R. Are CRISPR screens providing the next generation of therapeutic targets? *Cancer Res*. 2021;81(23):5806-5809. doi 10.1158/0008-5472.CAN-21-1784

Vishnoi K., Viswakarma N., Rana A., Rana B. Transcription factors in cancer development and therapy. *Cancers* (*Basel*). 2020;12(8): 2296. doi 10.3390/cancers12082296

Webber M.M., Quader S.T.A., Kleinman H.K., Bello-DeOcampo D., Storto P.D., Bice G., DeMendonca-Calaca W., Williams D.E. Human cell lines as an in vitro/in vivo model for prostate carcinogenesis and progression. *Prostate*. 2001;47(1):1-13. doi 10.1002/pros.1041

Weiner A.B., Faisal F.A., Davicioni E., Karnes R.J., Griend D.J.V., Lotan T.L., Schaeffer E.M. Somatic HOXB13 expression correlates with metastatic progression in men with localized prostate cancer following radical prostatectomy. *Eur Urol Oncol*. 2021;4(6):955-962. doi 10.1016/j.euo.2020.05.001

Xie X., Yu T., Li X., Zhang N., Foster L.J., Peng C., Huang W., He G. Recent advances in targeting the "undruggable" proteins: from drug discovery to clinical trials. *Signal Transduct Target Ther*. 2023;8(1): 335. doi 10.1038/s41392-023-01589-z

Yao J., Chen Y., Nguyen D.T., Thompson Z.J., Eroshkin A.M., Nerlakanti N., Patel A.K., … Coppola D., Zhang J., Perera R., Kim Y., Mahajan K. The homeobox gene, *HOXB13*, regulates a mitotic protein-kinase interaction network in metastatic prostate cancers. *Sci Rep*. 2019;9(1):9715. doi 10.1038/s41598-019-46064-4

Yarosh W., Barrientos T., Esmailpour T., Lin L., Carpenter P.M., Osann K., Anton-Culver H., Huang T. TBX3 is overexpressed in breast cancer and represses $p14^{ARF}$ by interacting with histone deacetylases. *Cancer Res*. 2008;68(3):693-699. doi 10.1158/0008-5472.CAN-07-5012

Yu M., Mazor T., Huang H., Huang H.-T., Kathrein K.L., Woo A.J., Chouinard C.R., … Roeder R.G., Kim C.F., Zon L.I., Fraenkel E., Cantor A.B. Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Mol Cell*. 2012; 45(3):330-343. doi 10.1016/j.molcel.2011.11.032

Zabalza C.V., Adam M., Burdelski C., Wilczak W., Wittmer C., Kraft S., Krech T., … Minner S., Simon R., Sauter G., Schlomm T., Tsourlakis M.C. HOXB13 overexpression is an independent predictor of early PSA recurrence in prostate cancer treated by radical prostatectomy. *Oncotarget*. 2015;6(14):12822-12834. doi 10.18632/oncotarget.3431

Zhang J., Lee D., Dhiman V., Jiang P., Xu J., McGillivray P., Yang H., … Cheng C., Yue F., Liu X.S., White K.P., Gerstein M. An integrative ENCODE resource for cancer genomics. *Nat Commun*. 2020;11(1):3696. doi 10.1038/s41467-020-14743-w

Zhuang J.-J., Liu Q., Wu D.-L., Tie L. Current strategies and progress for targeting the "undruggable" transcription factors. *Acta Pharmacol Sin*. 2022;43(10):2474-2481. doi 10.1038/s41401-021-00852-9

Conflict of interest. The authors declare no conflict of interest.
Received February 7, 2025. Revised April 8, 2025. Accepted April 17, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):753-768

doi 10.18699/vjgb-25-83

Karyological differentiation among bread wheat cultivars (*Triticum aestivum* L.) with distinct breeding statuses and growth habits

A.F. Muterko [1], E.D. Badaeva [2], E.V. Zuev [3], E.A. Salina [1]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Vavilov Institute of General Genetics of the Russian Academy of Sciences, Moscow, Russia
[3] Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia
muterko@gmail.com, muterko@bionet.nsc.ru

Abstract. The assessment of intraspecific variability of wheat has been relevant for years. Although most modern wheat cultivars are considered to be pure lines, the heterogeneity of varietal populations is one of the mechanisms for maintaining population homeostasis. It is possible that the high evolutionary stability of constitutive heterochromatin and its stable distribution within chromosomes will allow us to use karyological analysis not only for studying the genesis and taxonomy of *Triticum* L., but also for studying the intraspecific diversity of wheat. In this regard, a classification of 87 Russian cultivars of common wheat differing in breeding status (landraces and modern cultivars) and growth habit (spring and winter) was carried out using two alternative approaches for assessing karyograms. The first approach uses the qualitative assessment of karyograms based on the distribution of C-bands on chromosomes. We also proposed that quantification of karyograms based on the size of C-bands would make the classification of cultivars more adequate. The variability, informative value and resolution of diagnostic features, trends in grouping cultivars, and their associations with the breeding status and growth habit were studied. A high potential of karyotyping with C-banding in discriminating modern cultivars by growth habit, as well as in separating winter cultivars from landraces has been revealed. In terms of the tested karyological features, the homogeneity of modern cultivars was higher than that of local cultivars, and the homogeneity of winter wheat was higher than that of spring wheat. The obtained classification reflects the preservation of high similarity in the karyograms of modern spring cultivars and landraces, as well as the low distinguishability between the karyograms of landraces differing in growth habit. A comparative analysis of the classifications of 20 cultivars using C-banding and SNP genotyping (3,126 polymorphic markers) suggests that studying the karyotypic variability allows us to infer a more accurate differentiation of wheat varietal populations based on the breeding status than using SNP markers that detect genetic variability, especially when the number of diagnostic features is limited.

Key words: common winter and spring wheat; modern and local cultivars; landrace; breeding; karyological analysis; karyosystematics; C-banding

For citation: Muterko A.F., Badaeva E.D., Zuev E.V., Salina E.A. Karyological differentiation among bread wheat cultivars (*Triticum aestivum* L.) with distinct breeding statuses and growth habits. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):753-768. doi 10.18699/vjgb-25-83

Funding. The work was supported by Russian Science Foundation project No. 21-76-30003. The karyosystematics was carried out within the government-funded project FWNR-2022-0017.

Кариологическая дифференциация среди сортов мягкой пшеницы (*Triticum aestivum* L.), контрастных по селекционному статусу и типу развития

А.Ф. Мутерко [1], Е.Д. Бадаева [2], Е.В. Зуев [3], Е.А. Салина [1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия
[3] Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР), Санкт-Петербург, Россия
muterko@gmail.com, muterko@bionet.nsc.ru

Аннотация. На протяжении многих лет оценка внутривидовой изменчивости пшеницы не теряет своей актуальности. Хотя большинство современных сортов пшеницы относят к чистолинейным, гетерогенность сортовых популяций выступает одним из механизмов поддержания популяционного гомеостаза. Возможно, высокая эволюционная стабильность конститутивного гетерохроматина и его устойчивое распределение на хромосо-

мах позволят эффективно использовать кариологический анализ не только для исследования генезиса и таксономии рода *Triticum* L., но и для изучения внутривидового разнообразия пшеницы. В этой связи проведена классификация 87 российских сортов мягкой пшеницы различного селекционного статуса (староместные и современные сорта) и типа развития (яровые и озимые) на основании оценки кариограмм, выполненной с использованием двух подходов. Первый подход основан на качественной оценке кариограмм по распределению гетерохроматиновых С-блоков на хромосомах. Мы также предположили, что количественная оценка кариограмм по размеру индивидуальных С-блоков (второй подход) сделает классификацию сортов более адекватной. Исследовались вариабельность, информативность и разрешающая способность диагностических признаков, тенденции в группировании сортов, а также их ассоциации с селекционным статусом и типом развития. Результаты выявили высокий потенциал С-окраски в дискриминации современных сортов мягкой пшеницы по типу развития и обособлении их озимых форм от местных культур. Гомогенность современных сортов по тестируемым кариологическим признакам была выше, чем староместных, а озимых – чем яровых. Полученная классификация отражает сохранение высокой общности в кариограммах современных яровых культур и сортовых популяций местного возделывания, а также слабую различимость кариограмм староместных сортов, контрастных по типу развития. Сравнительный анализ классификаций 20 выборочных сортов по данным С-окрашивания и ОНП-генотипирования (3126 полиморфных маркеров) предполагает, что изучение кариотипической изменчивости помогает составить более верное представление о дифференциации сортовых совокупностей пшеницы по селекционному статусу, чем при использовании ОНП-маркеров, детектирующих генную изменчивость, особенно при ограниченном количестве диагностических признаков.

Ключевые слова: мягкая яровая и озимая пшеница; староместные и современные сорта; селекция; кариологический анализ; кариосистематика; С-окраска хромосом

Introduction

Cultivation of common wheat *Triticum aestivum* L. ($2n = 6x = 42$, BBAADD) in various eco-geographical regions with its exceptionally wide distribution has led to the accumulation of structural, genetic and physiological changes, and the emergence of a huge diversity of intraspecific forms (Zohary et al., 2012; Zhao et al., 2023). In this regard, the study of wheat intraspecific variability seems relevant. Although most modern wheat cultivars are classified as pure lines, their inherent polymorphism, heterogeneity of varietal populations, acts as one of the mechanisms for maintaining population homeostasis (Fadeeva, Narbut, 1969; Kudriavtsev, 2006; Serpolay-Besson et al., 2011), which leads to poorly reproducible results in their differentiation (Kudriavtsev, 2006; Metakovsky et al., 2024). The high evolutionary stability of heterochromatic blocks, as well as their stable distribution on chromosomes, creates the prerequisites for the successful use of chromosomal markers to solve such problems.

Previously, karyological analysis has proven itself successful in the study of the genesis and taxonomy of wheat within the genus (Iordansky et al., 1978a, b; Zurabishvili et al., 1978; Badaeva et al., 1986, 1994, 2007, 2015a, 2022; Gill et al., 1991; Jiang J. et al., 1993, 1994; Friebe, Gill, 1996; Dedkova et al., 2004, 2007, 2009); however, the validity and prospects of its application in assessing the population variability of this crop remain a subject of discussion. This is largely due to both the relatively high labor intensity of karyological analysis and the complexity of describing the karyotype of a variety in a form accessible for statistical processing. So, the karyogram of the model variety Chinese Spring serves as a standard for describing deletion lines (Endo, Gill, 1996) and compiling physical maps of chromosomes (Delaney et al., 1995; Mickelson-Young et al., 1995), but due to the absence of a number of C-bands present in other species or cultivars of wheat, it cannot be directly used to characterize intraspecific polymorphism. At the same time, although many authors have noted a wide variety of differential staining patterns of common wheat chromosomes (Iordansky et al., 1978a; Zurabishvili et al., 1978; Seal, 1982; Friebe, Gill, 1994), the karyograms they provide are not suitable for statistical processing and analysis of population structure.

Improvement of the fluorescent *in situ* hybridization method and the development of an oligoprobe system have significantly simplified and reduced the cost of the analysis and made it possible to study fairly large samples (Jiang M. et al., 2017; Huang et al., 2018; Guo et al., 2019; Hu et al., 2022). At the same time, the assessment of the frequency and distribution of polymorphic variants served to identify groups of closely related cultivars (Huang et al., 2018; Guo et al., 2019; Hu et al., 2022). Another approach, "chromosomal passportization", is based on a comparison of the karyotype of a specific sample with a generalized species ideogram (Badaeva et al., 1990). In contrast to the previously discussed approach, the diagnostic feature here is an individual block of constitutive heterochromatin (C-band). This approach was first used to assess the diversity of spelt wheat (Dedkova et al., 2004) and the European emmer group (Dedkova et al., 2009), and the population structure of *T. dicoccum* obtained with its use corresponded well to the existing taxonomy (Goncharov, 2012; Badaeva et al., 2015b). The population structure of *T. araraticum*, revealed using chromosomal analysis (Badaeva et al., 2022), was completely consistent with the data of a molecular genetic study of the same samples, performed using SSAP markers.

Depending on the breeding status, a distinction is made between landraces (local populations) and modern wheat cultivars. Local varietal populations are locally adapted, they are traditionally cultivated in isolated areas, and their seed production is carried out without deliberate hybridization and targeted change of genotype (Zeven, 1998). Nevertheless, the genetic diversity of landraces was maintained by cultivating a mixture of different genotypes or often even a mixture of different crops, which created the possibility of

exchanging genetic material between plants (Zeven, 1980, 1998; Feldman, 2001). The transition to scientifically based breeding of common wheat, based on targeted selection of parental genotypes, introgression of foreign genes, and the use of mutagenesis took place in the late 19th–early 20th centuries (Feldman, 2001). Since samples from geographically separated populations, as well as representatives of related taxa, were often used in the creation of cultivars (Mujeeb-Kazi et al., 2013; Sharma M. et al., 2020; Sharma S. et al., 2021; Boehm, Cai, 2024), the possibility of the formation of qualitatively new, specific karyotypes, isolated and maintained in the pool of breeding cultivars, cannot be ruled out. This circumstance actualizes the assessment of the potential of karyological analysis in the differentiation and classification of wheat varietal populations.

While the relationship between the karyotype of a given cultivar and its origin has natural grounds, it is fundamentally important to test the associations of the karyotype with other, less obvious – but significant in its formation – factors. One of these factors is the climate regime of the cultivation region, since it has a direct impact on the formation of the growth habit. In particular, winter cultivars are sown before the onset of cold winter weather, supporting the vernalization process under conditions of low temperature and shortened photoperiod. Their vegetation continues with the onset of the warm period of the year, and they ear earlier than spring cultivars, which are only sown in the spring. Thus, the difference in the timing of earing of winter and spring wheat can act as one of the mechanisms of reproductive isolation in the formation of specific features of karyotypes in cultivars with an alternative sowing season. However, we are not aware of any studies of this kind.

In the present study, the potential of constitutional heterochromatin karyotyping (C-staining) in discrimination of Russian-bred common wheat varietal populations is determined. The relationship between the classification obtained, the breeding status (landraces and modern), and the growth habit (winter and spring) is analysed. Two approaches to the extraction of diagnostic features are tested. The first approach is based on the qualitative assessment of karyograms by the distribution of heterochromatic C-bands on chromosomes. We also assumed that the quantitative assessment of karyograms by the size of individual C-bands would make the classification of cultivars more adequate. The advantages and disadvantages of karyosystematics of varietal populations are discussed in comparison with genotyping based on single nucleotide polymorphism (SNP genotyping).

Materials and methods

Plant material. Karyological analysis was performed on 87 cultivars of common wheat from different regions of Russia. The samples were 44 landraces (bred from landraces and received in the VIR collection or zoned in the USSR before 1940) and 43 modern cultivars (obtained after 1940 with hybridization) containing equal proportions of spring and winter samples (Supplementary Materials, Table S1)[1].

[1] Tables S1–S6 and Figures S1–S11 are available in Supplementary Materials at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Muterko_Engl_29_6.pdf

Karyological analysis. The standard C-banding technique was used to obtain and stain chromosome preparations (Badaeva et al., 1994). Chromosomes were classified according to the genetic nomenclature (Gill et al., 1991). Karyograms were evaluated using two approaches. Qualitative evaluation was based on discrimination of chromosome types by the presence of C-bands indicated on the reference karyogram. Quantitative evaluation involved visual determination of the size of C-bands on a six-point scale (from 0 to 5) in accordance with the previously proposed recommendations (Badaeva et al., 1990). For this purpose, the chromosomes of a specific variety were compared with the generalized species ideogram and each identified band was assigned a number from 1 to 5 depending on its size: 1 – small, 2 – small but clearly visible, 3 – medium, 4 – large, 5 – very large. If a band was missing in any position, it was assigned the value "0". Constant bands that did not vary in size were excluded from the analysis.

Data analysis. Cluster analysis was performed using hierarchical and K-means clustering on Euclidean distances calculated from the initial binary data. K-means clustering was performed on a standardized (centered and scaled) distance matrix. Gap statistics (function *fviz_nbclust*, R package *factoextra*) were used to determine the optimal number of clusters for initializing the K-means algorithm. Visualization of clusters on a plane was performed using the method of multidimensional scaling of the initial binary matrix (function *cmdscale*, R package *stats*) (Becker et al., 1988).

The choice of the agglomerative hierarchical clustering method was made among eight algorithms implemented in the *hclust function* (R package *stats*) using the following approaches. 1. The empirical evaluation of the method was based on its ability to separate cultivars by the considered characteristics (breeding status, growth habit and their combination). Since the largest number of characteristic states is four (four variants of combinations of the breeding status with the growth habit), the tested topology was divided into four clusters with a volume of at least three cultivars. In each cluster, the number of cultivars with the same type of the characteristic was counted and the largest was selected, then these values were summed up for all clusters. 2. To assess the parsimony of the cladogram, expressed by the smallest number of changes explaining its topology, the Fitch algorithm implemented in the *parsimony* function from the R package *phangorn* was used (Schliep, 2011).

Cladograms were calculated with the *hclust function* and visualized using the R package *dendextend* (Galili, 2015). The assessment of clade conservation on the dendrogram, support for their monophyly, was carried out using the bootstrap method (Felsenstein, 1985) with 1,000 iterations, implemented in the *boot.phylo* function from the R package *ape* (Paradis, Schliep, 2019). Topology entanglement was calculated using the *untangle* function (R package *dendextend*) in three rounds using the *random method* and an additional round using the *step2side* method. The topological distance between dendrograms (symmetric difference) (Robinson, Foulds, 1987) was calculated with the *dist.dendlist* function (R package *dendextend*).

The selection of groups of cultivars from the cladogram was carried out by cutting its topology to a given number of clusters using the *cutree* function from the R package *stats* (Becker et al., 1988), the centers of which were calculated for each variety from the distance matrix as the average value of the distances between the given variety and the cultivars from the current cluster. Standardized coordinates of the cluster centers were used in the factor analysis.

Factor analysis was performed using the principal component analysis, using the characteristics being tested as additional qualitative variables (*PCA* function, R package *FactoMineR*) (Husson et al., 2010). Pearson correlation coefficients (r) between additional variables (tested characteristics) and dimensions were calculated as square roots of *eta2*, and correlations between dimensions and values of these variables as square roots of *cos2*. Assuming a normal distribution of correlation coefficients (according to the central limit theorem), the asymptotic p-values for $r > 0.40$ with 87 observations were <0.01 (power = 0.9).

Multiple correspondence analysis was performed using the *MCA* function (R package *FactoMineR*) on the original binary data transformed into logical levels. The characteristics being tested were utilized as additional variables.

The statistical significance of overrepresentation assumptions was assessed using Fisher's exact test (function *fisher.test* from the R package *stats*), followed by filtering the results by p-value ≥ 0.05.

Logistic regression was calculated using the *glm* function (R package *stats*), using a quasi-binomial model with link function *logit*.

Primary data on SNP genotyping of 20 cultivars from the tested sample were obtained from the work of D.A. Afonnikov et al. (2024).

Results

The results of karyological differentiation of 87 Russian cultivars as well as the analysis of the conjugacy of their breeding status and growth habit with the resulting classification are consistently presented according to two types of diagnostic features: chromosome type and C-band size.

Qualitative assessment of karyograms. Chromosome type as a differentiating feature

Association of the chromosome type with the breeding status and the growth habit of cultivars

Using chromosome type as a diagnostic feature, a total of 205 unique karyograms were detected for chromosomes from the A, B, and D genomes (Fig. S1). Chromosomes 4D and 5D were excluded from the analysis, since the C-staining method used did not allow for reliable separation of their polymorphic variants. Although the frequency of occurrence of individual chromosome types varied greatly depending on the breeding status of cultivars, their growth habit, as well as the combinations of these characteristics (Fig. S2), no differences in their distribution by the characteristics tested were found between the genomes and groups of homoeologous chromosomes; in the groups of cultivars with alternative forms of the tested characteristics, an almost equal number of chromosome types from each genome or homoeologous group is presented, which indicates the absence of bias in the original data. 12 groups of associated chromosome types were found, characterized by identical sets in all cultivars (Table S2). In most cases, the groups were presented in pairs, three groups contained three associated types, and one four.

To assess the statistical significance of the association of chromosome types with tested characteristics, Fisher's exact test for overrepresentation was used in a sample of cultivars homogeneous for the trait being tested (Table S3). Most chromosome types (144, 70 %) were not informative, since the test for their overrepresentation resulted in a significance level of the alternative hypothesis $p \geq 0.05$. The number of overrepresented chromosome types in modern cultivars was more than that in landraces of identical growth habit, in the absence of differences between contrasting samples in terms of the tested characteristics (Fig. 1*a*, *c*).

Similar results were obtained when assessing the overrepresentation of chromosome types among winter and spring cultivars with identical breeding status, and, conversely, among landraces and modern cultivars of identical growth habit (Fig. 1*b*, *d*). It follows that the homogeneity of modern cultivars is higher than that of landraces, and that of winter cultivars is higher than that of spring cultivars.

Cluster analysis of cultivars by chromosome type

Cluster analysis was carried out using hierarchical and K-means clustering in order to select an approach that provides the most adequately interpretable idea of the differentiation of cultivars according to the tested characteristics.

According to gap statistics, the optimal number of clusters for K-means algorithm is three. Although visualization of the clusters on the principal coordinate plane confirmed good resolution (Fig. 2*a*), only the third cluster was highly homogenous in terms of the characteristics under consideration (Fig. 2*b*). It included 21 exclusively modern cultivars, predominantly of the winter growth habit (20 cultivars). The first cluster contained 30 cultivars, with a predominance of landraces (27 cultivars), equally represented by the growth habit (16 winter and 14 spring). The remaining cluster of 36 cultivars was the most heterogeneous (Fig. 2*b*).

It is important to note that, based on these results, the cultivars were optimally divided into three clusters, rather than the expected four, assumed by the number of contrasting forms of tested characteristics (landraces/modern, winter/spring, or their combinations) considered, and modern cultivars were divided by the growth habit, with the majority of winter cultivars (87 %) being allocated, while landraces were not distinguishable by this trait. At the same time, CS were grouped with landraces, and were mostly concentrated in the cluster with a predominance of spring landraces (cluster 2).

The choice of the agglomerative hierarchical clustering method was made empirically, according to the best separation of cultivars by the considered characteristics, as well as by the least number of changes explaining the tested topology (Table S4). Both approaches predict the best result for the cladogram topology calculated by the Ward method, taking into account the clustering criterion (method "ward.D2").

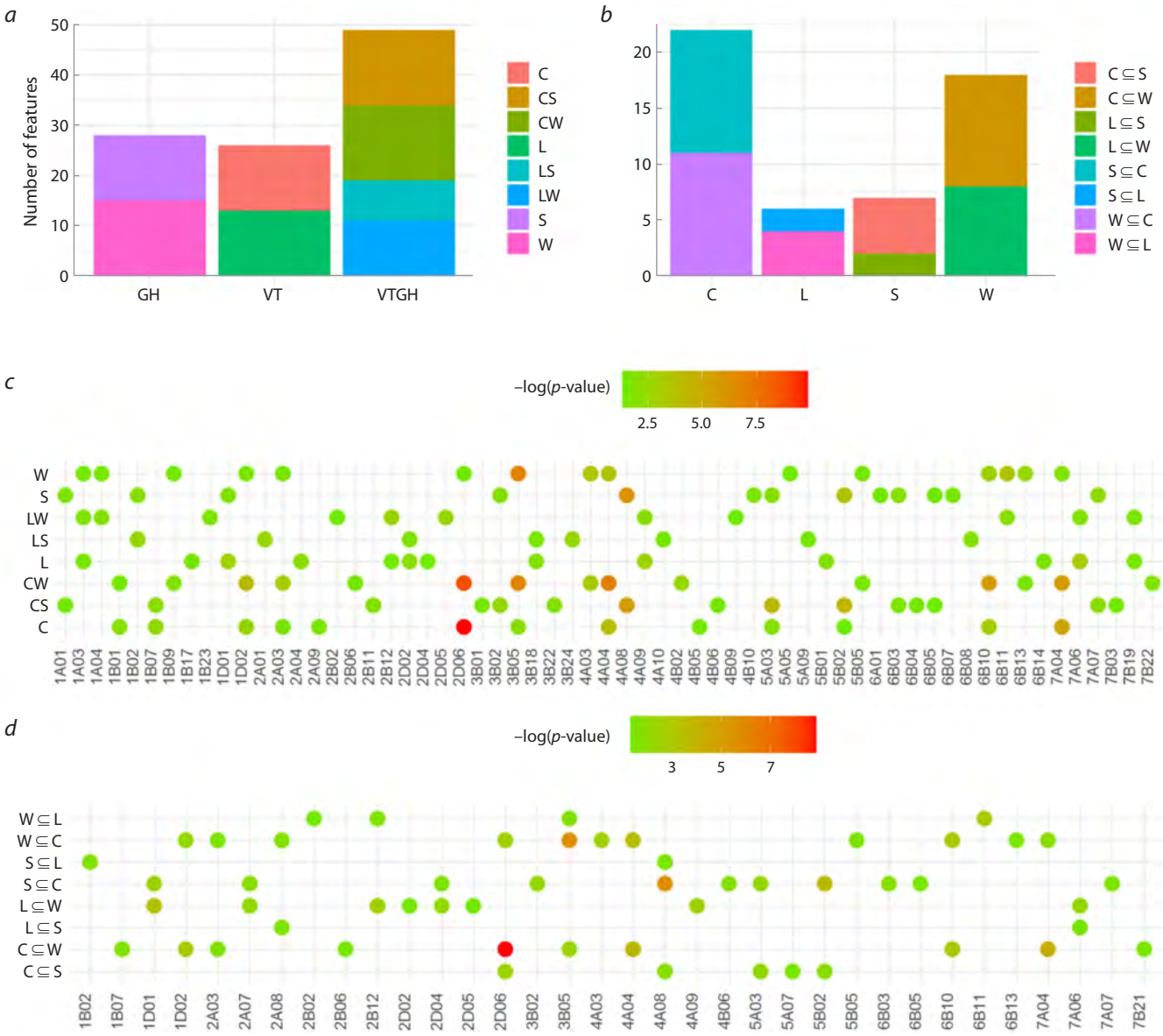


Fig. 1. Association of the chromosome type with the breeding status and the growth habit of cultivars.

The number of chromosome types associated with the breeding status (VT), growth habit (GH) and combination of these characteristics (VTGH) in the total sample (*a*) and among winter (W) and spring (S) landraces (W⊆L, S⊆L) and modern (W⊆C, S⊆C) cultivars, and *vice versa* among landraces (L) and modern cultivars (C) of spring (L⊆S, C⊆S) and winter (L⊆W, C⊆W) wheat (*b*). $p < 0.05$ test values for overrepresentation in the sample of cultivars homogeneous in the tested trait from the total population (*c*) and subsamples (*d*).

In the dendrogram of cultivars, two major clades, characterized by the greatest distance from each other, divide modern winter cultivars (22 of 23) and the remaining groups (Fig. 2*c*). More than 70 % of the dispersion in the average distances between cultivars and major clades is explained by this division, namely, the allocation of CW into a separate cluster. With further pruning of the topology, a small group of seven CS and a large cluster uniting almost all landraces (42 of 44) are identified. The completely homogeneous cluster CS (cluster 1), although characterized by a much greater distance from both landraces (cluster 3) and modern winter cultivars (cluster 2) than the distance of the latter two from each other, is more strongly attracted to the group of cultivars of predominantly landraces (the total length of branches between clusters 1 and 3 is 2.4 times shorter than that between clusters 1 and 2), among which there are already 40 % of CS (Fig. 2*e*). This tendency is more clearly reflected in the plane of the main factors of attraction of cultivars to clusters (Fig. 2*d*). Here, the first component is positively correlated with the inclusion of cultivars in cluster 2 ($r = 0.85$), and the second is negatively correlated with the inclusion of cultivars in cluster 1 ($r = –0.94$). In the principal component plot, most of the CS (75 %) are localized in the negative coordinates of the first dimension. CS are highly correlated with the second dimension ($r = 0.98$), the correlation with which is also more pronounced for LW ($r = 0.63$) and LS ($r = 0.31$) than for CW ($r = 0.13$). Thus, the major clade from clusters 1 and 3 unites most CS with landraces. It is

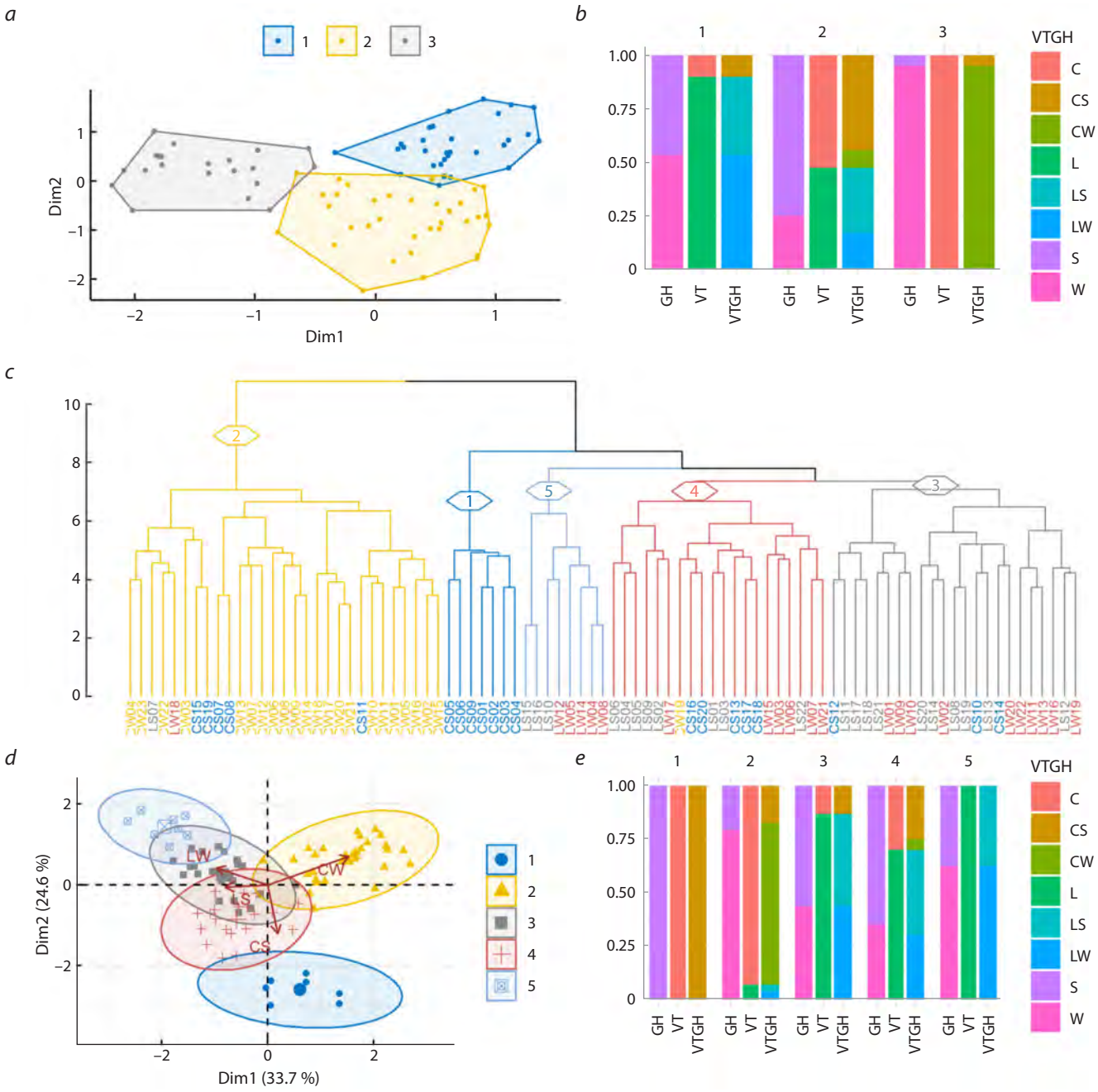


Fig. 2. Cluster analysis of cultivars by chromosome type.

a – visualization of clusters calculated by the K-means on the plane of principal coordinates; *b* – proportions of cultivars (W – winter, S – spring, L – landraces, C – modern) with alternative forms of the tested characteristics in these clusters; *c* – dendrogram of hierarchical clustering of cultivars calculated by the Ward method. For clarity, the first five clades with their representation on the plane of principal components (*d*) and the corresponding proportions of cultivars with alternative forms of the tested characteristics (*e*) are highlighted in color.

noteworthy that in cluster 3, CS are grouped with landraces of predominantly spring type, suggesting that modern spring cultivars have undergone less intensive breeding, retaining a greater similarity of karyograms with landraces. Despite the similarity of results between alternative cluster analysis approaches, hierarchical clustering provides a more adequately interpretable representation of variety differentiation. Support for monophyly of most clades was low, but statistically significant (9.3–99.3 %) for all top-ranking clades (Fig. S3).

Optimization of the branch length (cophenetic distances between cultivars) of the current dendrogram topology by the likelihood maximization method resulted in compression of clusters and their better resolution on the plane of principal coordinates (Fig. S4). The logarithm of likelihood (logLik) increases from –12,355 to –4,115. The separation of cultivars by breeding status becomes more convincing, but still imperfect, whereas landraces completely merge by the growth habit, but the splitting of modern cultivars by this characteristic increases.

Correspondence analysis of cultivars by chromosome type

In Figure 3*a*, the distribution of cultivars is presented on the plane of the first two dimensions. The proportion of explainable variance is evenly distributed between the main dimensions (3.7 and 3.6 %, respectively). However, the purpose of the present correspondence analysis is not to assess the contribution of individual chromosome types to the distribution of cultivars, but to study the relationship between this distribution and the tested characteristics, represented here by additional qualitative variables, based on their correlations with the dimensions.

The first dimension is highly correlated with the characteristic combining the breeding status and the growth habit of cultivars ($r = 0.66$), and to a greater extent, with their growth habit ($r = 0.47$) than with the breeding status ($r = 0.29$, $p > 0.14$). The characteristic combining the breeding status and the growth habit of cultivars is also best correlated with the second dimension ($r = 0.80$), but, unlike the first dimension, mainly due to the breeding status ($r = 0.64$), rather than the growth habit ($r = 0.43$). Thus, while the growth habit correlates equally with both dimensions, the correlations of the breeding status of cultivars with the main dimensions differ by more than two times. In addition, the characteristic combining the breeding status and the growth habit also tends more towards dividing cultivars by breeding status (the correlation with the second dimension is 21 % higher than with the first), suggesting that the separation of cultivars is more likely to be determined by their breeding status than by the growth habit. It is characteristic that the tendency to group by breeding status is most pronounced for CW ($r = 0.75$, with the second dimension) and LS ($r = 0.47$, with the second dimension), but not for CS ($r = 0.03$ [$p > 0.89$], with the second dimension). The correlations of CW and LS with the growth habit are so small ($r = 0.28$ and $r = 0.08$, with the first dimension) that they are statistically insignificant ($p > 0.17$ and 0.87), in contrast to CS ($r = 0.64$ [$p < 1e{-}4$], with the first dimension). Landrace winter cultivars (LW) equally weakly and statistically insignificantly correlate

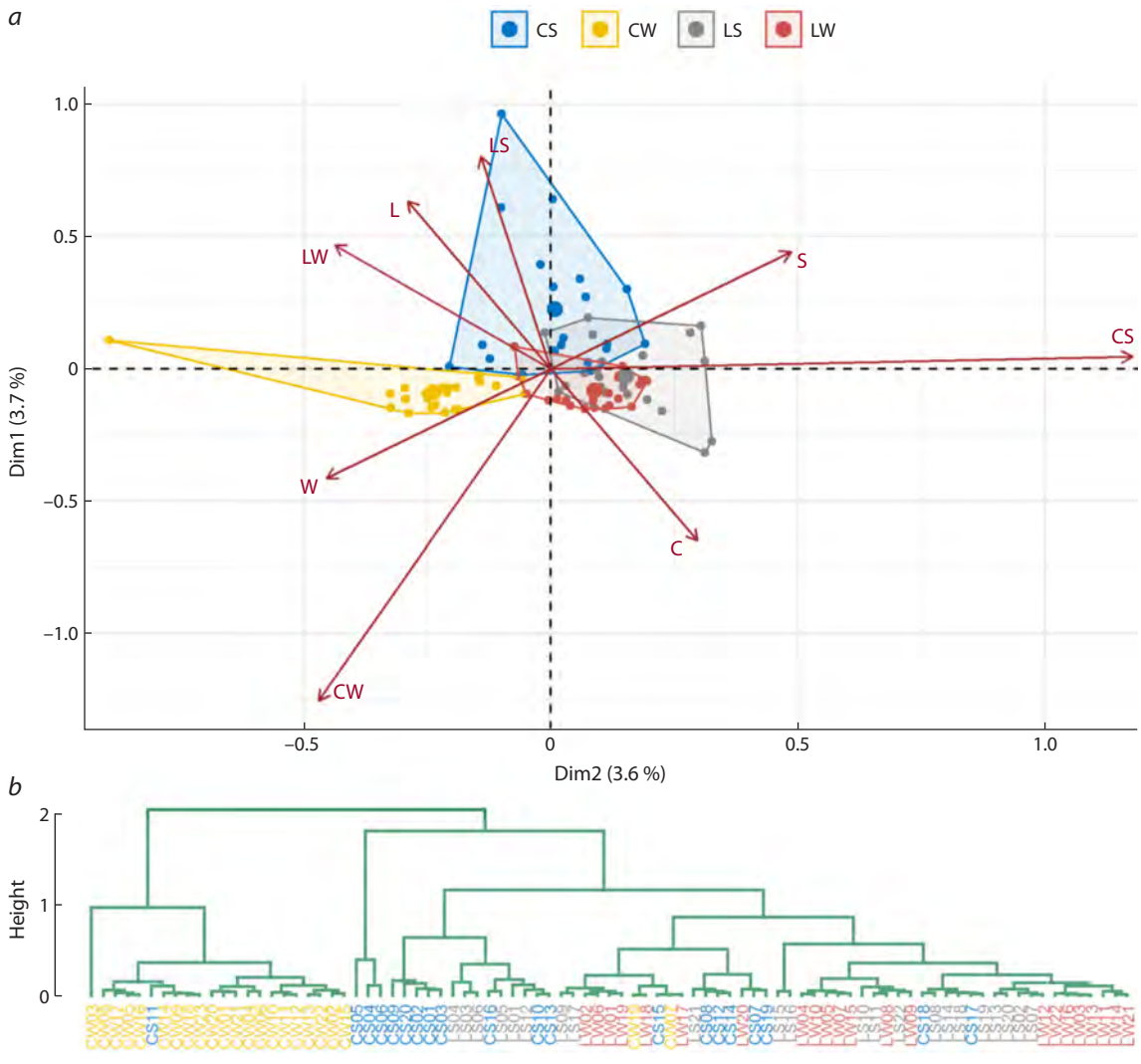


Fig. 3. Correspondence analysis of cultivars by chromosome type.

a – distribution of cultivars (W – winter, S – spring, L – landraces, C – modern) on the plane of the first two principal dimensions; *b* – dendrogram calculated from the coordinates of cultivars on the plane of principal dimensions.

with both dimensions (r = 0.27 and 0.25 [p > 0.20] with the second and first dimensions, respectively). Thus, the splitting of the sample is detected only among modern cultivars, but not landraces. In this case, the allocation of CW is mainly due to their breeding status, which does not, however, affect the distribution of CS.

On the plane of the main dimensions, it is also possible to see how much modern cultivars differ in the growth habit, the areas of clusters of winter and spring samples not overlapping, and how tightly the landraces are merged according to the same characteristic. With the exception of CW03, CW are the most homogeneous (CB = 0.42), they form a compact group in which a dense core of seven cultivars can be distinguished (CW02, CW09, CW14, CW15, CW18, CW21, and CW23). Landraces are also grouped relatively compactly (CB = 0.89), while CS form the most diffuse cluster (CB = 1.02), and therefore, are the most heterogeneous.

In view of the discovery of new factors that reveal hidden trends in the distribution of cultivars, a comparative analysis of their hierarchical clustering was carried out. In this regard, the coordinates of cultivars on the plane of principal dimensions were utilized to calculate the distance matrix, and clustering of cultivars was carried out using a similar method of calculating topology as in the original data (Ward method). In the new version of the cladogram (Fig. 3*b*), despite the complete restructuring of the topology (topology entanglement 97 %, topological distance 166 branches), the separation of modern cultivars has significantly improved. In particular, the CW cluster became more homogeneous, landraces and all spring cultivars except CS11 were completely excluded from it, sample CW03, carrying three chromosomal introgressions from *T. miguschovae*, stood out, and CW07 moved to the cluster of landraces. The cluster of CS cultivars split, but they still grouped with landraces, mainly of the spring growth habit. The distances between cultivars also changed, but their correlation remained at a high level (r = 0.80, p < 1e–5).

Quantitative assessment of karyograms. Size of C-bands as a differentiating feature

Association of C-bands with the breeding status and the growth habit of cultivars

During the quantitative assessment of karyograms by the size of C-bands, 98 diagnostic categories were identified (Fig. S5). The association of the C-bands with the tested trait was assessed using logistic regression, during which the statistical significance of the increase in the chances of detecting a C-band in cultivars with a given form of the tested characteristics was calculated with an increase in its size by one unit.

When assessing the odds ratio of detecting C-bands in cultivars that differ in breeding status, statistically significant results (p < 0.05) were obtained for 22 C-bands, with the number of C-bands more clearly detected in landraces being 1.4 times greater than that in modern ones. The same analysis revealed 27 C-bands associated with the growth habit of wheat, with most of them (60 %) being identified with the winter cultivars (Table S5). Similarly, among both winter and spring cultivars, the number of C-bands associated with landraces was greater than with modern ones, and with the winter growth habit, greater than with the spring growth habit, both among modern cultivars and landraces (Fig. 4*a*). However, in modern cultivars, twice as many C-bands associated with alternative growth habit were identified as in landraces, and in winter cultivars, the number of C-bands associated with alternative breeding status was detected to be greater than in spring cultivars. The same was true when testing the association of C-bands with combinations of the tested characteristics (Fig. 4*a*).

The obtained results are fully consistent with the association of the tested characteristics with the chromosome type (Fig. 1), confirming the higher homogeneity of modern cultivars than landraces, and winter cultivars than spring cultivars also in the quantitative assessment of karyograms. However, when comparing landraces and modern cultivars regardless of their growth habit or in subsamples of its alternative form, in landraces, a high probability of detection exists for a larger number of C-bands. The above-mentioned homogeneity of the sample of modern cultivars when testing a combination of characteristics is achieved due to a larger fraction of null alleles (the probability of the absence of a C-band on a chromosome, oddFC < 1), and in the case of testing subsamples from modern cultivars and landraces, it is due to better differentiation of karyograms of modern cultivars that differ in the growth habit. Thus, not only the greater disunity of spring and winter samples in the pool of modern cultivars than in landraces is confirmed, but also the ability to identify it in the course of associative analysis during the quantitative assessment of karyograms, as opposed to the qualitative assessment of karyograms.

It should be noted that the degree of association (the multiplicative factor of the odds ratio) was higher for most of the C-bands associated with CW (Fig. 4*b*). Thus, most of the C-bands (presence or absence) associated with modern cultivars or winter growth habit are specific for CW. Consequently, modern winter cultivars have twice as many specific karyological features (C-bands) that allow us to allocate them. In total, for all of the above-mentioned variants of analysis for the association of C-bands with the tested characteristics, less than half of them (45 %) were not informative.

Cluster analysis of cultivars by C-band size

Although the gap statistics algorithm predicts optimal seven clusters, the best resolution is provided with three. But even in this case, only two of them are almost completely separated (Fig. 5*a*, *b*). In one of these clusters, CW predominate (74 %), with 17 of the 23 modern winter cultivars included in this cluster. The other, largest cluster contains predominantly landraces (68 %), equally represented by the growth habit. The majority of CS (80 %) also entered here, confirming their gravitation towards landraces rather than modern cultivars. The high heterogeneity of the clusters by the tested characteristics makes this clustering approach not very useful for differentiation of tested cultivars.

The results of hierarchical clustering performed by the Ward method (predicting the best separation of cultivars

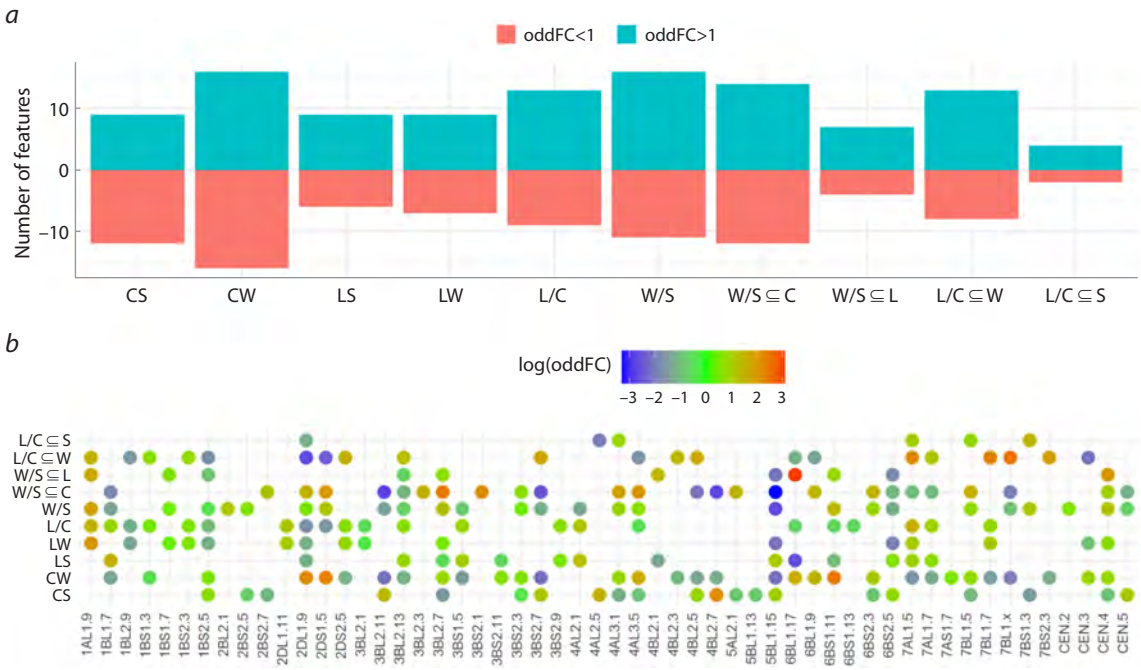


Fig. 4. Association of C-bands with the breeding status and the growth habit.

a – the number of C-bands associated with the breeding status, the growth habit and the combination of these characteristics in the total sample and in the subsamples of winter (W), spring (S), landraces (L) and modern (C) cultivars, as well as among winter and spring landraces (W⊆L, S⊆L) and modern (W⊆C, S⊆C) cultivars, and *vice versa*, among landraces and modern cultivars of spring (L⊆S, C⊆S) and winter (L⊆W, C⊆W) wheat; *b* – the multiplicative factor of the odds ratio for the C-band occurrence (oddFC) in the subset of cultivars similar by the tested characteristics from the total sample and subsamples.

by the tested characteristics and the assessment of topology parsimony, Table S6) are, on the whole (in terms of separation into major clades), similar to the above classification of cultivars by chromosome type, although the topologies of the corresponding dendrograms are completely entangled (up to 90 %, topological distance of 144 branches, Fig. S6, S7), and only 13 common clades, which, however, are untangled by 62 %, retain a topological distance of 14 branches (Fig. S8). In the new version, it should be noted that the CW clade is more homogeneous and is divided into two clusters of unequal volume (Fig. 5*c*). The validity of this division is confirmed by the high conjugacy of the sizes of a number of C-bands on chromosomes 1BS and 3BS (Fig. S9). In addition, CS and LW are also grouped more densely and uniformly. Finally, the topology of the dendrogram is distinguished by greater support for the monophyly of the clades (Q3 = 43.5 %, Fig. S10), which follows from the better conjugacy of the diagnostic features obtained during the quantitative assessment of the karyograms.

Factor analysis by C-band size

Factor analysis was carried out using the principal component (PC) method, utilizing the tested characteristics as additional variables. In the plane of the first two PCs (Fig. 6*a*), only CW is well separated, with a dense core of eight cultivars. Modern spring cultivars almost completely overlap with landraces, mainly also spring ones. This differs significantly from the distribution of these clusters observed during the analysis by chromosome type, where clusters of spring cultivars of alternative breeding status overlapped to a much lesser extent. Another distinctive feature of the results of this analysis is the emerging trend in the division of landraces by growth habit. Finally, in connection with the transition from a binary scale to a six-point scale, the density of clusters changes, they become more sparse, the resolution in the division of cultivars increases. Proportion of explainable variance falling on each of the components is less than 10 %, which confirms the weak consistency of diagnostic features. However, their total value (18.7 %) is 2.6 times greater than that obtained from the principal dimensions during the correspondence analysis of cultivars by chromosome type.

In this analysis, however, it is not the influence of combinations of features (size of C-bands) on the discrimination of cultivars that is of interest, but the identification of trends in the grouping of cultivars by the tested characteristics, which, however, weakly and statistically insignificantly ($p > 0.24$) correlate with the first PC. In this regard, it makes sense to consider the distribution of cultivars in the space formed by the second and third PCs (Fig. 6*b*, *c*), which are highly correlated with the tested characteristics. In particularly, the correlation of the growth habit is maximum with the second PC ($r = 0.71$), while the breeding status of cultivars is more strongly correlated with the third PC ($r = 0.57$). Since the correlation with the growth habit is higher than with the breeding status, the separation of cultivars by this characteristic is also better (*cos2* = 0.78 for winter and spring

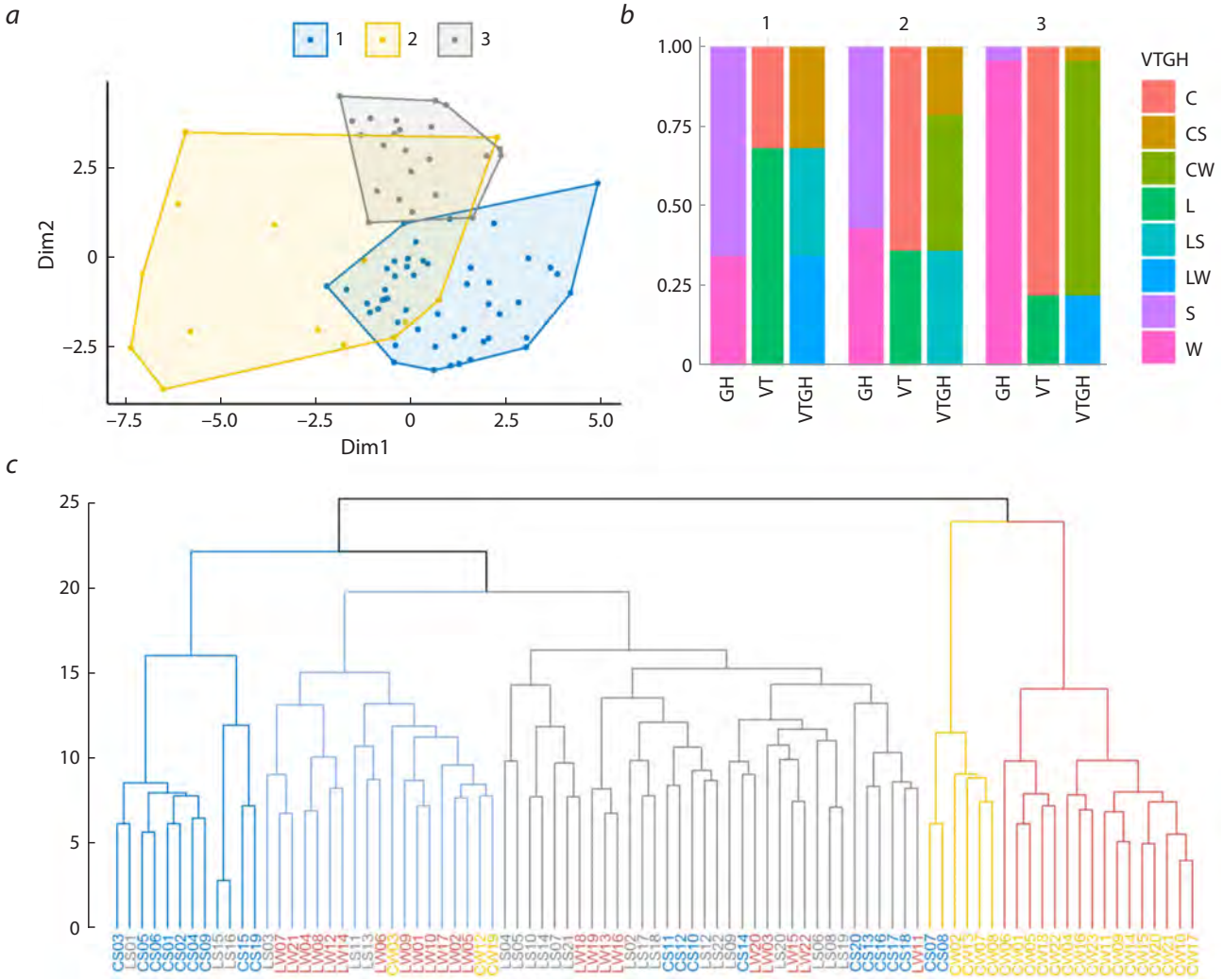


Fig. 5. Cluster analysis of cultivars by the size of C-bands.

a – visualization of clusters calculated with the K-means on the plane of principal coordinates; *b* – proportions of cultivars with alternative forms of the tested characteristics in these clusters (W – winter, S – spring, L – landraces, C – modern); *c* – dendrogram of hierarchical clustering of cultivars calculated by the Ward method (the first five clades are highlighted in color for clarity).

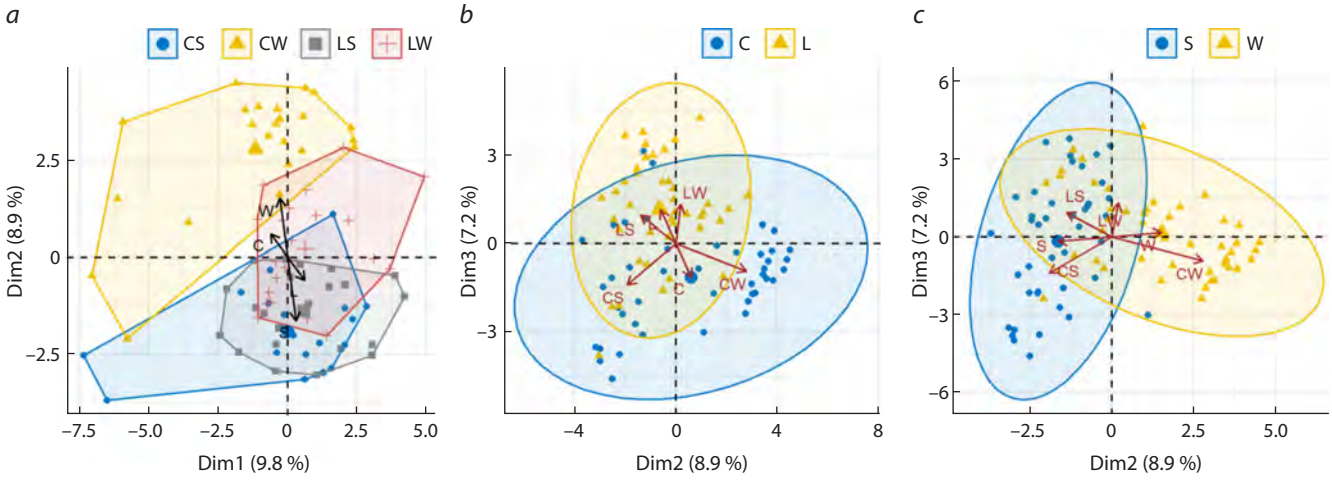


Fig. 6. Factor analysis by C-band size.

The distribution of cultivars is presented on the plane of the first two (*a*), as well as the second and third (*b*, *c*) principal components. In the latter case, groups of cultivars (W – winter, S – spring, L – landraces, C – modern) with identical breeding status (*b*) and growth habit (*c*) are distinguished.

types, and *cos2* = 0.52 for landrace and modern status); this is another difference from the correspondence analysis by chromosome type. Consequently, in addition to the clear allocation of CW, there is a tendency to separate cultivars by growth habit (Fig. 6*c*), which prevails over their breeding status (Fig. 6*b*), and this applies in particular to modern cultivars. CS are more strongly attracted to LS than CW to LW, as a result of which landraces are grouped more densely than modern ones. The movement of the LW cluster in the direction of the winter growth habit (W, the positive direction of the second dimension Dim2) shifts the perpendicularly directed influence of their landrace status (L, the positive direction of the third dimension Dim3); as a result, this cluster moves diagonally (Fig. 6*a*). It is interesting to note the influence of modern breeding on the isolation of winter wheat. On LW, the breeding status has a much stronger influence than the growth habit (*cos2* = 0.32, with Dim3, and *cos2* = 0.008, with Dim2), whereas in the CW cluster, the influence of the growth habit prevails (*cos2* = 0.08, with Dim3, and *cos2* = 0.74, with Dim2).

Comparative analysis of karyological classification with SNP genotyping

To assess the advantages and disadvantages of the karyological approach to classifying wheat cultivars in comparison with SNP genotyping, their resolution in differentiating a sample of 16 winter cultivars, contrasting in breeding status, and four modern spring cultivars was tested. The analysis was carried out on 120 features during assessment by chromosome type, 88 features during quantitative evaluation of karyograms by the size of C-bands, and 3,126 features during SNP genotyping; in the latter case, only unambiguously interpretable SNP markers were used, excluding heterozygotes.

When comparing the qualitative method of evaluating karyograms with SNP genotyping, the correlation of distances between cultivars was very low ($r = 0.22$, $p > 0.07$), the entanglement of the topologies of the dendrograms calculated by the Ward method was no less than 40 %, with a topological distance of 32 branches. Even the minimum estimate of the preservation of monophyly of the clades calculated based on the results of SNP genotyping (48.7 %) exceeded the third quartile when using the chromosome type as a diagnostic feature (Q3 = 34.5 %) (Fig. 7*a, c*). On the plane of the main dimensions, all three groups of cultivars are completely separated by chromosome type (Fig. 7*d*). According to the results of SNP genotyping, the cultivars are well separated only by growth habit, while the winter wheat group is not distinguishable by breeding status (Fig. 7*b*). In this regard, the advantage of karyological analysis by chromosome type in classifying cultivars by breeding status seems obvious, due to its greater resolution at the given discrimination criteria, despite the smaller number of diagnostic features and their low consistency.

The use of C-band size as a diagnostic feature led to an even greater decrease in the similarity between dendrograms, increasing the entanglement of topologies to more than 80 % and increasing the topological distance to 36 branches, with a complete loss of distance correlation. Otherwise, the differences between alternative approaches to assessing karyograms in relation to SNP genotyping are similar to the previously noted features of the quantitative method: better consistency of the original data in supporting the monophyly of the clades (Q3 = 56.4 %, Fig. 7*e*), allocation of CW as a more sparse cluster, and intersection of CS clusters with LW (Fig. 7*f*). Thus, regardless of the approach to evaluating karyograms, the significant differences between karyological classification and SNP genotyping include the distinct separation of modern winter cultivars and the CS tendency toward landraces, which is more pronounced in the classification by C-band size.

Discussion

The present study is aimed at assessing the potential of karyological analysis in differentiating such complex objects as varietal populations using modern cultivars and landraces as an example. In this regard, 1) two methods of obtaining diagnostic features from karyograms were tested for suitability for discrimination, based on the limits of their variability (degree of polymorphism), informativeness and resolution; 2) an assessment of the advantages and disadvantages of this approach was carried out in comparison with SNP genotyping; 3) an analysis of karyological classifications was performed for the presence of trends in the distribution of cultivars in accordance with their breeding status and growth habit, as well as an assessment of their contribution to the observed clustering and evaluation of its statistical significance.

The diagnostic features used were based on the variability of constitutive heterochromatin blocks. Two alternative approaches to its assessment were tested for resolution in discrimination of Russian common wheat cultivars by their breeding status and growth habit. The approach based on the qualitative assessment of the karyogram uses a binary scale reflecting the presence/absence of the C-band on the chromosome (distribution of C-bands along the chromosome length). The disadvantage of this approach is that chromosomes differing only in one C-band will be classified as different types, and the difference between them will be assessed in the same way as between chromosomes with significantly different karyograms. For instance, in the classification based on this method, the CW03 sample with three unique introgressions from *T. miguschovae* is distinguished as a separate branch on the dendrogram and is highly isolated from the other cultivars in the space of the main dimensions during the correspondence analysis.

Assuming a reduction in the number of distinguished categories due to the unification of similar chromosome types, and, consequently, a more adequate classification of cultivars, a transition was made to a six-point scale for assessing karyograms by the size of individual C-bands (quantitative assessment). Indeed, when using this approach, the number of categories allocated was halved (from 205 to 98), but, on average, the resolution of cultivars increased by the same amount (Fig. S11). The latter is due to a threefold expansion of the range of the evaluation scale, which allows for a more in-depth study of the genesis of cultivars, since the higher the resolution of the method, the more common

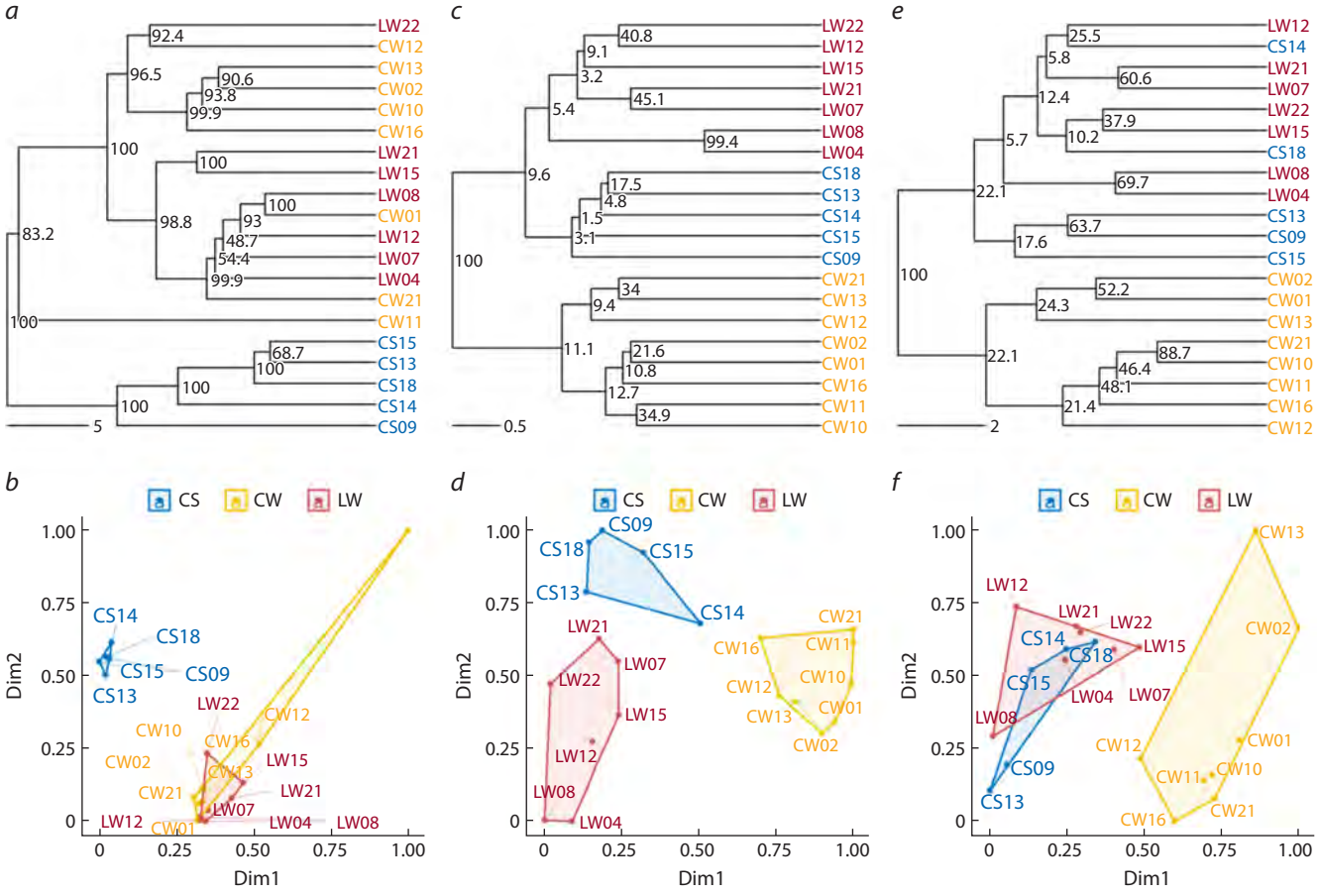


Fig. 7. Comparative analysis of karyological classification with SNP genotyping.

Dendrograms and distribution of cultivars (W – winter, S – spring, L – landraces, C – modern) on the plane of principal coordinates are presented according to the results of SNP genotyping (*a* and *b*), chromosome type analysis (*c* and *d*), and C-band size (*e* and *f*).

features among the classified samples it is able to detect. In the course of the principal component analysis, this led not only to an increase in the proportion of explainable variance in the karyograms, but also to the revelation of the importance of less obvious factors in the systematization of the studied material. In particular, while in the classification of cultivars by chromosome type (qualitative assessment of karyograms), the "breeding factor" is of primary importance, in the classification based on the quantitative assessment of karyograms (size of C-bands), a more significant contribution of the growth habit is found. In hierarchical clustering, the advantage of moving to a six-point scale is expressed in a denser and more homogeneous grouping of cultivars by the considered characteristics, justifying the expected increase in the adequacy of their classification. Support for the monophyly of the clades was also higher, indicating a better consistency of the initial data (conjugacy of diagnostic features). Finally, in the study of the association of alternative differentiating features with the breeding status and the growth habit of cultivars, despite the reduction of diagnostic categories, the uninformative fraction decreased from 70 to 45 %. However, despite the above-mentioned advantages of this approach, its resolution in discriminating modern spring cultivars from landraces is significantly inferior to the qualitative approach in assessing karyograms. In addition, in interpreting the results, one should accept the risks of the influence of the subjective factor when assessing the size of C-bands.

Since C-bands represent segments of inactive chromatin characterized by a high degree of evolutionary conservatism, their use as classification features is more justified in the study of the genesis of cultivars than the specificity of their genotypes. We also assume that the karyological features of this type are predominantly selection-neutral and are not subject to the adaptive action of natural selection. In this regard, any association of a karyogram and a genotype with a specific growth habit reflects only the commonality of their origin. In view of the peculiarities of intensive selection, it is likely that such associations will be more strongly expressed among modern cultivars than among landraces. Since varietal populations are locally adapted to the ecological and geographical conditions of the cultivation region, a relationship is expected between the wheat growth habit and its origin. We also hypothesized that more homogeneous modern cultivars, which were often obtained using similar sets of elite founder cultivars, would be characterized by a higher correlation of

the karyogram with the growth habit than heterogeneous and locally adapted landraces. This assumption is based on the fact that genetic drift events such as the "bottleneck" and the "founder effect" occur during domestication, and modern cultivars experience the impact of these evolutionary events again when they are separated from the pool of locally cultivated wheats (Ladizinsky, 1985; Tanksley, McCouch, 1997; Feldman, Levy, 2023). For instance, a study of a sample of Chinese bread wheat cultivars based on FISH analysis revealed a significant role of the "founder effect" in their clustering (Huang et al., 2018; Hu et al., 2022). As a diagnostic feature, the authors of these works took the similarity of the distribution patterns of several polymorphic DNA probes, which corresponds to the qualitative feature "chromosome type" in our study. In addition, the more intensively the cultivars are involved in breeding, the more they diverge and become isolated, which leads to a narrowing of the genetic diversity of modern cultivars (Reif et al., 2005; Haudry et al., 2007; Sindhu, 2022). Based on the fact that the number of winter common wheat cultivars approved for use in the territory of the Russian Federation significantly exceeds the number of spring cultivars (432 and 336, respectively; State Register..., 2024), it can be assumed that the breeding of modern winter wheat is carried out more intensively, and it should be more isolated from landraces. Indeed, the factor analysis showed a strong association of the growth habit with the classification of modern cultivars by chromosome type, while the karyograms of landraces are poorly distinguishable by this characteristic. At the same time, with an increase in the resolution of the method, the association of the growth habit with chromosomal differentiation slightly increased, with a tendency to separate landraces by growth habit. Earlier, when analyzing 20 Russian cultivars using SSR and ISBP markers, a division into subclusters of winter and spring forms was also noted (Adonina et al., 2017).

The study revealed different effects of modern breeding on the isolation of winter and spring wheat. In particularly, modern winter cultivars, regardless of the type of differentiating feature (chromosome type or C-band size) and classification approach, due to the greater similarity of karyotypes, are distinguished as a strictly delimited group, in which a dense core of 7–8 cultivars with a very similar karyogram is found. Thus, the growth habit of these cultivars is not only associated with the corresponding alleles of the vernalization genes, but reflects a karyotype that is separate from spring and landrace cultivars, formed in the course of long-term and targeted breeding. This observation confirms the greater commonality in the genealogy of modern winter cultivars (Novoselskaya-Dragovich et al., 2015). This is also evidenced by their high homogeneity, according to the number of associated karyological features of both types, and intervarietal distances. Indeed, most of the CW included in the present study had the Bezostaya 1 variety in their pedigree, which could probably also have affected the isolation of this group.

Modern spring cultivars, unlike winter cultivars, are probably less intensively involved in breeding. They have experienced less influence of artificial selection and have retained more affinity with landraces, mainly also of the spring growth habit. The latter is especially clearly evident in the classification based on the quantitative assessment of karyograms where the cluster of modern spring cultivars is almost completely overlapped with landraces, while none of them are included in the cluster of modern winter cultivars (Fig. 3*a*, 6*a*). Factor analysis also emphasizes the predominance of the growth habit over the breeding status in the distribution of cultivars. It should be noted that, despite the closeness of modern spring cultivars to landraces, both approaches to karyological classification reveal a tendency in their isolation from the latter. This is manifested both in the homogeneity of the CS clustering among landraces and in the allocation of some of them into an independent clade. Considering the complete separation of CW and CS, as well as the tendency in the separation of landraces by this characteristic, the growth habit appears to be a significant factor in the differentiation of cultivated wheat, the reproductive isolation of varietal populations, and their eco-geographical distribution.

When comparing the karyological method with SNP genotyping, both its key disadvantages and advantages in classifying wheat cultivars have been identified. Among the disadvantages, the limited number of diagnostic features and their low consistency are mainly highlighted. If the quantitative limitation is due to the nature of the features of this type, then the weak support for the monophyly of clades on the dendrogram is associated not only with their low conjugacy in this sample, but also follows from the first disadvantage (Rokas, Carroll, 2005). The low consistency of karyological characters can be explained by the high frequency of C-band recombination caused by both their high disunity on the chromosome and combinatorial events in the distribution of homologous chromosomes during meiosis (Blary, Jenczewski, 2019; Koo et al., 2020; Mason, Wendel, 2020; Fan et al., 2021), as well as changes in chromosome staining patterns as a result of introgression and chromosomal rearrangements. A certain contribution is made by the technology of preparation and staining of metaphase plate preparations, as well as a considerable amount of subjectivity in assessing the size of C-bands.

Considerations of the monophyly of clades come from the systematics of species based on their phylogeny. However, the assumption of complete reproductive isolation, usually applied to species, is unacceptable for freely interbreeding wheat cultivars, and in the obtaining of cultivars, as was noted in the introductory part of the article, samples from geographically separated populations and even representatives of related taxa were often used. The assessment of the monophyly of clades in the dendrograms presented in the current study reflects mainly the degree of conjugacy of diagnostic characters. Indeed, the support for the monophyly of the clades was somewhat higher when using the quantitative karyotyping method. And although the statistical reliability of monophyly was significant for all clades of the highest rank, regardless of the approach to assessing the karyograms,

it sharply decreased to complete disappearance when moving to clades of lower ranks. At the same time, the justification for monophyly was incomparably stronger and statistically significant for all clades of the dendrogram calculated using SNP genotyping, in which the number of diagnostic features was an order of magnitude greater, and their conjugacy was higher, due to the higher mapping density on chromosomes.

Nevertheless, the value of differentiating features in classification problems is determined by their ability to discriminate a sample by the required qualities. While SNP markers provide an idea of allelic, gene variability, polymorphism by chromosomal markers reflects the variability of the karyotype as a whole. Since the structure of heterochromatic C-bands is not subject to genetic recombination and conversion, they are evolutionarily more stable. Chromosomal aberrations, also related to chromosomal markers, lead to significant disturbances in the meiosis of heterozygotes, thereby leveling the heterogeneity of the population and contributing to the preservation of the authenticity of the original populations over time, as, for example, noted for the group of European emmer with the marker 5B-7A translocation (Dedkova et al., 2009; Badaeva et al., 2015b). Thus, chromosomal markers seem to be more reliable for systematization of wheat varietal sets from the point of view of reliability and reproducibility of classification, especially when the number of diagnostic features is limited. Combinatorial variability at the chromosome level forms new karyotypes (classification categories in karyological differentiation), but does not change the combinations of features in linkage groups (often reflected in conjugated markers in SNP genotyping). As a result, a larger number of agreed diagnostic features, in itself, does not yet mean better discrimination of their carriers, since the latter also depends on the type of these features. Indeed, comparison of SNP and chromosomal markers, in particular when using the chromosome type, but not the size of C-bands as a diagnostic feature, confirmed the superiority of the latter in discrimination of 20 selected wheat cultivars both by breeding status and by growth habit, despite their low consistency.

It is significant that the use of chromosomal markers reliably differentiates winter wheat cultivars depending on their breeding status, whereas in the case of SNP genotyping, these same cultivars are grouped together. As noted above, this circumstance may be due to the difference in the type of variability detected by SNP and chromosomal markers; in other words, karyotypes of different origin may not differ in the alleles of many genes (hence the high consistency of SNP markers). It is possible that an increase in the number of SNP markers will increase the resolution of this method in discriminating the tested sample of cultivars by breeding status. Nevertheless, the number of such markers already exceeds by more than an order of magnitude the minimum set of karyological characteristics sufficient for the successful resolution of this problem, emphasizing the advantage of the karyological method, especially with a limited number of diagnostic features.

Conclusion

The present study revealed a high potential of karyotyping by constitutive heterochromatin in discrimination of modern Russian cultivars of common wheat by the growth habit, and isolation of their winter samples from landraces. At the same time, the homogeneity of modern cultivars by the tested karyological features is higher than that of landraces, and that of winter cultivars is higher than that of spring cultivars. The resulting classification reflects the preservation of high similarity in the karyotypes of modern spring cultivars and varietal populations of landraces, as well as weak distinguishability of the karyotypes of the latter, contrasting by the growth habit. In connection with a more unambiguous association of differential chromosome staining patterns with the origin of cultivars, as well as their high evolutionary stability, it is assumed that the analysis of karyotypic variability helps to form a more accurate idea of the differentiation of wheat varietal sets according to this trait than when using SNP markers that detect gene variability, especially with a limited number of diagnostic characteristics.

References

Adonina I.G., Leonova I.N., Badaeva E.D., Salina E.A. Genotyping of hexaploid wheat varieties from different Russian regions. *Russ J Genet Appl Res*. 2017;7(1):6-13. doi 10.1134/S2079059717010014

Afonnikova S.D., Kiseleva A.A., Fedyaeva A.V., Komyshev E.G., Koval V.S., Afonnikov D.A., Salina E.A. Identification of novel loci precisely modulating pre-harvest sprouting resistance and red color components of the seed coat in *T. aestivum* L. *Plants*. 2024;13:1309. doi 10.3390/plants13101309

Badaeva E.D., Shkutina F.M., Bogdevich I.N., Badaev N.S. Comparative study of *Triticum aestivum* and *T. timopheevi* genomes using C-banding technique. *Plant Syst Evol*. 1986;154:183-194. doi 10.1007/BF00990122

Badaeva E.D., Sozinova L.F., Badaev N.S., Muravenko O.V., Zelenin A.V. "Chromosomal passport" of *Triticum aestivum* L. *em Thell*. cv. Chinese Spring and standardization of chromosomal analysis of cereals. *Cereal Res Commun*. 1990;18(4):273-281

Badaeva E.D., Badaev N.S., Gill B.S., Filatenko A.A. Intraspecific karyotype divergence in *Triticum araraticum* (*Poaceae*). *Plant Syst Evol*. 1994;192:117-145. doi 10.1007/BF00985912

Badaeva E.D., Dedkova O.S., Gay G., Pukhalskyi V.A., Zelenin A.V., Bernard S., Bernard M. Chromosomal rearrangements in wheat: their types and distribution. *Genome*. 2007;50(10):907-926. doi 10.1139/g07-072

Badaeva E.D., Dedkova O.S., Pukhalskyi V.A., Zelenin A.V. Chromosomal changes over the course of polyploid wheat evolution and domestication. In: Ogihara Y., Takumi S., Handa H. (Eds) Advances in Wheat Genetics: From Genome to Field. Tokyo: Springer, 2015a; 83-89. doi 10.1007/978-4-431-55675-6_9

Badaeva E.D., Keilwagen J., Knüpffer 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*. 2015b;10(5):e0128556. doi 10.1371/journal.pone.0128556

Badaeva E.D., Konovalov F.A., Knüpffer H., Fricano A., Ruban A.S., Kehel Z., Zoshchuk S.A., … Filatenko A., Bogaard A., Jones G., Özkan H., Kilian B. Genetic diversity, distribution and domestication history of the neglected GGA^tA^t genepool of wheat. *Theor Appl Genet*. 2022;135(3):755-776. doi 10.1007/s00122-021-03912-0

Becker R.A., Chambers J.M., Wilks A.R. The New S Language: A Programming Environment for Data Analysis and Graphics. Berlin: Springer, 1988

Blary A., Jenczewski E. Manipulation of crossover frequency and distribution for plant breeding. *Theor Appl Genet*. 2019;132(3):575-592. doi 10.1007/s00122-018-3240-1

Boehm J., Jr., Cai X. Enrichment and diversification of the wheat genome via alien introgression. *Plants*. 2024;13(3):339. doi 10.3390/plants13030339

Dedkova O.S., Badaeva E.D., Mitrofanova O.P., Zelenin A.V., Pukhalskiy V.A. Analysis of intraspecific divergence of hexaploid wheat *Triticum spelta* L. by C-banding of chromosomes. *Russ J Genet*. 2004;40:1111-1126. doi 10.1023/B:RUGE.0000044755.18085.7e

Dedkova O., Badaeva E., Mitrofanova O., Bilinskaya E., Pukhalskiy V. Analysis of intraspecific diversity of cultivated emmer *Triticum dicoccum* (Schrank.) Schuebl using C-banding technique. *Russ J Genet*. 2007;43:1271-1285. doi 10.1134/S1022795407110105

Dedkova O., Badaeva E., Amosova A., Martynov S., Ruanet V., Mitrofanova O., Pukhal'skiy V. Diversity and the origin of the European population of *Triticum dicoccum* (Schrank) Schuebl. As revealed by chromosome analysis. *Russ J Genet*. 2009;45:1082-1091. doi 10.1134/S1022795409090099

Delaney D.E., Nasuda S., Endo T.R., Gill B.S., Hulbert S.H. Cytologically based physical maps of the group-2 chromosomes of wheat. *Theor Appl Genet*. 1995;91(4):568-573. doi 10.1007/BF00223281

Endo T.R., Gill B.S. The deletion stocks of common wheat. *J Hered*. 1996;87(4):295-307. doi 10.1093/oxfordjournals.jhered.a023003

Fadeeva T.S., Narbut S.I. Genetic characteristics of varieties and characteristics of pure-line varieties. *Vestnik Leningradskogo Gosudarstvennogo Universiteta = Herald of the Leningrad State University.* 1969;21(4):123-131 (in Russian)

Fan C., Hao M., Jia Z., Neri C., Chen X., Chen W., Liu D., Lukaszewski A.J. Some characteristics of crossing over in induced recombination between chromosomes of wheat and rye. *Plant J*. 2021;105(6):1665-1676. doi 10.1111/tpj.15140

Feldman M. Origin of cultivated wheat. In: Bonjean A.P., Angus W.J. (Eds) The World Wheat Book: A History of Wheat Breeding. Technique & Doc, 2001;3-56

Feldman M., Levy A.A. Wheat Evolution and Domestication. Springer, 2023. doi 10.1007/978-3-031-30175-9

Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39(4):783-791. doi 10.1111/j.1558-5646.1985.tb00420.x

Friebe B., Gill B.S. C-band polymorphism and structural rearrangements detected in common wheat (*Triticum aestivum*). *Euphytica*. 1994;78:1-5. doi 10.1007/BF00021392

Friebe B., Gill B.S. Chromosome banding and genome analysis in diploid and cultivated polyploid wheats. In: Jauhar P.P. (Ed.) Methods in Genome Analysis in Plants. CRC Press, 1996;39-60

Galili T. dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*. 2015;31(22):3718-3720. doi 10.1093/bioinformatics/btv428

Gill B.S., Friebe B., Endo T.R. Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*). *Genome*. 1991;34(5):830-839. doi 10.1139/g91-128

Goncharov N.P. Comparative Genetics of Wheats and their Related Species. Novosibirsk: Acad. Publ. House "Geo", 2012 (in Russian)

Guo J., Gao D., Gong W., Li H., Li J., Li G., Song J., Liu J., Yang Z., Liu C. Genetic diversity in common wheat lines revealed by fluorescence in situ hybridization. *Plant Syst Evol*. 2019;305:247-254. doi 10.1007/s00606-019-1567-y

Haudry A., Cenci A., Ravel C., Bataillon T., Brunel D., Poncet C., Hochu I., Poirier S., Santoni S., Glémin S., David J. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol Biol Evol*. 2007;24(7):1506-1517. doi 10.1093/molbev/msm077

Hu Z., Luo J., Wan L., Luo J., Li Y., Fu S., Liu D., Hao M., Tang Z. Chromosomes polymorphisms of Sichuan wheat cultivars displayed by ND-FISH landmarks. *Cereal Res Commun*. 2022;50:253-262. doi 10.1007/s42976-021-00173-x

Huang X., Zhu M., Zhuang L., Zhang S., Wang J., Chen X., Wang D., … Chu C., Du P., Qi Z., Wang H., Chen P. Structural chromosome rearrangements and polymorphisms identified in Chinese wheat cultivars by high-resolution multiplex oligonucleotide FISH. *Theor Appl Genet*. 2018;131(9):1967-1986. doi 10.1007/s00122-018-3126-2

Husson F., Le S., Pages J. Exploratory Multivariate Analysis by Example Using R. Chapman and Hall, 2010

Iordansky A.B., Zurabishvili T.B., Badaev N.S. Linear differentiation of cereal chromosomes. I. Common wheat and its supposed ancestors. *Theor Appl Genet*. 1978a;51(4):145-152. doi 10.1007/BF00273138

Iordansky A.B., Zurabishvili T.G., Badaev N.S. Linear differentation of cereal chromosomes: III. Rye, triticale and 'Aurora' variety. *Theor Appl Genet*. 1978b;51(6):281-288. doi 10.1007/BF00274816

Jiang J., Friebe B., Gill B.S. Recent advances in alien gene transfer in wheat. *Euphytica*. 1993;73:199-212. doi 10.1007/BF00036700

Jiang J., Friebe B., Gill B.S. Chromosome painting of Amigo wheat. *Theor Appl Genet*. 1994;89(7-8):811-813. doi 10.1007/BF00224501

Jiang M., Xaio Z.Q., Fu S.L., Tang Z.X. FISH karyotype of 85 common wheat cultivars/lines displayed by ND-FISH using oligonucleotide probes. *Cereal Res Commun*. 2017;45:549-563. doi 10.1556/0806.45.2017.049

Koo D.H., Friebe B., Gill B.S. Homoeologous recombination: a novel and efficient system for broadening the genetic variability in wheat. *Agronomy*. 2020;10(8):1059. doi 10.3390/agronomy10081059

Kudriavtsev A.M. Intravarietal heterogeneity of durum wheat is an important component of species biodiversity. *Russ J Genet*. 2006;42(10):1437-1440. doi 10.1134/S1022795406100139

Ladizinsky G. Founder effect in crop-plant evolution. *Econ Bot*. 1985;39:191-199. doi 10.1007/BF02907844

Mason A.S., Wendel J.F. Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. *Front Genet*. 2020;11:1014. doi 10.3389/fgene.2020.01014

Metakovsky E., Melnik V.A., Pascual L., Wrigley C.W. How important are genetic diversity and cultivar uniformity in wheat? The case of Gliadins. *Genes* (*Basel*). 2024;15(7):927. doi 10.3390/genes15070927

Mickelson-Young L., Endo T.R., Gill B.S. A cytogenetic ladder map of the wheat homoeologous group 4 chromosomes. *Theor Appl Genet*. 1995;90(7-8):1007-1011. doi 10.1007/BF00222914

Mujeeb-Kazi A., Kazi A.G., Dundas I., Rasheed A., Ogbonnaya F.C., Kishii M., Bonnett D., Wang R.C., Xu S., Chen P., Mahmood T., Bux H., Farrakh S. Genetic diversity for wheat improvement as a conduit to food security. In: Sparks D.L. (Ed.) Advances in Agronomy. Acad. Press, Burlington, 2013;179-257. doi 10.1016/B978-0-12-417187-9.00004-8

Novoselskaya-Dragovich A.Y., Bespalova L.A., Shishkina A.A., Melnik V.A., Upelniek V.P., Fisenko A.V., Dedova L.V., Kudryavtsev A.M. Genetic diversity of common wheat varieties at the gliadin-coding loci. *Russ J Genet*. 2015;51(3):262-271. doi 10.1134/S1022795415030102

Paradis E., Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019;35(3):526-528. doi 10.1093/bioinformatics/bty633

Reif J.C., Zhang P., Dreisigacker S., Warburton M.L., van Ginkel M., Hoisington D., Bohn M., Melchinger A.E. Wheat genetic diversity trends during domestication and breeding. *Theor Appl Genet*. 2005;110(5):859-864. doi 10.1007/s00122-004-1881-8

Robinson D.F., Foulds L.R. Comparison of phylogenetic trees. *Math Biosci*. 1987;53(1-2):131-147. doi 10.1016/0025-5564(81)90043-2

Rokas A., Carroll S.B. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Mol Biol Evol*. 2005;22(5):1337-1344. doi 10.1093/molbev/msi121

Schliep K.P. phangorn: phylogenetic analysis in R. *Bioinformatics*. 2011;27(4):592-593. doi 10.1093/bioinformatics/btq706

Seal A.G. C-banded wheat chromosomes in wheat and triticale. *Theor Appl Genet*. 1982;63(1):39-47. doi 10.1007/BF00303488

Serpolay-Besson E., Dawson J., Chable V., van Bueren E.L., Osman A., Pino S., Silveri D., Goldringer I. Diversity of different farmer and modern wheat varieties cultivated in contrasting organic farming conditions in western Europe and implications for European seed and variety legislation. *Org Agric*. 2011;1(3):127-145. doi 10.1007/s13165-011-0011-6

Sharma M., Punya S., Gupta B.B. Role of wild relatives for development of climate-resilient varieties. In: Salgotra R.K., Zargar S.M. (Eds) Rediscovery of Genetic and Genomic Resources for Future Food Security. Springer, Singapore, 2020;303-314. doi 10.1007/978-981-15-0156-2_11

Sharma S., Schulthess A.W., Bassi F.M., Badaeva E.D., Neumann K., Graner A., Özkan H., Werner P., Knüpffer H., Kilian B. Introducing beneficial alleles from plant genetic resources into the wheat germplasm. *Biology* (*Basel*). 2021;10(10):982. doi 10.3390/biology10100982

Sindhu A. Wheat genetic diversity trends as a result of domestication and plant breeding. *Pharma Innovation.* 2022;11(6):427-433

State Register of Varieties and Hybrids of Agricultural Plants Admitted for Usage (National List): official publication. Moscow: Rosinformagrotech Publ., 2024 (in Russian)

Tanksley S.D., McCouch S.R. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*. 1997;277(5329):1063-1066. doi 10.1126/science.277.5329.1063

Zeven A.C. Polyploidy and domestication: the origin and survival of polyploids in cytotype mixtures. In: Lewis W.H. (Ed.) Polyploidy. Basic Life Sciences. Vol. 13. Springer, 1980;385-407. doi 10.1007/978-1-4613-3069-1_20

Zeven A.C. Landraces: a review of definitions and classifications. *Euphytica*. 1998;104:127-139. doi 10.1023/A:1018683119237

Zhao X., Guo Y., Kang L., Yin C., Bi A., Xu D., Zhang Z., … Kear P., Wang J., Liu Z., Fu X., Lu F. Population genomics unravels the Holocene history of bread wheat and its relatives. *Nat Plants*. 2023;9(3):403-419. doi 10.1038/s41477-023-01367-3

Zohary D., Hopf M., Weiss E. Domestication of Plants in the Old World. Oxford Univ. Press, 2012

Zurabishvili T.G., Iordansky A.B., Badaev N.S. Linear differentiation of cereal chromosomes. II. Polyploid wheats. *Theor Appl Genet*. 1978;51(5):201-210. doi 10.1007/BF00273766

Conflict of interest. The authors declare no conflict of interest.
Received March 6, 2025. Revised May 19, 2025. Accepted May 26, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):769-778

doi 10.18699/vjgb-25-84

Mapping loci and genes controlling heading and maturity time in common wheat under long-day conditions and assessing their effects on yield-related traits

A.A. Kiseleva, A.I. Stasyuk, I.N. Leonova, E.A. Salina

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
antkiseleva@bionet.nsc.ru

Abstract. The duration of the vegetation period significantly impacts yield formation and is one of the important characteristics of spring common wheat (*Triticum aestivum* L.) varieties. The primary developmental phases influencing the vegetation period include the time from seedling emergence to heading and from heading to maturity. To identify genes and loci associated with these traits under long-day conditions typical of Western Siberia and to assess their impact on yield components, we conducted QTL mapping followed by an evaluation of yield-related traits in lines carrying different alleles of key heading time genes. For mapping, we used an F_2 population derived from a cross between the varieties Obskaya 2 and Tulun 15, which contrast in their heading and maturity times. QTL analysis identified a novel locus on the long arm of chromosome 7B associated with maturity time, as well as two loci on chromosome 2D and the short arm of chromosome 7B associated with heading time. Gene analysis within these loci revealed candidate genes for the "seedling-maturity" trait, with expression patterns corresponding to the known maturity time regulator *NAM-1*. The localization of loci for the "seedling-to-heading" trait suggested their correspondence to the well-known genes *Ppd-D1* and *Vrn-B3*. Analysis of progeny carrying the *Ppd-D1a* and *Vrn-B3a* allele combination demonstrated that *Ppd-D1a* had a stronger effect on heading time than *Vrn-B3a*, and their combined presence resulted in the earliest heading – on average, five days earlier than in lines with the *Ppd-D1b* and *vrn-B3* alleles. Evaluation of yield-related traits (number and weight of grains per main spike and per plant, and 1,000-grain weight) indicated that *Ppd-D1* was significantly associated with all traits, with the *Ppd-D1*a allele generally exerting a negative effect on yield. In contrast, *Vrn-B3* had a comparatively smaller effect on yield traits than *Ppd-D1*.

Key words: common wheat; heading time; maturity time; yield traits; QTL mapping; *Ppd-D1*; *Vrn-B3*

For citation: Kiseleva A.A., Stasyuk A.I., Leonova I.N., Salina E.A. Mapping loci and genes controlling heading and maturity time in common wheat under long-day conditions and assessing their effects on yield-related traits. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):769-778. doi 10.18699/vjgb-25-84

Funding. The QTL analysis and the study of loci associated with heading and maturity time were conducted within the framework of the budget project "Youth Laboratories" FWNR-2024-0009. The evaluation of the effects of the *Ppd-D1a* and *Vrn-B3a* alleles on productivity traits was supported by the Russian Science Foundation No. 21-76-30003-П, https://rscf.ru/project/21-76-30003/.

Картирование локусов и генов, определяющих время колошения и созревания яровой мягкой пшеницы в условиях длинного дня, и оценка их влияния на урожайность

А.А. Киселёва, А.И. Стасюк, И.Н. Леонова, Е.А. Салина

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
antkiseleva@bionet.nsc.ru

Аннотация. Продолжительность вегетационного периода оказывает значительное влияние на формирование урожая и является одной из важных характеристик сортов яровой мягкой пшеницы (*Triticum aestivum* L.). К основным межфазным периодам, влияющим на продолжительность вегетации, относятся время от всходов до колошения и от колошения до созревания. Для выявления генов и локусов, ассоциированных с этими признаками в условиях длинного дня, характерного для Западной Сибири, и оценки их влияния на структуру урожая мы провели картирование QTL с последующей оценкой признаков урожайности у линий, несущих различные аллели генов, определяющих скорость развития. В качестве картирующей популяции мы использовали растения F_2, полученные от скрещивания контрастных по скорости развития сортов Обская 2 и

Тулун 15. QTL анализ выявил новый локус на длинном плече хромосомы 7B, ассоциированный со временем созревания, и два локуса на хромосоме 2D и коротком плече хромосомы 7B, ассоциированных со временем колошения. Анализ генов, входящих в состав локусов, позволил выявить гены-кандидаты для признака «всходы–созревание», паттерны экспрессии которых соответствовали генам известного регулятора скорости созревания *NAM-1*. Локализация локусов для признака «всходы–колошение» позволила предположить, что они соответствуют известным генам *Ppd-D1* и *Vrn-B3*. Анализ потомства линий с сочетанием аллелей *Ppd-D1a* и *Vrn-B3a* показал, что *Ppd-D1a* оказывает больший эффект на время колошения, чем *Vrn-B3a*, а сочетание этих двух аллелей приводит к наиболее раннему колошению, в среднем на пять дней раньше линий с аллелями *Ppd-D1b* и *vrn-B3*. Оценка признаков структуры урожая (количество и масса зерен с главного колоса и с растения, масса 1000 зерен) показала, что ген *Ppd-D1* ассоциирован со всеми признаками на высоком уровне значимости, при этом в большинстве случаев аллель *Ppd-D1a* негативно влиял на урожайность. Ген *Vrn-B3* влиял на признаки урожайности в меньшей степени по сравнению с *Ppd-D1*.
Ключевые слова: мягкая пшеница; время колошения; время созревания; урожайность; QTL картирование; *Ppd-D1*; *Vrn-B3*

Introduction

Wheat yield and its adaptability to diverse environmental conditions are largely determined by the duration of key developmental stages. The most critical stages in wheat development are the transition from vegetative to reproductive growth – namely, heading time – and maturity time. In variety evaluations, the prediction of heading time is most often based on the allelic composition of key regulatory genes, including *Vrn-A1*, *Vrn-B1*, *Vrn-D1*, *Ppd-D1*, and, less frequently, *Vrn-B3* (*TaFT-1*), as these genes exert the most substantial influence on this trait (Zhang Y. et al., 2010; Kiss et al., 2014; Chen S. et al., 2018; Mizuno et al., 2022; Palomino, Cabrera, 2023). The primary genes associated with maturity time are those belonging to the *NAM-1* family (Hagenblad et al., 2012; Alhabbar et al., 2018b).

Most *Ppd-1* alleles conferring photoperiod insensitivity are characterized by structural changes in the promoter region, such as deletions or insertions, which affect various regulatory sequences (Beales et al., 2007; Wilhelm et al., 2009; Nishida et al., 2013). Besides, dominant alleles known for *Ppd-B1* are characterized by an increased number of copies (Díaz et al., 2012). Among all *Ppd-1* genes, the dominant *Ppd-D1a* allele is currently the most widely utilized in global wheat breeding programs (Seki et al., 2011). According to Z. Guo et al., the *Ppd-D1a* allele has been identified in 33 % of common wheat cultivars in South America, 45.5 % of those cultivated in Southern Europe, and 8 % of varieties in Northern and Western Europe (Guo et al., 2010). The highest frequency of this allele has been reported in Asia, where it is present in 57.4 % of wheat cultivars grown in China. Among Japanese cultivars, 84 % carry this dominant allele (Seki et al., 2011). In contrast, the *Ppd-D1a* allele remains relatively rare among cultivars developed through Russian breeding programs, despite its potential not only to accelerate heading but also to positively influence other agronomic traits (Likhenko et al., 2014; Lysenko et al., 2014). In most studies, the effect of this allele has been examined under short-day conditions, where it shortens the time to heading by 20 to 30 days. However, there is limited evidence suggesting that the *Ppd-D1a* allele can also accelerate heading by 3 to 5 days even under long-day conditions (Worland et al., 1998; Kiseleva et al., 2014).

Another important gene is *Vrn-B3*, which serves as a central regulator of heading time. The *Vrn-B3a* allele promotes early flowering, with its expression enhanced by the insertion of a 5,300 bp retroelement into the promoter region (Yan et al., 2006). This allele is very rarely found in cultivated wheat varieties (Zhang X.K. et al., 2008; Iqbal et al., 2011; Chen F. et al., 2013; Lysenko et al., 2014). Additionally, four other alleles of this gene – designated as b, c, d, and e – have been identified, although they exhibit much weaker effects on heading time (Chen F. et al., 2013; Berezhnaya et al., 2021).

The wild-type allele of the *NAM-B1* gene is associated with earlier maturity; however, it is rarely found in modern cultivars due to its negative impact on yield (Lundström et al., 2017). Alleles of its homoeolog, *NAM-A1* – specifically *NAM-A1a* and *NAM-A1b* – have also been identified and are similarly associated with earlier maturity (Alhabbar et al., 2018a).

Environmental conditions significantly influence the developmental rate of common wheat. Climatic conditions vary considerably across different regions where wheat is cultivated in Russia. Despite the clear importance of investigating the regulation of heading and maturity times under long-day conditions – typical of most regions in Russia – the genetic mechanisms underlying these traits under such photoperiods remain poorly understood. For example, several studies have demonstrated the influence of *Vrn-1*, *Vrn-B3*, and *NAM-A1* on maturity time (Zaitseva, Lemesh, 2015; Alhabbar et al., 2018a; Whittal et al., 2018). However, in our previous research, no association was found between the allelic state of these genes and maturity time in spring wheat cultivars under the conditions of Western Siberia. Instead, novel loci associated with this trait were identified on chromosomes 2A, 3B, 4A, 5B, 7A, and 7B (Kiseleva et al., 2023).

Thus, the genetic control of wheat developmental rates is highly dependent on growing conditions, with different genes influencing the trait in distinct climatic zones. In Western Siberia, there is a particular need for early-maturing, high-yielding common wheat cultivars, as most varieties currently registered for this region are mid-season types. Cultivating spring wheat with a range of maturity times allows for a more flexible harvest schedule, which is crucial for minimizing yield losses due to over-ripening (Belan et al., 2021). This highlights the need for further analysis of known loci and identification of new loci and genes controlling the duration of key growth stages, as well as breeding of wheat cultivars and lines with heading and maturity times adapted to specific environmental conditions.

The objective of our study was to identify loci and their associated genes related to the duration of major develop-

mental phases in spring common wheat and to assess their effects on yield under the environmental conditions of Western Siberia.

Materials and methods

Plant material. The mapping population was developed from a cross between two spring common wheat cultivars, Obskaya 2 and Tulun 15. Obskaya 2 belongs to the group of mid-season cultivars and is characterized by high yield potential and baking quality comparable to that of premium wheat. Tulun 15 is an early-maturing cultivar with high grain quality, although it has lower yield performance compared to Obskaya 2. Hybridization of the parental cultivars, subsequent self-pollination of the F_1 hybrids, and cultivation of the F_2 generation were carried out under the greenhouse conditions of the Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (ICG SB RAS). The subsequent F_3 and F_4 generations were also obtained through self-pollination and grown under field conditions (Fig. S1)[1].

Phenotypic analysis. The F_3 and F_4 generation plants derived from the cross between Obskaya 2 and Tulun 15 were sown in 2018 and 2019 at the experimental field of the Siberian Research Institute of Plant Production and Breeding (Krasnoobsk, Novosibirsk Region; 54.914070°N, 82.975379°E).

Heading was defined as the stage when half of the wheat spike had emerged from the sheath, and heading time was recorded as the duration (in days) from seedling emergence to heading. Maturity was determined based on the hardening of the grain and the yellowing and drying of the spikes and stems. The grain filling period was calculated as the difference between maturity time and heading time. The soil at the experimental field was leached chernozem. Field trials were established on plots 0.5 meters wide, with 20 seeds sown per row. Each sample was sown in two rows with a 20 cm row spacing. Mature plants were harvested in bundles, dried, and subsequently used for yield component analysis. The number of grains per main spike, the grain weight per main spike, the total number and weight of grains per plant, and the 1000-grain weight were evaluated. Structural analysis was performed on 20 plants per sample.

Weather conditions in the Novosibirsk Region during the growing seasons deviated from the long-term averages. In May 2018, the mean monthly temperature was 7 °C, compared to the long-term average of 12.5 °C, while rainfall reached 82 mm – 2.5 times higher than the norm. In June, July, and August 2018, temperatures were close to long-term averages. Rainfall in June and July did not differ significantly from the norm; however, August was characterized by warm and dry conditions, with only 35 mm of rainfall compared to the average of 53 mm. In 2019, the temperature regime throughout the growing season was consistent with long-term averages. Moisture availability in May 2019 was comparable to the long-term norm. However, June and August experienced a moisture deficit, with only 26 mm and 22 mm of precipitation, respectively, compared to the averages of 59 mm and 53 mm. In contrast, July 2019 was warm and humid, with 98 mm of rainfall compared to the average of 69 mm.

[1] Figures S1, S2 and Tables S1–S8 are available at: https://vavilovj-icg.ru/download/pict-2025-29/appx26.xlsx

DNA extraction and PCR. Genomic DNA was extracted from wheat leaf tissue using a modified protocol based on the method published by J. Plaschke et al. (Plaschke et al., 1995). Allele-specific primers previously reported in the literature were used to identify the alleles of the *Vrn-1*, *Vrn-B3*, and *Ppd-D1* genes (Yan et al., 2004, 2006; Fu et al., 2005; Beales et al., 2007; Shcherban et al., 2012). PCR amplification was performed using a T100 Thermal Cycler (Bio-Rad, USA) and BioMaster HS-Taq PCR-Color reagents (Biolabmix, Russia) following the protocols described in the corresponding publications.

Genotyping, genetic map construction, and QTL mapping. High-throughput SNP genotyping was performed on DNA from the F_2 mapping population of common wheat (84 lines) derived from the cross between Obskaya 2 and Tulun 15, using the Illumina Infinium 20K Wheat chip (TraitGenetics GmbH, Germany). A total of 17,267 markers were analyzed.

Genetic map construction was performed using the MultiPoint UltraDense software (Mester et al., 2003). Markers with more than 25 errors or with a segregation distortion (χ^2) greater than 42 were removed. The minimum size for a group of co-segregating markers (linked markers mapped to the same position) was set at 2. Clustering was carried out with a recombination fraction threshold of 1.5. Marker ordering within clusters was conducted using the GES (guided evolutionary strategy) algorithm with jackknife resampling. To generate stable maps, monotony control was applied, involving the removal of outlier markers followed by sequential elimination of destabilizing markers.

Using the constructed genetic maps and phenotypic data, QTL mapping was conducted to identify loci controlling heading time, maturity time, and grain filling period in the population derived from the cross between Obskaya 2 and Tulun 15. QTL mapping was performed using the MultiQTL software based on the CIM (composite interval mapping) algorithm.

Gene prioritization from identified loci. For the functional characterization of candidate genes, sequences were annotated using the IWGSC RefSeq v1.0 reference genome. Gene expression patterns were assessed based on transcriptome sequencing data from the common wheat cultivar Azhurnaya, obtained from various plant tissues during development from germination to full maturity (Ramírez-González et al., 2018). As reference transcripts, sequences for *NAM-A1* (*TraesCS6A02G108300.1/.2*) and *NAM-D1* (*TraesCS6D02G096300.1*) were used, since *NAM-B1* in the Chinese Spring (CS) cultivar is represented by a non-functional allele and does not have an annotated ID in the RefSeq genome annotation.

Statistical analysis. Descriptive statistics, ANOVA, Tukey test, and Pearson correlation coefficients (r) were calculated using R software. All plots were generated with the R package ggplot2 (Wickham, 2016).

Results

Evaluation of developmental phase duration in the mapping population

Heading time in the population ranged from 34 to 41 days. Descriptive statistics for the three traits are presented in Table S1. The distribution of heading time was approximately normal (Fig. 1), while the distributions of maturity time and

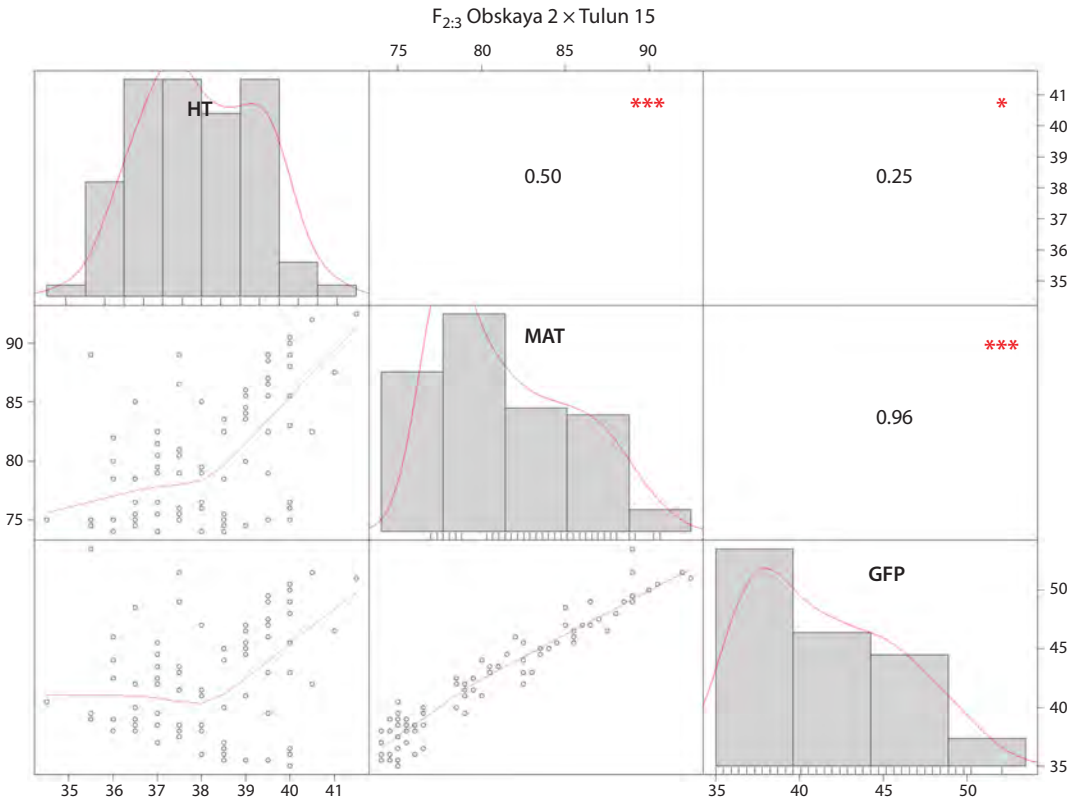


Fig. 1. Pearson correlation coefficients between traits related to developmental rate.

HT – heading time; MAT – maturity time; GFP – grain filling period. Significant differences are indicated by asterisks: * $p < 0.05$; *** $p < 0.001$.

grain filling period were skewed toward lower values. A very strong correlation was observed between maturity time and grain filling period, with an r^2 of 0.96.

Genetic maps and QTL mapping

For the construction of genetic maps for the Obskaya 2 × Tulun 15 population, 3,323 polymorphic markers were selected. Of these, 2,629 markers were mapped, including 534 skeleton markers. A total of 25 linkage groups were developed, with chromosomes 3B, 3D, and 5D represented by multiple groups. Summary data are presented in Table S2, and graphical representations of the maps are shown in Figure S2.

Using the CIM model, significant loci associated with the duration of developmental phases in common wheat were mapped to chromosomes 2D and 7B (Fig. 2). The locus on chromosome 2D was associated exclusively with heading time and accounted for 37 % of the phenotypic variation (PEV) for this trait. A locus on the short arm of chromosome 7B was also associated with heading time, explaining 20 % of the variation. Additionally, two loci on the long arm of chromosome 7B were associated with maturity time and grain filling period; these loci overlapped. The PEV values were 11.5 % for maturity time and 18 % for grain filling period.

The locus on chromosome 2D associated with heading time was located in the interval between markers BS00022276_51 (position 29454345 on RefSeq v1.0) and wsnp_CAP12_c1503_764765 (position 35683599 on RefSeq v1.0). Thus, the most likely candidate gene for this QTL is *Ppd-D1* (position 33952048–33956269), the physical location of which on the reference map precisely corresponds to the interval between the identified markers.

The locus on chromosome 7B associated with heading time was located in the interval between markers Tdurum_contig5352_556 (position 5061935 on RefSeq v1.0) and AX-95248379 (position 12717101 on RefSeq v1.0). Thus, the most likely candidate gene for this QTL is *Vrn-B3* (position 9702354–9704354), the physical location of which on the reference map falls within the interval defined by these markers.

For the loci associated with maturity time and grain filling period on the long arm of chromosome 7B, no known genes were identified. A search of the WheatQTLdb database also did not reveal any loci with similar positions.

Candidate genes within the maturity time and grain filling period locus

The *QMat.icg-7BL* locus was localized to the interval between 712618516 and 721195460 bp on RefSeq v1.0 and contains 141 genes (Table S3). Analysis of gene expression patterns in various tissues during plant development identified several candidate genes (Table S4).

A total of eight genes were identified that are predominantly expressed in the flag leaf, the fifth leaf after heading, or in the grain: *TraesCS7B02G455300*, *TraesCS7B02G459500*, *TraesCS7B02G459600*, *TraesCS7B02G460500*, *TraesCS7B02G460300*, *TraesCS7B02G454000*, *TraesCS7B02G461300*, and *TraesCS7B02G461400*.

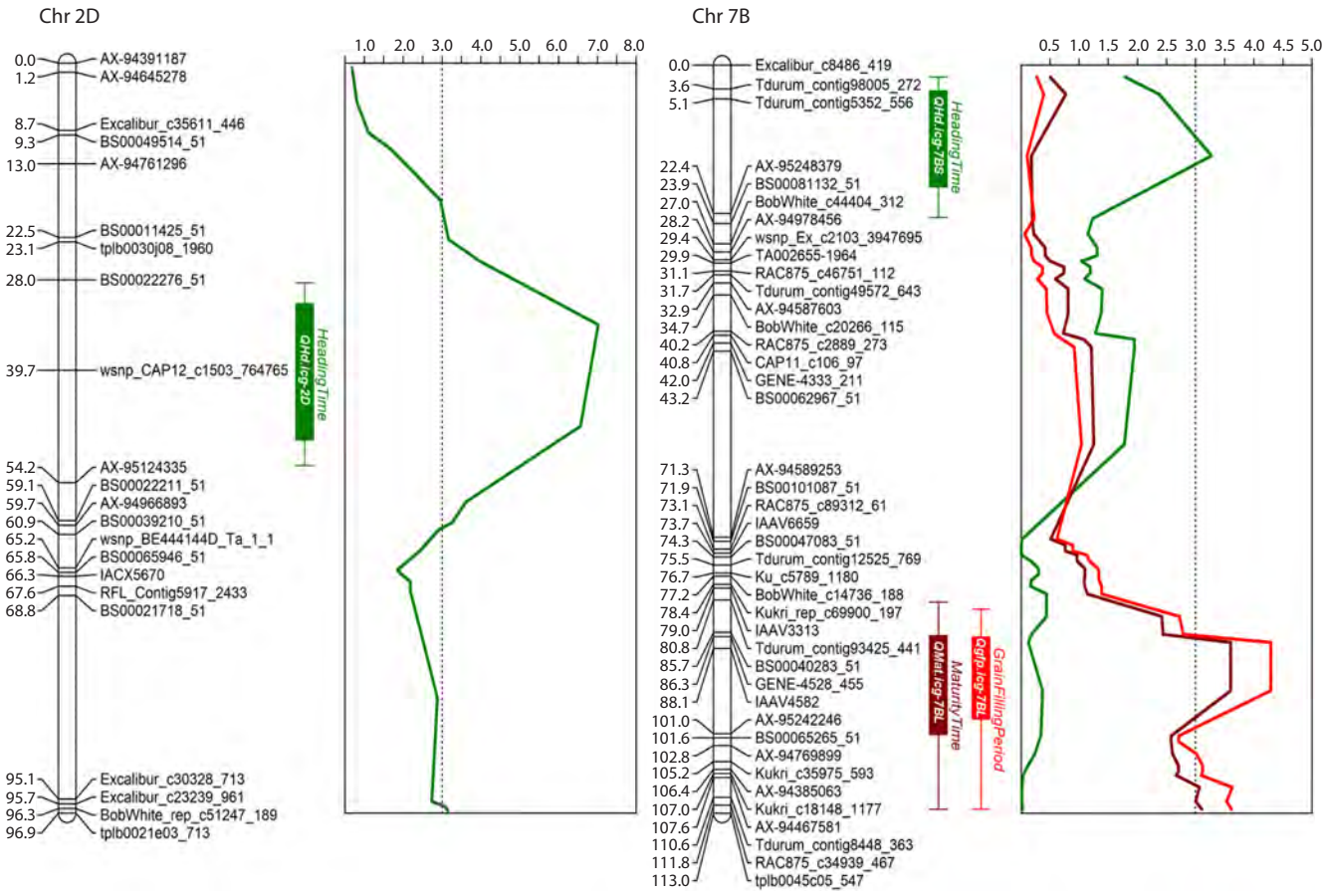


Fig. 2. Genetic maps of chromosomes 2D and 7B (showing only skeleton markers) with loci associated with heading time (green), maturity time (maroon), and grain filling period (red) indicated.

Selection of plants with different alleles of *Ppd-D1* and *Vrn-B3*

The *Ppd-D1* and *Vrn-B3* genes were identified through QTL analysis as the primary candidates controlling heading time in the studied population. Therefore, PCR was conducted to determine their allelic composition. Genotyping revealed that the early-maturing cultivar Tulun 15 carries the *Ppd-D1a* and *Vrn-B3a* alleles, whereas Obskaya 2 carries the *Ppd-D1b* and *vrn-B3* alleles. Both cultivars shared the same allelic composition for the *Vrn-1* genes: *Vrn-A1a*, *Vrn-B1c*, and *vrn-D1*.

Subsequently, the F_2 population plants were genotyped. As a result, 34 plants in which *Ppd-D1* and *Vrn-B3* were in a homozygous state were selected. Based on their allelic composition, the selected plants were divided into four groups (Fig. S1). The F_3 and F_4 progeny of these plants were sown in the field to determine the heading times for each group.

Evaluation of heading time in the F_3 and F_4 populations

Evaluation of the period from seedling emergence to heading in 2018 showed that plants from Group 1, carrying the dominant alleles *Ppd-D1a* and *Vrn-B3a*, headed the earliest – on average at 34.5 days (Fig. 3). Plants carrying the *Ppd-D1a* allele and the recessive *vrn-B3* allele (Group 2) headed 2.8 days later than Group 1, at 37.3 days. Plants from Group 3 (*Ppd-D1b* and *Vrn-B3a*) headed 38 days after emergence. The latest heading was observed in Group 4 plants (40.3 days), in which both genes were in the recessive state. Although the parental cultivar Obskaya 2 had the same allelic composition as Group 4, it headed even later, at 42.9 days. The Tulun 15 cultivar, despite carrying the same allelic combination as Group 1 (*Ppd-D1a* and *Vrn-B3a*), headed 3 days later than the Group 1 plants.

Evaluation of the duration from seedling emergence to heading in 2019 showed that the shortest period (37.1 days) was observed in the parental cultivar Tulun 15 (Fig. 3). Plants from Group 1, which had the same allelic composition for the *Vrn* and *Ppd* genes, headed 39 days after emergence, which was 5.8 days earlier than the plants from Group 4 (homozygous for the recessive *vrn-B3* and *Ppd-D1b* alleles). In Group 2, carrying the dominant *Ppd-D1a* allele and the recessive *vrn-B3* allele, the heading period was 2.5 days longer than in Group 1, totaling 41.5 days. Even later heading (42.8 days) was recorded in Group 3, where plants carried the recessive *Ppd-D1b* allele and the dominant *Vrn-B3a* allele. Among the hybrid population, plants in Group 4, with the recessive *Ppd-D1b* and *vrn-B3* alleles, headed the latest – 44.8 days after emergence. The parental cultivar Obskaya 2 exhibited the longest period from emergence to heading – 46.8 days.

ANOVA confirmed that the alleles of *Ppd-D1* and *Vrn-B3* significantly influenced heading time in both years of the study (Table S5). The presence of the *Ppd-D1a* allele accelerated

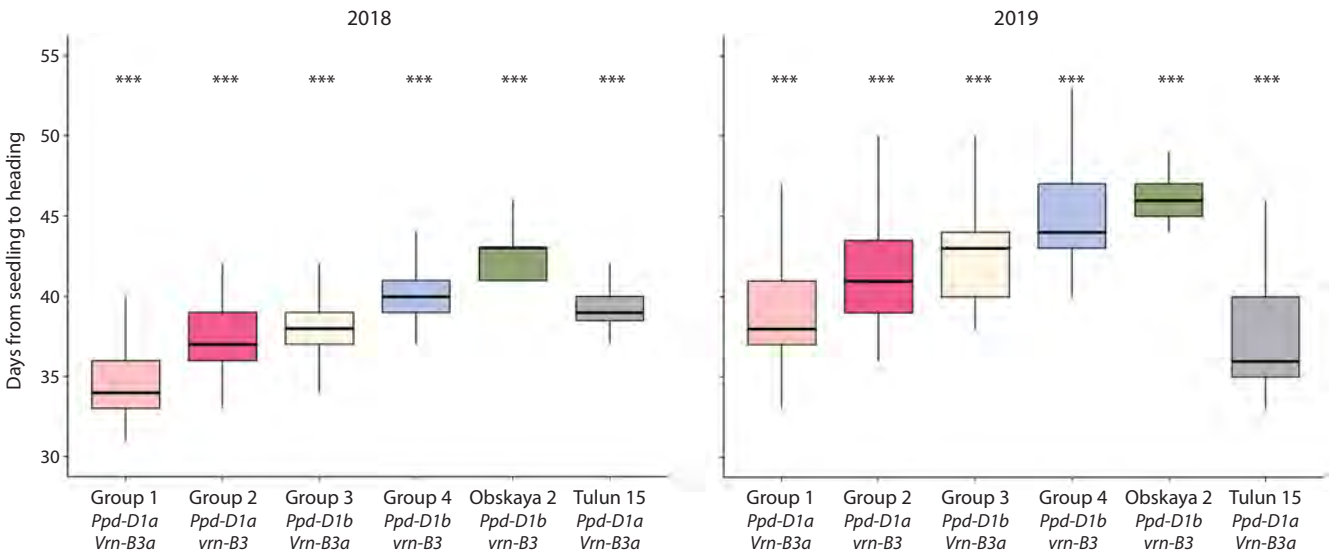


Fig. 3. Comparison of the duration from seedling to heading in F_3 and F_4 plants derived from the cross between Obskaya 2 and Tulun 15.
*** Significant differences between hybrids and parental cultivars at $p < 0.001$.

heading by 3.5 days in 2018 and by 4.4 days in 2019. The *Vrn-B3a* allele accelerated heading by 2.3 days in 2018 and by 2.5 days in 2019. Moreover, the combination of these two alleles resulted in heading 5.5 days earlier in 2018 and 6.5 days earlier in 2019 compared to the combination of the recessive alleles of these genes.

Effect of the *Ppd-D1* and *Vrn-B3* alleles on yield components

In both years, several yield parameters were evaluated, including the number of grains per main spike, grain weight per main spike, total number and weight of grains per plant, and 1000-grain weight (Table S6). According to ANOVA results, group classification based on the combination of *Ppd-D1* and *Vrn-B3* alleles had a significant effect on all traits, while the year of cultivation significantly affected all traits except for the number of grains per plant and grain weight per plant (Table S7).

The number of grains per main spike in the Obskaya 2 cultivar was the highest in 2018 – averaging 39.25 grains – and was significantly greater ($p < 0.001$) than in all four studied groups. The lowest number of grains per spike was observed in plants from Group 1 (25.65 grains), which differed significantly from Group 3 (29.08 grains) and Group 4 (28.56 grains). In Group 2 and in the Tulun 15 cultivar, the number of grains per spike was 26.83 and 31.60, respectively; however, no significant differences were detected between these and other groups. In 2019, significant differences were observed only between Group 1 (38.68 grains) and Group 3 (43.67 grains).

The grain weight per main spike in the Obskaya 2 cultivar was significantly higher than in all other plants in the experiment in both years: 1.99 g in 2018 and 2.21 g in 2019. Plants of the Tulun 15 cultivar had the lowest grain weight per spike in both 2018 (0.94 g) and 2019 (1.38 g). However, in 2018, there were no significant differences between Tulun 15 and the four groups, whereas in 2019, significant differences were observed between Tulun 15 and Group 3 (1.76 g). Comparing the groups, in 2018, Groups 1 (1.04 g) and 2 (1.02 g) had significantly lower grain weight per spike compared to Groups 3 (1.25 g) and 4 (1.20 g). In 2019, the grain weight per spike in Groups 1 (1.51 g) and 2 (1.40 g) was also significantly lower than in Group 3 (1.76 g).

The number of grains per plant in 2018 was the lowest in Group 1 (45.54 grains) and differed significantly compared to Group 2 (63.18 grains), Group 3 (66.52 grains), Group 4 (59.91 grains), and the Obskaya 2 cultivar (72.30 grains). No significant differences were observed between Group 1 and the Tulun 15 cultivar (54.60 grains). In 2019, there were no significant differences among all the studied plants for this trait.

The grain weight per plant in 2018 was highest in Obskaya 2 (3.56 g) and differed significantly from all other groups. Significant differences were also observed between Group 1 plants (1.78 g) and those in Group 3 (2.71 g) and Group 4 (2.44 g). In 2019, the Obskaya 2 cultivar again showed a significantly higher grain weight per plant (3.53 g) compared to all other plants in the experiment. The grain weight per plant in Groups 1, 2, 3, 4, and in Tulun 15 was 2.03, 2.13, 2.39, 2.28, and 2.34 g, respectively; however, no significant differences among these groups were detected in 2019.

The highest 1000-grain weight was observed in the Obskaya 2 cultivar in both years – 48.95 g in 2018 and 45.94 g in 2019 – and was significantly greater than in all other plants in the experiment. The lowest 1000-grain weight was recorded in the Tulun 15 cultivar: 29.58 g in 2018 and 34.15 g in 2019. In 2018, significant differences were observed between Tulun 15 and plants from Groups 1 (38.86 g), 2 (36.51 g), and 4 (38.62 g), while in 2019, significant differences were found between Tulun 15 and Group 3 (39.28 g). In 2018, no significant differences were found between Group 1 (38.86 g) and Groups 2 (36.51 g), 3 (40.96 g), and 4 (38.62 g), or between Groups 3 and 4. In 2019, no significant differences were detected between Group 1 (37.70 g) and Groups 3 (39.28 g)

and 4 (37.15 g). However, the 1000-grain weight of the Group 2 plants (32.93 g) was significantly lower than that of Groups 1, 3, and 4.

All these observations were supported by factorial analysis, where the allele combinations of *Ppd-D1* and *Vrn-B3* were used as factors (Table S7). ANOVA revealed that the allelic state of *Ppd-D1* was significantly associated with most of the evaluated traits in both years of the study, with a high level of significance (except for the number of grains per plant in 2019). In contrast, the allelic state of *Vrn-B3* demonstrated a lower level of significance for all traits compared to *Ppd-D1*. In 2019, *Vrn-B3* showed a significant association only with the 1000-grain weight. Overall, it can be concluded that the lines from Groups 3 and 4 possess a higher yield potential compared to other groups, although still lower than that of the original cultivar Obskaya 2.

Discussion

A new locus associated with maturity time

The correlation between maturity time and grain filling period was very high (0.96), indicating that in this population and under long-day conditions, grain filling period contributes the most to overall maturity time rather than heading time, despite the range for heading time being about seven days – a considerable variation. Previous studies have also reported that maturity time does not always depend on heading time and grain filling period and may be influenced by independent mechanisms (May, Van Sanford, 1992; Kajimura et al., 2011). However, genes specifically associated with maturity time in common wheat, apart from the *NAM-1* gene family, are largely unknown (Hagenblad et al., 2012).

In this study, we identified the *QMat.icg-7BL* locus on the long arm of chromosome 7B, associated with maturity time, localized within the interval 712618516–721195460 bp (RefSeq v1.0). Although a considerable number of studies have focused on identifying markers and loci associated with maturity time, describing loci on most wheat chromosomes except 3A and 6A, they have resulted in only a few associations for this trait being reported for chromosome 7B (Kulwal et al., 2003; McCartney et al., 2005; Huang et al., 2006; Kamran et al., 2013; Yu et al., 2015; Perez-Lara et al., 2016; Zou et al., 2017). We hypothesized that this locus might coincide with a previously described maturity time locus identified through GWAS in a population of domestic spring wheat cultivars (Kiseleva et al., 2023). However, the QTL mapped in the present study was located closer to the telomere and did not overlap with the previously identified locus. Similarly, comparison with another previously reported locus on chromosome 7B associated with maturity time (Kulwal et al., 2003) also revealed no overlap. Thus, we can conclude that we have identified a novel locus associated with maturity time.

Within the boundaries of this locus, 141 genes were identified. Based on the analysis of expression patterns in various tissues during plant development, several candidate genes associated with maturity time were identified. The *TraesCS7B02G455300* gene exhibited an expression pattern most similar to that of *NAM-A1* and *NAM-D1*, with a peak in the flag leaf at the full maturity stage. This gene encodes 12-oxophytodienoate reductase 1, a key enzyme involved in jasmonic acid biosynthesis. Previously, it was described as one of the candidate genes involved in the regulation of stem density (Taria et al., 2025). The *TraesCS7B02G459500*, *TraesCS7B02G459600*, *TraesCS7B02G460500*, and *TraesCS7B02G460300* genes also exhibited expression in the flag leaf after the heading stage and showed similar expression patterns in the fifth leaf. *TraesCS7B02G454000* showed increased expression during maturation in the fifth leaf and had detectable expression in the first leaf only at the tillering stage. *TraesCS7B02G461300* and *TraesCS7B02G461400* are annotated as Pseudo-Response Regulators, belonging to the same gene family as *Ppd-1*, one of the main genes regulating heading time. These genes exhibited relatively low expression levels but were specific to the grain at the milk and dough stages of development.

The *Ppd-D1a* and *Vrn-B3a* alleles significantly influence heading time under long-day conditions

With the same allelic combination of *Vrn-A1a*, *Vrn-B1c*, and *vrn-D1*, the presence of the dominant *Ppd-D1a* and *Vrn-B3a* alleles results in the earliest heading under long-day conditions. When the dominant *Ppd-D1a* allele is combined with the recessive *vrn-B3* allele, heading time is delayed by 2.5–3 days. In plants carrying the recessive *Ppd-D1b* allele and the dominant *Vrn-B3a* allele, the emergence-to-heading period is further extended by an additional 1–1.3 days.

The obtained results indicate that the *Ppd-D1a* allele exerts a stronger influence on the rate of transition to the generative phase of wheat development than the *Vrn-B3a* allele. This finding is consistent with the QTL analysis results, which showed that the locus on chromosome 2D explains a larger percentage of the variation in this trait.

Plants in which both genes are in the recessive form (Group 4) transition to heading significantly later than plants from the other groups. The Obskaya 2 cultivar, which also carries the recessive *Ppd-D1b* and *vrn-B3* alleles, heads even later – by an additional 2 to 2.5 days. Although the QTL analysis did not reveal other significant loci associated with heading time, this could be due to the presence of minor loci that were not detected with sufficient significance in the analysis, and that may have been inherited by Group 4 plants from the early-maturing cultivar Tulun 15. In addition to the genes studied here, other known genes such as *TaELF3*, *PhyC*, *PhyB*, *WPCL*, and numerous QTLs distributed across all chromosomes have been shown to influence the transition rate to the generative phase (Chen A. et al., 2014; Milec et al., 2014; Mizuno et al., 2016; Pearce et al., 2016; Wang et al., 2016; Zikhali et al., 2016).

When comparing heading time assessments across the two years (Fig. 3), it can be observed that in the second year of the study, heading occurred 4 to 5 days later in all hybrid groups and in the Obskaya 2 cultivar. The exception was the Tulun 15 cultivar, in which the duration period from emergence to heading remained unchanged at 37 days. This increase in heading time is likely due to differences in weather conditions between 2018 and 2019. According to data from gismeteo.ru, the average air temperature in June 2018 was 21.3 °C, whereas in June 2019 it was 18.5 °C – 2.8 °C lower. The stability of heading time in the Tulun 15 cultivar may indicate its high level of environmental plasticity.

We observed that the effects of combining the *Ppd-D1a* and *Vrn-B3a* alleles are additive (Table S5). According to the established model of floral meristem initiation leading to heading (Li C. et al., 2024), *Ppd-D1* acts as a primary inducer of the *Vrn-B3* gene. It is possible that *Ppd-D1a* does not directly influence *Vrn-B3a* because *Vrn-B3a* expression is already enhanced due to promoter region modifications (insertion events). Thus, it can be hypothesized that in the presence of the *Vrn-B3a* allele, *Ppd-D1a* affects heading time through the *Vrn-3* homeologs located on chromosomes 7A and 7D. Although few studies have specifically investigated the involvement of these genes in heading time regulation, (Bonnin et al., 2008) demonstrated that nucleotide polymorphisms in *Vrn-A3* and *Vrn-D3* were associated with variations in heading time.

Effect of early heading alleles on other agronomically important traits

The effects of *Ppd-1* genes on various agriculturally significant traits have been demonstrated in several previous studies. For example, (Boden et al., 2015) showed that these genes play a crucial role in inflorescence architecture and the development of paired spikelets in wheat. A more complex influence of *Ppd-1* on spike traits, including spike length, number of spikelets, and anther length, has also been reported (Okada et al., 2019), as well as effects on tiller number and spikelet number (Li W.L. et al., 2002). Several studies have further noted the influence of *Ppd-1* on the number of grains per main spike and 1000-grain weight (Wu et al., 2021). Our results showed that the *Ppd-D1a* allele had a significant negative effect on traits such as the number and weight of grains per spike and per plant, as well as on 1000-grain weight.

There is limited information available on the effects of *Vrn-3* genes on these traits. According to our data, the effect of the *Vrn-B3a* allele on the traits studied was less pronounced compared to *Ppd-D1a*, and its expression was more strongly influenced by growing conditions.

In most cases, when the differences were significant (p-value < 0.001), the dominant alleles of the studied genes were associated with lower trait values (fewer and lighter grains). Moreover, the trait values for the parental cultivar Tulun 15 were even lower, suggesting the involvement of additional genetic mechanisms regulating these traits independently of the duration of the vegetative phase.

Selection of lines from different maturity groups with high productivity traits

Among the F_4 generation plants, a search was conducted for highly productive lines across all four groups differing in heading time. Selection was based on traits such as grain weight per plant, which reflects yield potential, and 1000-grain weight, which reflects grain size. Additionally, visual assessment of the plants in the field was taken into account. It is well established that wheat yield has a strong correlation with the length of the vegetative period. In our experiment, variation was observed within each group for both grain weight per plant and 1000-grain weight. Although Group 1, characterized by the earliest heading, generally exhibited lower productivity traits, some early-heading lines from this group demonstrated grain weight per plant and 1000-grain weight values comparable to those of Group 4, which had the latest heading time, and substantially exceeded the early-maturing parental cultivar Tulun 15. Lines with good trait values were also identified in Group 3. In Group 2, several lines showed competitive grain weight per plant; however, the 1000-grain weight was the lowest among all groups. Based on this analysis, 19 lines from all four groups were selected as promising candidates for further breeding efforts (Table S8).

Conclusion

The results obtained allow us to conclude that the *Ppd-D1a* and *Vrn-B3a* alleles have a significant impact on the heading time of spring common wheat under long-day conditions. *Ppd-D1a* accelerates heading more strongly but also exerts a more pronounced negative effect on traits related to productivity. It is hypothesized that in the presence of the *Vrn-B3a* allele, *Ppd-D1a* may influence heading time through its homeologs *Vrn-A3* and *Vrn-D3*. The practical significance of this study lies in the development of new promising breeding lines of spring wheat with heading times optimized for many regions of Russia and other parts of the world characterized by long photoperiods.

References

Alhabbar Z., Islam S., Yang R., Diepeveen D., Anwar M., Balotf S., Sultana N., Maddern R., She M., Zhang J., Ma W., Juhasz A. Associations of *NAM-A1* alleles with the onset of senescence and nitrogen use efficiency under Western Australian conditions. *Euphytica*. 2018a;214(10):180. doi 10.1007/s10681-018-2266-4

Alhabbar Z., Yang R., Juhasz A., Xin H., She M., Anwar M., Sultana N., Diepeveen D., Ma W., Islam S. *NAM* gene allelic composition and its relation to grain-filling duration and nitrogen utilisation efficiency of Australian wheat. *PLoS One*. 2018b;13(10):e0205448. doi 10.1371/journal.pone.0205448

Beales J., Turner A., Griffiths S., Snape J.W., Laurie D.A. A *Pseudo-Response Regulator* is misexpressed in the photoperiod insensitive *Ppd-D1* a mutant of wheat (*Triticum aestivum* L.). *Theor Appl Genet.* 2007;115(5):721-733. doi 10.1007/s00122-007-0603-4

Belan I.A., Rosseeva L.P., Blokhina N.P., Grigoriev Y.P., Mukhina Y.V., Trubacheeva N.V., Pershina L.A. Resource potential of soft spring wheat varieties for the conditions of Western Siberia and Omsk region (analytical review). *Agrarnaya Nauka Evro-Severo-Vostoka = Agric Science Euro-North-East.* 2021;22(4):449-465. doi 10.30766/2072-9081.2021.22.4.449-465 (in Russian)

Berezhnaya A., Kiseleva A., Leonova I., Salina E. Allelic variation analysis at the vernalization response and photoperiod genes in Russian wheat varieties identified two novel alleles of *Vrn-B3*. *Biomolecules*. 2021;11(12):1897. doi 10.3390/biom11121897

Boden S.A., Cavanagh C., Cullis B.R., Ramm K., Greenwood J., Jean Finnegan E., Trevaskis B., Swain S.M. *Ppd-1* is a key regulator of inflorescence architecture and paired spikelet development in wheat. *Nat Plants*. 2015;1(2):14016. doi 10.1038/nplants.2014.16

Bonnin I., Rousset M., Madur D., Sourdille P., Dupuits C., Brunel D., Goldringer I. FT genome A and D polymorphisms are associated with the variation of earliness components in hexaploid wheat. *Theor Appl Genet.* 2008;116(3):383-394. doi 10.1007/s00122-007-0676-0

Chen A., Li C., Hu W., Lau M.Y., Lin H., Rockwell N.C., Martin S.S., Jernstedt J.A., Lagarias J.C., Dubcovsky J. Phytochrome C plays a major role in the acceleration of wheat flowering under long-day photoperiod. *Proc Natl Acad Sci USA.* 2014;111(28):10037-10044. doi 10.1073/pnas.1409795111

Chen F., Gao M., Zhang J., Zuo A., Shang X., Cui D. Molecular characterization of vernalization and response genes in bread wheat from the Yellow and Huai Valley of China. *BMC Plant Biol.* 2013;13:199. doi 10.1186/1471-2229-13-199

Chen S., Wang J., Deng G., Chen L., Cheng X., Xu H., Zhan K. Interactive effects of multiple vernalization (*Vrn-1*)- and photoperiod (*Ppd-1*)-related genes on the growth habit of bread wheat and their association with heading and flowering time. *BMC Plant Biol.* 2018;18(1):374. doi 10.1186/s12870-018-1587-8

Díaz A., Zikhali M., Turner A.S., Isaac P., Laurie D.A. Copy number variation affecting the *Photoperiod-B1* and *Vernalization-A1* genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS One*. 2012;7(3):e33234. doi 10.1371/journal.pone.0033234

Fu D., Szucs P., Yan L., Helguera M., Skinner J.S., von Zitzewitz J., Hayes P.M., Dubcovsky J. Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. *Mol Genet Genomics*. 2005;273(1):54-65. doi 10.1007/s00438-004-1095-4

Guo Z., Song Y., Zhou R., Ren Z., Jia J. Discovery, evaluation and distribution of haplotypes of the wheat *Ppd-D1* gene. *New Phytol.* 2010;185(3):841-851. doi 10.1111/j.1469-8137.2009.03099.x

Hagenblad J., Asplund L., Balfourier F., Ravel C., Leino M.W. Strong presence of the high grain protein content allele of *NAM-B1* in Fennoscandian wheat. *Theor Appl Genet.* 2012;125(8):1677-1686. doi 10.1007/s00122-012-1943-2

Huang X.Q., Cloutier S., Lycar L., Radovanovic N., Humphreys D.G., Noll J.S., Somers D.J., Brown P.D. Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). *Theor Appl Genet.* 2006;113(4):753-766. doi 10.1007/s00122-006-0346-7

Iqbal M., Shahzad A., Ahmed I. Allelic variation at the *Vrn-A1, Vrn-B1, Vrn-D1, Vrn-B3* and *Ppd-D1a* loci of Pakistani spring wheat cultivars. *Electron J Biotechnol.* 2011;14(1):1-8. doi 10.2225/vol14-issue1-fulltext-6

Kajimura T., Murai K., Takumi S. Distinct genetic regulation of flowering time and grain-filling period based on empirical study of D-genome diversity in synthetic hexaploid wheat lines. *Breed Sci.* 2011;61(2):130-141. doi 10.1270/jsbbs.61.130

Kamran A., Iqbal M., Navabi A., Randhawa H., Pozniak C., Spaner D. Earliness per se QTLs and their interaction with the photoperiod insensitive allele *Ppd-D1a* in the Cutler × AC Barrie spring wheat population. *Theor Appl Genet.* 2013;126(8):1965-1976. doi 10.1007/s00122-013-2110-0

Kiseleva A.A., Eggi E.E., Koshkin V.A., Sitnikov M.N., Roder M., Salina E.A., Potokina E.K. Detection of genetic determinants that define the difference in photoperiod sensitivity of *Triticum aestivum* L. near-isogenic lines. *Russ J Genet.* 2014;50(7):701-711. doi 10.1134/S102279541405007X

Kiseleva A.A., Leonova I.N., Ageeva E.V., Likhenko I.E., Salina E.A. Identification of genetic loci for early maturity in spring bread wheat using the association analysis and gene dissection. *PeerJ*. 2023;11:e16109. doi 10.7717/peerj.16109

Kiss T., Balla K., Veisz O., Láng L., Bedő Z., Griffiths S., Isaac P., Karsai I. Allele frequencies in the *VRN-A1, VRN-B1* and *VRN-D1* vernalization response and *PPD-B1* and *PPD-D1* photoperiod sensitivity genes, and their effects on heading in a diverse set of wheat cultivars (*Triticum aestivum* L.). *Mol Breed.* 2014;34(2):297-310. doi 10.1007/s11032-014-0034-2

Kulwal P.L., Roy J.K., Balyan H.S., Gupta P.K. QTL mapping for growth and leaf characters in bread wheat. *Plant Sci.* 2003;164(2):267-277. doi 10.1016/S0168-9452(02)00409-0

Li C., Lin H., Debernardi J.M., Zhang C., Dubcovsky J. *GIGANTEA* accelerates wheat heading time through gene interactions converging on *FLOWERING LOCUS T1*. *Plant J.* 2024;118(2):519-533. doi 10.1111/tpj.16622

Li W.L., Nelson J.C., Chu C.Y., Shi L.H., Huang S.H., Liu D.J. Chromosomal locations and genetic relationships of tiller and spike characters in wheat. *Euphytica*. 2002;125:357-366. doi 10.1023/A:1016069809977

Likhenko I.E., Stasyuk A.I., Shcherban A.B., Zyryanova A.F., Likhenko N.I., Salina E.A. Analysis of the allelic variation of the *Vrn-1* and *Ppd-1* genes in Siberian early and medium early varieties of spring wheat. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2014;18(4/1):691-703 (in Russian)

Lundström M., Leino M.W., Hagenblad J. Evolutionary history of the *NAM-B1* gene in wild and domesticated tetraploid wheat. *BMC Genet.* 2017;18(1):118. doi 10.1186/s12863-017-0566-7

Lysenko N.S., Kiseleva A.A., Mitrofanova O.P., Potokina E.K. VIR World Collection Catalogue. Iss. 815. Bread Wheat: Molecular Testing of the Vrn and Ppd Alleles in the Selection Varieties Approved for Use in the Russian Federation. St. Petersburg: VIR Publ., 2014 (in Russian)

May L., Van Sanford D.A. Selection for early heading and correlated response in maturity of soft red winter wheat. *Crop Sci.* 1992;32(1):47-51. doi 10.2135/cropsci1992.0011183X003200010011x

McCartney C.A., Somers D.J., Humphreys D.G., Lukow O., Ames N., Noll J., Cloutier S., McCallum B.D. Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452 × 'AC Domain'. *Genome*. 2005;48(5):870-883. doi 10.1139/G05-055

Mester D., Ronin Y., Hu Y., Peng J., Nevo E., Korol A. Efficient multipoint mapping: making use of dominant repulsion-phase markers. *Theor Appl Genet.* 2003;107(6):1102-1112. doi 10.1007/s00122-003-1305-1

Milec Z., Valárik M., Bartoš J., Safář J. Can a late bloomer become an early bird? Tools for flowering time adjustment. *Biotechnol Adv.* 2014;32(1):200-214. doi 10.1016/j.biotechadv.2013.09.008

Mizuno N., Kinoshita M., Kinoshita S., Nishida H., Fujita M., Kato K., Murai K., Nasuda S. Loss-of-function mutations in three homoeologous *PHYTOCLOCK 1* genes in common wheat are associated with the extra-early flowering phenotype. *PLoS One*. 2016;11(10):e0165618. doi 10.1371/journal.pone.0165618

Mizuno N., Matsunaka H., Yanaka M., Nakata M., Nakamura K., Nakamaru A., Kiribuchi-Otobe C., Ishikawa G., Chono M., Hatta K., Fujita M., Kobayashi F. Allelic variations of *Vrn-1* and *Ppd-1* genes in Japanese wheat varieties reveal the genotype-environment interaction for heading time. *Breed Sci.* 2022;72(5):343-354. doi 10.1270/jsbbs.22017

Nishida H., Yoshida T., Kawakami K., Fujita M., Long B., Akashi Y., Laurie D.A., Kato K. Structural variation in the 5′ upstream region of photoperiod-insensitive alleles *Ppd-A1a* and *Ppd-B1a* identified in hexaploid wheat (*Triticum aestivum* L.), and their effect on heading time. *Mol Breed.* 2013;31(1):27-37. doi 10.1007/s11032-012-9765-0

Okada T., Jayasinghe J.E.A.R.M., Eckermann P., Watson-Haigh N.S., Warner P., Hendrikse Y., Baes M., … Albertsen M., Wolters P., Fleury D., Baumann U., Whitford R. Effects of *Rht-B1* and *Ppd-D1* loci on pollinator traits in wheat. *Theor Appl Genet.* 2019;132(7):1965-1979. doi 10.1007/s00122-019-03329-w

Palomino C., Cabrera A. Evaluation of the allelic variations in vernalisation (*VRN1*) and photoperiod (*PPD1*) genes and genetic diversity in a Spanish spelt wheat collection. *Int J Mol Sci.* 2023;24(22):16041. doi 10.3390/ijms242216041

Pearce S., Kippes N., Chen A., Debernardi J.M., Dubcovsky J. RNA-seq studies using wheat *PHYTOCHROME B* and *PHYTOCHROME C* mutants reveal shared and specific functions in the regulation of flowering and shade-avoidance pathways. *BMC Plant Biol.* 2016;16(1):141. doi 10.1186/s12870-016-0831-3

Perez-Lara E., Semagn K., Chen H., Iqbal M., N'Diaye A., Kamran A., Navabi A., Pozniak C., Spaner D. QTLs associated with agronomic traits in the Cutler × AC Barrie spring wheat mapping population using single nucleotide polymorphic markers. *PLoS One*. 2016;11(8):e0160623. doi 10.1371/journal.pone.0160623

Plaschke J., Ganal M.W., Röder M.S. Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor Appl Genet.* 1995;91(6-7):1001-1007. doi 10.1007/BF00223912

Ramírez-González R.H., Borrill P., Lang D., Harrington S.A., Brinton J., Venturini L., Davey M., … Sharpe A.G., Paux E., Spannagl M., Bräutigam A., Uauy C. The transcriptional landscape of polyploid wheat. *Science.* 2018;361(6403):eaar6089. doi 10.1126/science.aar6089

Seki M., Chono M., Matsunaka H., Fujita M., Oda S., Kubo K., Kiribuchi-Otobe C., Kojima H., Nishida H., Kato K. Distribution of photoperiod-insensitive alleles *Ppd-B1a* and *Ppd-D1a* and their effect on heading time in Japanese wheat cultivars. *Breed Sci.* 2011;61(4):405-412. doi 10.1270/jsbbs.61.405

Shcherban A.B., Efremova T.T., Salina E.A. Identification of a new *Vrn-B1* allele using two near-isogenic wheat lines with difference in heading time. *Mol Breed.* 2012;29(3):675-685. doi 10.1007/s11032-011-9581-y

Taria S., Arora A., Krishna H., Manjunath K.K., Kumar Sudhir, Singh B., Meena S., Malakondaiah A.C., Kousalya S., Padaria J.C., Singh P.K., Alam B., Kumar Sushil, Arunachalam A. Mapping of the QTLs governing stem-specific weight for stem reserve mobilisation in wheat (*Triticum aestivum* L.) under combined heat and drought stress. *Plant Physiol Rep*. 2025. doi 10.1007/s40502-025-00854-3

Wang J., Wen W., Hanif M., Xia X., Wang H., Liu S., Liu J., Yang L., Cao S., He Z. *TaELF3-1DL,* a homolog of *ELF3,* is associated with heading date in bread wheat. *Mol Breed.* 2016;36(12):161. doi 10.1007/s11032-016-0585-5

Whittal A., Kaviani M., Graf R., Humphreys G., Navabi A. Allelic variation of vernalization and photoperiod response genes in a diverse set of North American high latitude winter wheat genotypes. *PLoS One*. 2018;13(8):e0203068. doi 10.1371/journal.pone.0203068

Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer, 2016. doi 10.1007/978-3-319-24277-4

Wilhelm E.P., Turner A.S., Laurie D.A. Photoperiod insensitive *Ppd-A1a* mutations in tetraploid wheat (*Triticum durum* Desf.). *Theor Appl Genet.* 2009;118(2):285-294. doi 10.1007/s00122-008-0898-9

Worland A.J.J., Börner A., Korzun V., Li W.M.M., Petrovíc S., Sayers E.J.J. The influence of photoperiod genes on the adaptability of European winter wheats. *Euphytica*. 1998;100(1/3):385-394. doi 10.1023/A:1018327700985

Wu Y., Liu J., Hu G., Xue H., Xu H., Zhao C., Qin R., Cui F., Sun H. Functional analysis of the "Green Revolution" gene *Photoperiod-1* and its selection trends during bread wheat breeding. *Front Plant Sci.* 2021;12:745411. doi 10.3389/fpls.2021.745411

Yan L., Helguera M., Kato K., Fukuyama S., Sherman J., Dubcovsky J. Allelic variation at the *VRN-1* promoter region in polyploid wheat. *Theor Appl Genet.* 2004;109(8):1677-1686. doi 10.1007/s00122-004-1796-4

Yan L., Fu D., Li C., Blechl A., Tranquilli G., Bonafede M., Sanchez A., Valarik M., Yasuda S., Dubcovsky J. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. *Proc Natl Acad Sci USA.* 2006;103(51):19581-19586. doi 10.1073/pnas.0607142103

Yu M., Chen G.-Y., Pu Z.-E., Zhang L.-Q., Liu D.-C., Lan X.-J., Wei Y.-M., Zheng Y.-L. Quantitative trait locus mapping for growth duration and its timing components in wheat. *Mol Breed.* 2015;35(1):44. doi 10.1007/s11032-015-0201-0

Zaitseva O.I., Lemesh V.A. Allelic composition in the *Vrn-A1, Vrn-B1,* and *Vrn-B3* genes of double haploid lines of hexaploid triticale. *Russ J Genet.* 2015;51(7):653-660. doi 10.1134/S1022795415070145

Zhang X.K., Xiao Y.G., Zhang Y., Xia X.C., Dubcovsky J., He Z.H. Allelic variation at the vernalization genes *Vrn-A1, Vrn-B1, Vrn-D1,* and *Vrn-B3* in Chinese wheat cultivars and their association with growth habit. *Crop Sci.* 2008;48(2):458-470. doi 10.2135/cropsci2007.06.0355

Zhang Y., Liu W.C., Li J., Wei H.T., Hu X.R., Li Y.J., Lu B.R., Yang W.Y. Distribution and selective effects of *Vrn-A1, Vrn-B1,* and *Vrn-D1* genes in derivative varieties from four cornerstone breeding parents of wheat in China. *Agric Sci China*. 2010;9(10):1389-1399. doi 10.1016/S1671-2927(09)60230-3

Zikhali M., Wingen L.U., Griffiths S. Delimitation of the *Earliness per se D1* (*Eps-D1*) flowering gene to a subtelomeric chromosomal deletion in bread wheat (*Triticum aestivum*). *J Exp Bot.* 2016;67(1):287-299. doi 10.1093/jxb/erv458

Zou J., Semagn K., Iqbal M., N'Diaye A., Chen H., Asif M., Navabi A., Perez-Lara E., Pozniak C., Yang R.-C., Randhawa H., Spaner D. Mapping QTLs controlling agronomic traits in the 'Attila' × 'CDC Go' spring wheat population under organic management using 90K SNP array. *Crop Sci.* 2017;57(1):365-377. doi 10.2135/cropsci2016.06.0459

Conflict of interest. The authors declare no conflict of interest.
Received March 13, 2025. Revised May 15, 2025. Accepted May 16, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):779-788

doi 10.18699/vjgb-25-85

Prebreeding studies of near-isogenic spring bread wheat lines, differing by presence or absence of the 3R(3D) chromosomal substitution from the triticale cultivar Satu

S.N. Sibikeev [1], I.G. Adonina [2], A.E. Druzhin [1], Z.E. Fitileva [1], O.A. Baranova [3]

[1] Federal Center of Agriculture Research of the South-East Region, Saratov, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] All-Russian Research Institute of Plant Protection, Pushkin, St. Petersburg, Russia
sibikeev_sergey@mail.ru

Abstract. One of the sources of resistance to leaf and stem rust pathogens for bread wheat is the Australian spring triticale cultivar Satu, which carries highly effective linked *SrSatu/LrSatu* genes localized on chromosome 3R. However, they are little used in the practical breeding of *Triticum aestivum* L. The main reason for that is a low level of knowledge regarding the 3R(3D) chromosomal substitution. This paper presents the results of a comparative study of the agronomic value of near-isogenic spring bread wheat siblings, L16 and L17 = Satu/Saratovskaya 70//Saratovskaya 74/3/Saratovskaya 74, differing by presence (L16 (3R(3D))) or absence (L17 (3D3D)) of chromosome 3R from Satu in 2023–2024. The 3R(3D) chromosomal substitution in L16 was detected by cytogenetic analysis combining GISH with labeled *Secale cereale* genomic DNA and FISH with probes pSc119.2, pAs1. Line L16 is highly resistant to *Puccinia triticina* and *P. graminis*, including the Ug99 race. PCR analysis with DNA markers of *Sr* genes revealed the non-identity of the resistance gene in L16 to *Sr* genes: *Sr2*, *Sr24*, *Sr25*, *Sr28*, *Sr31*, *Sr32*, *Sr36*, *Sr38*, *Sr39*, *Sr47* and *Sr57*. L16 was inferior to both L17 and the standard cultivar Saratovskaya 76 in terms of 1,000-grain weight. An analysis of productivity elements of the main ear revealed that the 3R(3D) substitution in L16 significantly reduced the length of the ear, increased the density of the ear and did not significantly affect the number of spikelets and the number of grains per ear and the grain weight per ear. The grain protein content in L16 did not significantly differ from its L17 siblings or Saratovskaya 76. Similarly, there were no significant differences in gluten content. However, gluten in L16 was weaker in comparison with line L17 and Saratovskaya 76. According to the complex trait of SDS sedimentation, L16 was inferior to L17, but did not significantly differ from the standard cultivar. According to the alveograph, L16 had significantly lower dough elasticity and flour strength, but in comparison with the standard cultivar, the decrease in flour strength was not significant. L16 showed a higher bread volume than Saratovskaya 76, but did not significantly differ from its L17 sibling. There was no difference in porosity for all three samples. In general, in terms of the complex of agronomically valuable traits, the spring bread wheat line L16 (3R(3D)) requires further work to improve its breeding value.

Key words: triticale Satu; near isogenic lines of bread wheat; 3R(3D) substitution; resistance to leaf and stem rust; influence for productivity and grain quality

For citation: Sibikeev S.N., Adonina I.G., Druzhin A.E., Fitileva Z.E., Baranova O.A. Prebreeding studies of near-isogenic spring bread wheat lines, differing by presence or absence of the 3R(3D) chromosomal substitution from the triticale cultivar Satu. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):779-788. doi 10.18699/vjgb-25-85

Funding. Cytogenetic analysis was carried out with the support of the Ministry of Science and Higher Education project FWNR-2022-0017.

Пребридинговые исследования почти изогенных линий яровой мягкой пшеницы, отличающихся по наличию/отсутствию хромосомного замещения 3R(3D) от сорта тритикале Satu

С.Н. Сибикеев [1], И.Г. Адонина [2], А.Е. Дружин [1], З.Е. Фитилева [1], О.А. Баранова [3]

[1] Федеральный аграрный научный центр Юго-Востока, Саратов, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Всероссийский научно-исследовательский институт защиты растений, Пушкин, Санкт-Петербург, Россия
sibikeev_sergey@mail.ru

Аннотация. Одним из источников устойчивости к возбудителям листовой и стеблевой ржавчин для мягкой пшеницы является австралийский сорт ярового тритикале Satu, несущий высокоэффективные сцепленные

гены *SrSatu*/*LrSatu*, локализованные в хромосоме 3R. Однако они мало используются в практической селекции *Triticum aestivum* L. из-за недостаточной изученности влияния этих генов на продуктивность и качество зерна. В настоящей работе представлены результаты сравнительного исследования агрономической ценности почти изогенных линий-сибсов яровой мягкой пшеницы Л16 и Л17, полученных с участием сорта Satu и различающихся наличием (Л16) или отсутствием (Л17 (3D3D)) замещения 3R/3D. Хромосомное замещение 3R(3D) у Л16 выявлено при цитогенетическом анализе, сочетающем GISH с меченой геномной ДНК *Secale cereale* и FISH с зондами pSc119.2, pAs1. Линия Л16 высокоустойчива к *Puccinia triticina* и *P. graminis*, включая линейку расы Ug99. ПЦР-анализом с ДНК-маркерами генов *Sr* установлена неидентичность гена устойчивости у Л16 генам *Sr*: *Sr2*, *Sr24*, *Sr25*, *Sr28*, *Sr31*, *Sr32*, *Sr36*, *Sr38*, *Sr39*, *Sr47* и *Sr57*. Урожайность зерна у Л16 в оба года исследований была ниже, чем у Л17 и сорта-стандарта Саратовская 76. По массе 1000 зерен Л16 уступила как Л17, так и Саратовской 76. Анализ элементов продуктивности главного колоса показал, что замещение 3R(3D) у Л16 значимо уменьшило длину колоса, повысив его плотность, и практически не повлияло на количество колосков и зерен в колосе и массу зерна с колоса. По содержанию белка в зерне линия Л16 значимо не отличалась ни от своего сибса Л17, ни от сорта Саратовская 76. Схожим образом не обнаружились значимые различия по содержанию клейковины. Однако клейковина у Л16 была более слабой по сравнению как с Л17, так и с сортом Саратовская 76. По комплексному показателю SDS-седиментации Л16 уступила Л17, но незначимо различалась с сортом-стандартом. По показателям альвеографа у Л16 более низкие упругость теста и сила муки, но по сравнению с сортом-стандартом понижение силы муки незначимое. По объему хлеба Л16 с 3D(3R) имела большее значение, чем Саратовская 76, но незначимо отличалась от своего сибса Л17 с 3D3D. По пористости все три образца не отличались друг от друга. В целом по комплексу хозяйственно ценных признаков линия яровой мягкой пшеницы Л16 (3R(3D)) требует дальнейшей работы по улучшению ее селекционной ценности.
Ключевые слова: тритикале Satu; почти изогенные линии мягкой пшеницы; замещение 3R(3D); устойчивость к листовой и стеблевой ржавчинам; влияние на продуктивность и качество зерна

Introduction

Bread wheat (*Triticum aestivum* L.) is one of the main food crops. Of the total world grain production, wheat accounts for over 27 %. Bread wheat has the largest areas of crops and is the main food product for a third of the world's population. A significant erosion of the gene pool of this crop for genes of resistance to disease and pests occurred due to intensive breeding to increase productivity that took place in the 20th century (Dymchenko et al., 1990).

Pathogens have greater genetic variability in their biological characteristics, and two or three epiphytoties and a crop area of 50–100 thousand hectares are sufficient to overcome the resistance gene of the host plant. At present, the main diseases in the bread wheat growing zones continue to be stem, leaf and stripe rust, powdery mildew, leaf and ear septoria, and various types of viral infection. The susceptibility of bread wheat cultivars to diseases leads to huge losses in grain yield and a decrease in bread-making quality traits (Sibikeev, Krupnov, 2007).

Wild relatives of bread wheat have many genes of agronomic interest and can be valuable sources of resistance to diseases, insects, and extreme environmental factors (Sibikeev et al., 2019). To protect bread wheat from pathogens and, first of all, from rust diseases, the resistance genes localized in alien chromosomes and translocations are widely used. Thus, out of 82 identified *Lr* genes, 39 were transferred from "wild" relatives, and out of 63 *Sr* genes, 26 were introgressed (McIntosh et al., 2013, 2018, 2022). The following genes were transferred and identified from *Secale cereale* L.: *Sr27*, *Sr31*, *Sr50*, *Sr1RS^{Amigo}* (McIntosh et al., 2013), and from triticale: *SrSatu*, *SrBj*, *SrNin*, *SrLa1*, *SrLa2* and *SrVen* (McIntosh et al., 1995; Adhikari, 1996).

One of the donors of leaf and stem rust pathogens resistance for bread wheat is the Australian spring triticale cultivar Satu, which carries highly effective linked genes *SrSatu/LrSatu*, localized in chromosome 3R. Moreover, *SrSatu* is presumably allelic to the *Sr27* gene, which is found in many triticale varieties (McIntosh et al., 1995). The *SrSatu* gene is highly effective against the races of *Puccinia graminis* f. sp. *tritici* Erikss. & Henning from the USA, Kenya and South Africa, including the lineage of the Ug99 race, namely TTKSK, TTKST, TTTSK, TRTTF, TTTTF, RKQQC, QTHJC, TPMKS, TKTTF, MCCFC (Rahmatov et al., 2016). In the Russian Federation, the *SrSatu* gene is effective against the *P. graminis* populations of the Middle and Lower Volga regions and the Northwestern region (Baranova et al., 2023, 2024). However, *SrSatu* is extremely rarely used in bread wheat breeding. The main reason for this is insufficient knowledge of the effect the 3R chromosome with the *SrSatu/LrSatu* genes has on cytological stability, the main traits of grain productivity, and bread-making quality traits.

For use of alien substitutions and translocations in wheat breeding, prebreeding studies are necessary to determine their cytological stability in the genome, the effect on the adaptive properties of plants, as well as on elements of productivity, grain yield and the quality of the final product (Sibikeev, Druzhin, 2015).

The purpose of our research was to determine, based on the results of studying the near-isogenic siblings of spring bread wheat L16 (3R(3D)) and L17 (3D3D), their cytological stability and potential for practical breeding both in terms of effectiveness against rust diseases and in terms of their impact on grain productivity and bread-making quality traits.

Materials and methods

The material used included near-isogenic siblings of spring bread wheat L16 and L17, which were obtained by crossing the Australian cultivar of spring hexaploid triticale Satu, resistant

to leaf and stem rust pathogens, and the cultivars of spring bread wheat Saratovskaya 70 and Saratovskaya 74, susceptible to the indicated types of rust. Pedigree: Satu/Saratovskaya 70 //Saratovskaya 74/3/Saratovskaya 74. Sibling lines L16 and L17 were obtained by the method of forced heterozygotes. They were maintained in the heterozygous state until the seventh generation; then the homozygosity of the sib plants was confirmed over three generations. Both lines belong to the albidum subspecies and differ from each other for their resistance to leaf and stem rust pathogens (L16 is resistant and L17 is susceptible to rust). Thus, the marker trait of the *SrSatu/LrSatu* genes from triticale in L16 was resistance to leaf and stem rust pathogens. To characterize the grain productivity and the bread-making quality traits, the L16 and L17 lines were compared with each other, as well as with the spring bread wheat cultivar Saratovskaya 76 – this standard was adopted by the State Commission of the Russian Federation for Breeding Achievements Test and Protection in the Saratov region.

Cytogenetic studies. Karyotype analysis of the lines was performed using fluorescent *in situ* hybridization (FISH) probes based on repeating sequences of pSc119.2 (Bedbrook et al., 1980) and pAs1 (Rayburn, Gill, 1987) on mitotic metaphase chromosomes. Mitotic chromosome preparations were prepared from the root meristem of seedlings according to the method (Badaeva et al., 2017). FISH was performed using the method described in (Salina et al., 2006) with minor modifications. Genomic *in situ* hybridization (GISH) was performed using labeled *S. cereale* genomic DNA as a probe in combination with a 10–30-fold excess of unlabeled fragmented *T. aestivum* DNA according to a previously published work (Schubert et al., 1998). The preparations were analyzed using an Axio Imager M1 microscope (Zeiss, Germany) equipped with a ProgRes MF CCD digital camera (Jenoptik, Germany) using the Isis image analysis program (Meta Systems, Germany). The work was carried out at the Center for Collective Use for Microscopic Analysis of Biological Objects of the Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia).

Cytological stability was evaluated by studying the behavior of chromosomes in microsporogenesis in meiosis. Microsporogenesis was studied on temporary squashed preparations. The ears of the L16 and L17 lines were cut before leaving the leaf sheath and fixed in a mixture of 96 % ethanol and glacial acetic acid (3:1). A day after fixation, the material was transferred to 70 % ethyl alcohol, where it was stored until analysis at +2–4 °C. Schiff reagent was used for staining. For each line, 100–200 microsporocytes of meiotic stages (metaphase I and II, anaphase I and II, telophase I and II, tetrads) were analyzed. Slides were analyzed on an Axio Scope A 1 microscope (Carl Zeiss) with N-ACHROPLAN 40×/0.65 and N-ACHROPLAN 100×/1.25 0;1 objectives.

Phytopathological studies. Since L17 was selected as a sibling of L16 susceptible to *P. triticina* and *P. graminis*, only L16 was used in phytopathological studies. To evaluate the resistance of the L16 line to the stem rust pathogen in laboratory conditions of the All-Russian Institute of Plant Protection, populations collected in 2022 in the Arsk district of the Republic of Tatarstan (from the Nadira cultivar) and in the Samoylovka district of the Saratov region (from the Voevoda cultivar) were used. Virulence analysis of *P. graminis* f. sp. *tritici* was carried out using a standard set of 20 differentiator lines (*Sr5*, *Sr21*, *Sr9e*, *Sr7b*, *Sr11*, *Sr6*, *Sr8a*, *Sr9g*, *Sr36*, *Sr9b*, *Sr30*, *Sr17*, *Sr9a*, *Sr9d*, *Sr10*, *SrTmp*, *Sr24*, *Sr31*, *Sr38*, *SrMcN*), as well as additional lines with *Sr* genes (*Sr2compl*, *Sr8b*, *Sr12*, *Sr13*, *Sr15*, *Sr20*, *Sr22*, *Sr25*, *Sr26*, *Sr27*, *Sr28*, *Sr29*, *Sr32*, *Sr33*, *Sr35*, *Sr37*, *Sr39*, *Sr40*, *Sr44*, *SrWLD*, *Sr24+31*, *Sr36+31*, *Sr24+36*, *Sr7a+12*, *Sr17+13*, *Sr7b+18*, *Sr26+9g* and *Sr33+5*), cultivars Avrora (*Sr31*) and Khakasskaya (susceptible control). The virulence analysis of pathogen populations from the Nadira and Voevoda cultivars was described by us earlier (Baranova et al., 2023).

The propagation of stem rust pathogen populations and the plants analysis for resistance at the seedling stage were carried out using methods accepted in world practice (Jin et al., 2007). The seedlings reaction to inoculation with a suspension of stem rust pathogen spores was checked on the 12th day using the standard 4-point scale of E.C. Stakman et al. (1962). The resistance/susceptibility of the sample was determined based on the infection types in two replicates. Plants with infection types “0”, “0;”, “1”, “2” were considered as resistant; plants with infection types “3”, “4”, “X” were considered as susceptible.

Resistance to race Ug99 (TTKSK) was tested at the adult plant stage in 2023 in the plant pathology nurseries of the International Maize and Wheat Improvement Center (CIMMYT) in Kenya, in the Kenya Agricultural and Livestock Research Organization (KALRO) in Njoro. A modified Cobb scale (Peterson et al., 1948) was used to evaluate plant response. The main distinguishing trait of race Ug99 pathotypes is virulence towards *Sr31* genetic carriers. The severity of cultivars with the *Sr31* gene being affected in KALRO phytopathological plant nurseries during the vegetation period of 2023 was: for the Prokhorovka cultivar (*Sr31*) – 60 % (60MSS), for the Yugo-Vostochnaya 2 cultivar (*Sr31*) – 80 % (80S), for the Saratovskaya 74 cultivar (without identified *Sr* genes) – 80 % (80S), for the Saratovskaya 70 cultivar (without identified *Sr* genes) – 40 % (40MSS).

Molecular genetic analysis. For PCR analysis, DNA was isolated from five-day-old wheat seedlings using the cetyltrimethylammonium bromide (CTAB) method (Murray, Thompson, 1980). To identify resistance genes (*Sr2*, *Sr24*, *Sr25*, *Sr28*, *Sr31*, *Sr32*, *Sr36*, *Sr38*, *Sr39*, *Sr47* and *Sr57*), DNA markers recommended for marker-assisted breeding (MAS) were used: *Sr2* – CAPS marker csSr2 (Mago et al., 2011); *Sr24/Lr24* – STS markers Sr24#12 and Sr24#50 (Mago et al., 2005); *Sr25/Lr19* – STS marker Gb (Prins et al., 2001); *Sr26* – STS marker Sr26#43 (Mago et al., 2005); *Sr28* – DaRT marker wPt-7004-PCR and SSR marker *Xwmc332* (Rouse et al., 2012); *Sr31/Lr26* – STS marker of SCM9 (Weng et al., 2007); *Sr32* – STS marker csSr32#2 (Mago et al., 2013); *Sr36* – SSR marker of *Xstm773-2* (Tsilo et al., 2008); *Sr38/Lr37* – STS marker of VENTRIUP-LN2 (Helguera et al., 2003); *Sr39/Lr35* – STS marker Sr39#22 (Mago et al., 2009); *Sr47* – *Xgwm501*, *Xgpw4043* (Faris et al., 2008; Klindworth et al, 2012); *Sr57/Lr34* – STS marker of csLV34 (Lagudah et al., 2006).

Amplification was performed on C1000 Thermal Cycler (BioRad) amplifiers, amplification products were separated in 2 % agarose and 8 % polyacrylamide gels stained with ethidium bromide. Isogenic lines and cultivars with known *Sr* genes were positive controls, the susceptible cultivar Khakasskaya was the negative control, and the PCR mixture without added DNA served as contamination controls. GeneRulerTM 50 bp DNA Ladder (Fementas) was used as a molecular weight marker. Amplification products were visualized using the ChemiDoc XRS+ gel documentation system (Bio-Rad). PCR was performed in two replicates.

The evaluation of grain productivity traits, physical and bread making quality traits in the L16, L17 lines and the standard cultivar Saratovskaya 76 was carried out in 2023 and 2024. The experimental material was randomly sown in plots of 7 m^2 in three replicates in the experimental field of the Federal Center of Agriculture Research of the South-East Region in Saratov. The seeding rate was 400 grains per 1 m^2.

In addition to phenological observations and direct evaluation of grain yield, studies of the elements productivity of the main spike – spike length, number of spikelets and grains, grain weight per spike, spike density, grains for one spikelet (as a general fertility trait) – of the studied lines and the standard cultivar were carried out. These analyses were conducted for 15 spikes of L16, L17 and Saratovskaya 76. Bread-making quality traits were evaluated by the content of crude gluten, the strength of which was determined using the Gluten Deformation Meter IDK-3M (OOO PLAUN), as well as by the Chopin alveograph (Chopin Technologies) traits with the baking of experimental samples of bread. The protein content in grain from the 2023 and 2024 harvests was determined using the Foss Infratec TM 1241 grain analyzer (Foss Analytical A/S).

The meteorological traits of the years of cultivation according to Selyaninov's hydrothermal coefficient (www.agrometeo.online/articles/gtk.htm, accessed 28.01.2025) by months of the growing season showed the following. In 2023, the hydrothermal coefficient for May was 0.8, in June, 1.1, in July, 0.6 and in August, 0.4. In 2024, the hydrothermal coefficient for April was 0.3, for May, 0.1, for June, 0.8 and for July, 0.1. Of the two years of research, 2023 was the more favorable in terms of meteorological conditions. In both years of research, leaf rust epiphytoties were observed.

The data obtained for the L16 and L17 lines and the standard cultivar Saratovskaya 76 were subjected to a one-way variance analysis with multiple comparisons according to Duncan; an analysis of the genotype–environment interaction was also carried out using the Agros-2.09 breeding and genetic software package (Martynov, 1999).

Results

Cytogenetic analysis of spring bread wheat lines L16 and L17

Line karyotyping was performed by FISH using a combination of pSc119.2 and pAs1 probes for chromosome identification (Schneider et al., 2003). Cytogenetic analysis of the L16 line revealed the absence of a pair of 3D chromosomes (they are determined by characteristic pAs1 signals) (Fig. 1*a*) and showed a pair of large chromosomes with bright pSc119.2 signals at the ends of the arms, which is typical for rye 3R chromosomes. Chromosomal substitution 3R(3D) in the L16 line was also confirmed by GISH with *S. cereale* DNA as a probe (Fig. 1*a*). Analysis of the L17 line did not reveal any chromosomal rearrangements (Fig. 1*b*).

Cytological stability of the L16 and L17 lines

The main factor limiting the practical application of distant hybrids is their instability, leading to a rapid loss of alien genetic material. This instability is based on disturbances in the meiotic cycle of hybrid plants, causing the formation of non-functional gametes (Orlovskaya et al., 2015). It is known that the genome of distant hybrids and amphidiploids contains systems of genetic meiosis control of different parental species, which not only act independently in the hybrid genome, but also mutually influence each other (Naranjo et al., 1979; Lelley, Larter, 1980; Orlovskaya et al., 2015). In this regard, it is necessary to determine the cytological stability of both distant hybrids and introgressive lines. Since the L16 line is characterized by the 3R(3D) substitution, and chromosome 3D is the carrier of the *Ph2* gene (McIntosh et al., 2013), there is reason to expect disturbances in the meiotic cycle. In addition, we conducted a comparative analysis of the meiotic stability in the L16 and L17 lines.

For evaluation of meiosis stability, an integral trait is used – the meiotic index, which is the percentage of normal tetrads to the total number of cells studied. Studies have shown that L16 (3R(3D)) has a meiotic index of 95 %, while L17 (3D3D) has a meiotic index of 94 %, i. e. there are no differences between the lines in this trait. With a meiotic index of 90 % or higher, the plant is cytologically stable, i. e. both lines studied are stable. However, some tetrads of these lines have 1–2 inclusions and triads. In earlier stages of meiosis, asynchrony was found in both lines: earlier divergence of one bivalent in metaphase I, two lagging chromosomes in anaphase I and two chromosomes not included in the nuclei in telophase I. Among the mononuclear pollen, micropollen in the L16 and L17 lines was 1.5 and 2.1 %, respectively.

Phytopathological analysis of line L16 for resistance to stem and leaf rust pathogens. Identification of stem rust resistance genes using molecular markers

During the production of near isogenic sibs, the L17 line was selected as stably susceptible to both stem and leaf rust pathogens at all stages of plant growth (infection type IT = 33^+) under greenhouse and field conditions. In this regard, the evaluation of resistance to *P. graminis* f. *tritici* and *P. triticina* was carried out on L16 with the substitution of 3R(3D) and the standard cultivar Saratovskaya 76 (Table 1).

It should be noted that the analysis of the virulence of the *P. graminis* f. *tritici* population from the Favorit cultivar showed that the following genes and their combinations are effective: *Sr2compl*, *Sr13*, *Sr22*, *Sr26*, *Sr27*, *Sr31*, *Sr32*, *Sr33*, *Sr35*, *Sr39*, *Sr24+Sr31*, *Sr36+Sr31*, *Sr26+Sr9g*, *Sr17+Sr13*, *Sr33+Sr5*. The pathogen population from the Favorit cultivar on the line with *Sr27* (the gene transferred from *S. cereale*

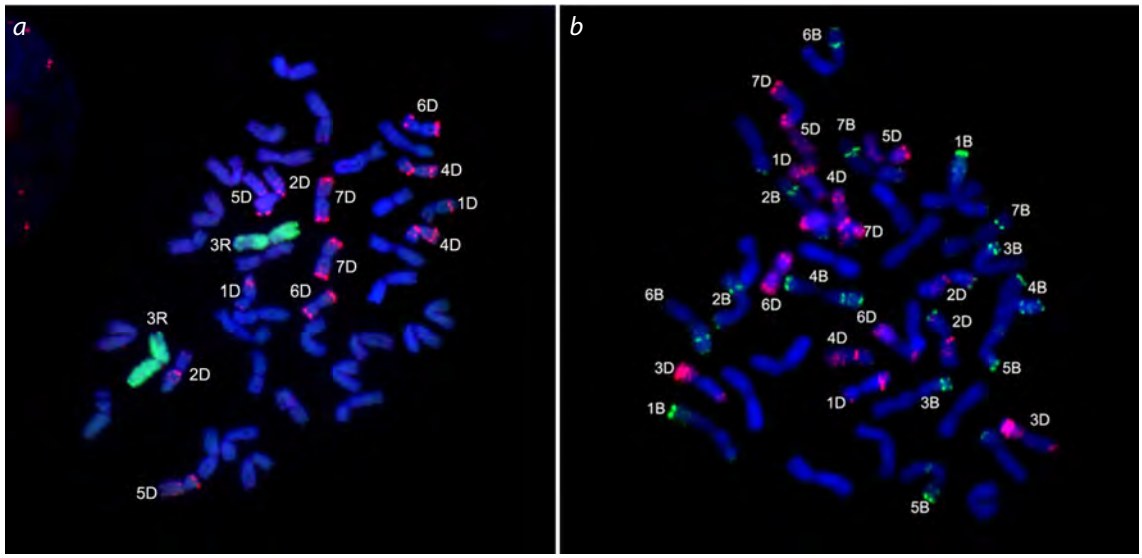


Fig. 1. Results of *in situ* hybridization with different combinations of probes on metaphase chromosomes of spring bread wheat lines L16 (*a*) and L17 (*b*).

Probes: (*a*, *b*) pAs1 (red signal); *a* – GISH with *S. cereale* genomic DNA (green signal), *b* – pSc119.2 (green signal).

and localized in the 3R chromosome, widely present in triticale cultivars (McIntosh et al., 1995)) gave reaction type "2" (IT = 2), and from the populations from the Nadira and Voevoda cultivars – "2$^+$" (IT = 2$^+$) and "1" (IT = 1), respectively.

As can be seen from Table 1, L16 showed high resistance to all populations of *P. graminis* f. *tritici*, including the Ug99 race. Under natural epiphytoties conditions of *P. triticina* in 2023 and 2024, L16 showed resistance to the local Saratov population. At the same time, the standard cultivar Saratovskaya 76 was susceptible to both populations of *P. graminis* f. *tritici* (the exception was the population from the cultivar Nadira IT = 0;) and to *P. triticina*.

High resistance to both rust pathogens, to all populations of pathogens from different parts of Russia and Kenya makes disease resistance in L16 attractive for breeding work. It should be noted that L16 in the Njoro KALRO nursery, Kenya, also showed resistance to the local population of *P. striiformis* f. sp. *tritici* West. – 5R, and the Saratovskaya 76 cultivar to – 5M. The results of *Sr* genes identification in the analyzed L16 line using molecular markers for the *Sr2*, *Sr24*, *Sr25*, *Sr28*, *Sr31*, *Sr32*, *Sr36*, *Sr38*, *Sr39*, *Sr47* and *Sr57* genes showed their absence.

Phenology, grain productivity and bread making quality traits in lines L16, L17 and standard cultivar Saratovskaya 76

Figure 2 shows the traits of the germination-earing stage duration, plant height, lodging resistance in the vegetation seasons of 2023 and 2024 for L16, L17 and the standard cultivar. For the vegetation seasons of 2023 and 2024, the germination–earing stage duration in the sibling lines L16 and L17 and the standard cultivar Saratovskaya 76 was almost the same and the differences were insignificant. Thus, the substitution of

Table 1. Characteristics of spring bread wheat line L16 and the standard cultivar Saratovskaya 76 for resistance to *P. graminis* f. *tritici* and *P. triticina* in field conditions (natural infection background) and laboratory conditions (artificial infection)

Cultivar, line	Infection type (*P. graminis* f. *tritici*)				Infection type (*P. triticina*)	
	Populations from cultivars*			Ug99 (TTKSK)**	Saratov local population, field conditions***	
	Nadira	Voevoda	Favorit		2023	2024
Saratovskaya 76	0;	3, 4	3	30MS	3	3
L16	1	0;	0;	0	0;	0;

* Populations of *P. graminis* f. *tritici* collected from spring bread wheat cultivars Nadira (Arsk district of the Republic of Tatarstan), Voevoda (Samoilovka district of the Saratov region) and Favorit (Arkadak district of the Saratov region). Laboratory evaluation at the 3-leaf stage.
** Evaluation of the Ug99 race pathotypes was carried out at Njoro KALRO phythopalogical nursery, Kenya.
*** Evaluation was carried out during natural epiphytotics of *P. triticina* in the experimental field of the Federal Center of Agriculture Research of the South-East Region.

the 3D chromosome of bread wheat by the 3R chromosome from the spring triticale cultivar Satu has almost no effect on the duration of the germination–earing period. In 2023, plant height of the L16 line, a sib with the 3R(3D) substitution, was significantly lower than that of its pair with normal 3D3D chromosome composition and the standard cultivar Saratovskaya 76, and in 2024, it was significantly lower than in the L17 line, but was at the level of the standard cultivar.

Thus, the substitution of the 3D chromosome by the 3R chromosome led to a decrease in plant height, which affected the evaluation of lodging resistance in the 2023 growing season, which was significantly higher in the sib with the 3R(3D) substitution compared to the sib pair and the standard cultivar. In 2024, the lodging resistance of L17 was significantly lower than that of Saratovskaya 76, and L16 did not differ from either L17 or the standard cultivar. However, L16 with the 3R(3D) chromosomal substitution exceeded the value of L17 in absolute lodging resistance. Analysis of the genotype–environment interactions between the L16, L17 lines and the Saratovskaya 76 cultivar for the germination–earing stage duration, plant height and lodging resistance showed that these interactions are not significant. Comparisons of L16 and L17 for grain yield and elements of main spike productivity made it possible to identify the effect of alien substitution with the 3R chromosome on these traits (Table 2). It was found out that the 3R(3D) substitution in both years of research reduced grain yield both in comparison with L17 (3D3D) and with the standard cultivar Saratovskaya 76. One of the factors that reduced grain productivity was the lower 1,000-grain weight of L16 in 2023 and 2024 (Table 2).

The analysis of the main spike productivity elements revealed that the substitution 3R(3D) reduces the spike length, increasing its density. However, L16 did not differ from L17 and the Saratovskaya 76 cultivar in the number of spikelets, grains, and grain weight per spike. In terms of grain content per one spikelet (a generalized fertility criterion), L16 did not differ from either L17 or the Saratovskaya 76 cultivar. The analysis of the genotype-environment interaction between the L16, L17 lines and the Saratovskaya 76 cultivar for all traits of spike productivity structure, as well as 1,000-grain weight and grain yield, showed that this interaction was not significant.

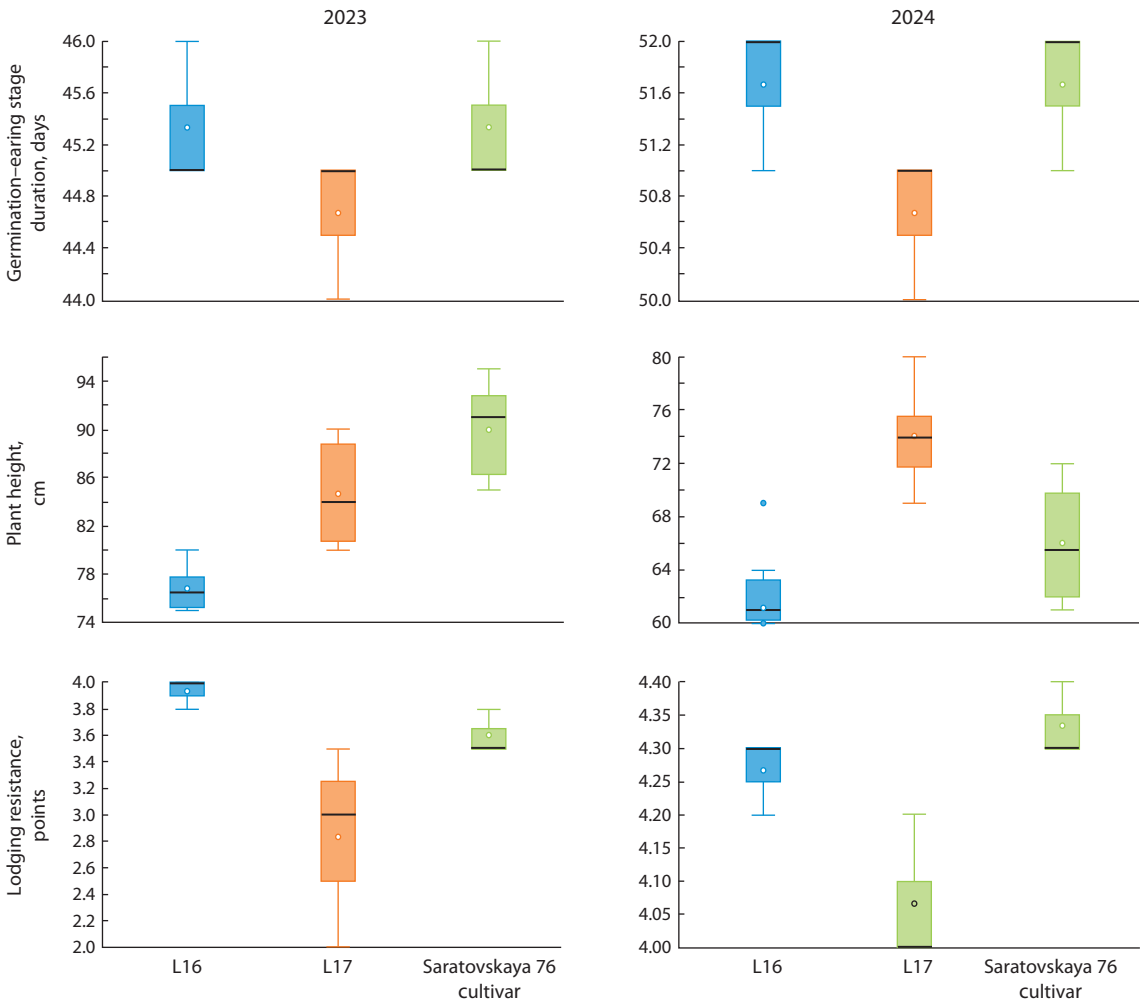


Fig. 2. Traits of the germination–earing stage duration, plant height, lodging resistance in the growing seasons of 2023 and 2024 for L16, L17 and the standard cultivar Saratovskaya 76.

Table 2. The grain yield, 1,000-grain weight and elements of the main ear productivity of spring bread wheat lines L16, L17 and the Saratovskaya 76 cultivar for the growing seasons of 2023 and 2024

Cultivar, line	Elements of the main ear productivity, average for 2023–2024						1,000-grain weight			Grain yield, t/ha		
	Ear length, mm	Number, pcs.		Ear density	Grain weight per ear, g	Grains for one spikelet, pcs.	2023	2024	Average	2023	2024	Average
		spikelets	grains per ear									
L16	67.8	17.3	42.4	2.4	1.3	2.5	32.7	29.6	31.2	2.078	1.698	1.888
L17	92.8	16.5	39.4	1.7	1.5	2.4	40.0	38.0	39.0	2.569	2.565	2.567
$LSD_{0.5}^{*}$	22.9	NS	NS	0.6	NS	NS	3.7	3.3	7.0	0.206	0.501	NS
S76	98.6	17.3	42.4	1.7	1.6	2.5	38.9	31.6	35.3	2.776	2.042	2.409
$LSD_{0.5}^{**}$	23.5	NS***	NS	0.6	NS	NS	2.8	NS	3.5	0.622	0.300	NS

* The least significant difference for the 5 % significance level between L16 and L17.
** The least significant difference for the 5 % significance level between L16 and Saratovskaya 76.
*** NS – no difference.

Table 3. Bread making quality traits in spring bread wheat lines L16, L17 and the standard cultivar Saratovskaya 76 on average for the growing seasons of 2023 and 2024

Cultivar, line	Protein content in grain, %	Gluten		SDS, ml	Alveograph*			Bread**		
		%	IDK 3, units of device		P, mm	P/L	W, units of alveograph	V, ml	porosity, point	crumb color
L16	17.1	36.7	88	67	60	0.7	205	820	4.8	White
L17	17.5	37.4	78	86	91	1.0	345	890	5.0	White
$LSD_{0.5}^{**}$	NS*****	NS	8	13	20	NS	70	NS	NS	
S76	16.2	31.5	79	75	109	2.0	261	715	4.8	White
$LSD_{0.5}^{****}$	NS	NS	8	NS	25	NS	NS	100	NS	

* Alveograph indices: P – dough elasticity, P/L – ratio of dough elasticity to extensibility, W – flour strength; ** Bread evaluation indices: V – bread volume;
*** The least significant difference for 5 % significance level between L16 and L17; **** The least significant difference for 5 % significance level between L16 and Saratovskaya 76; ***** NS – no difference.

The analysis of bread-making quality traits showed the following results. Grain protein content in L16 (3R(3D)) did not differ significantly from either its sib L17 (3D3D) or the Saratovskaya 76 cultivar. Similarly, no significant differences were found in gluten content. However, according to the IDK-3M device, L16 gluten was weaker compared to both L17 and Saratovskaya 76 (Table 3).

According to the SDS sedimentation indices, which characterize physical dough properties, L16 was inferior to L17, but did not differ significantly from the standard cultivar Saratovskaya 76. For the alveograph indices, L16 had significantly lower dough elasticity and flour strength, but compared to the standard cultivar, the decrease in flour strength was insignificant (Table 3). L16 had a higher bread volume than Saratovskaya 76, but did not differ significantly from its sibling L17. In terms of porosity, all three samples did not differ significantly from each other, but the highest score (5.0) was found in the L17 line. The substitution 3R(3D) did not change the color of the bread crumb, L16 (3R(3D)) had a white crumb, like L17 (3D3D) and Saratovskaya 76.

Discussion

As noted above, triticale cultivars are attractive among breeders because of a set of valuable agronomic traits for breeding bread wheat, including disease resistance genes. Thus, the following genes of resistance to the stem rust pathogen have been identified in triticale: *Sr27*, *SrSatu*, *SrBj*, *SrNin*, *SrLa1*, *SrLa2* and *SrVen* (McIntosh et al., 1995; Adhikari, McIntosh, 1998). Moreover, the *SrBj* and *SrVen* genes control moderate susceptibility at the seedling stage in field conditions, and the *SrLa1* and *SrLa2* genes determine resistance at the seedling stage both in field and in greenhouse conditions (Adhikari, McIntosh, 1998). The *Sr27* and *SrSatu* genes control resistance throughout the growing season of plants (Singh, McIntosh, 1988). Despite the *Sr27* and *SrSatu* genes being considered as allelic (Singh, McIntosh, 1988), they differ in

their effectiveness against the stem rust pathogen. Currently, in the Saratov population, there is an increase in the content of virulent pathotypes of *P. graminis* f. *tritici* to the *Sr27* gene (in 2016 – 10 %, in 2019 – 20 %, in 2020 – 90 %), while pathotypes virulent to *SrSatu* are not detected (Kon'kova, 2021).

The results of our studies of the *SrSatu* gene in L16 confirm the conclusions about its effectiveness. Moreover, its efficiency against the population of *P. graminis* f. *tritici* from the Nadira cultivar collected in Tatarstan was revealed, IT = 1, while in the line with *Sr27*, IT = 2^+. These results coincide with the data of testing other lines of spring bread wheat L968 = Satu/S70// S74/3/S70/4/S70 and L935 = Satu/S70//S70 with 3R(3D) from crossings of the triticale cultivar Satu with spring bread wheat cultivars bred by the Federal Center of Agriculture Research of the South-East Region. These lines showed a reaction type of "0" to the population from the Nadira cultivar (Baranova, unpublished data) and a rating of "–0" to the Ug99 race of *P. graminis* f. *tritici* (Baranova et al., 2024). In our studies, the L16 line was also resistant to the Saratov populations of *P. triticina* both during the selection of sibs with 3R(3D) and 3D3D chromosomal composition and in the field experiments in 2023 and 2024.

Thus, the *SrSatu/LrSatu* genes in chromosome 3R of the L16 line are highly effective against populations from the Lower and Middle Volga regions of *P. graminis* f. *tritici* and *P. triticina*, as well as against the Ug99 pathotypes of *P. graminis* f. *tritici*. Consequently, when chromosome 3R was transferred from the triticale cultivar Satu to the genotypes of spring bread wheat line L16, the expression of the *SrSatu/ LrSatu* resistance genes was not disrupted. Our attempts to identify *Sr* genes in the analyzed L16 line using molecular markers for the *Sr2*, *Sr24*, *Sr25*, *Sr28*, *Sr31*, *Sr32*, *Sr36*, *Sr38*, *Sr39*, *Sr47* and *Sr57* genes showed their absence. Thus, it was shown that L16 carries its own (*SrSatu*) unidentified resistance gene. Unfortunately, to date, there is no DNA marker for the *Sr27* and *SrSatu* genes (McIntosh et al., 2013).

In our studies of cytological stability in lines L16 and L17, a number of disturbances during the passage of meiosis phases were revealed. However, according to meiotic indices in L16 (3R(3D)) and L17 (3D3D), 95 and 94 %, respectively, these lines are characterized as stable. There are grounds to assume that the absence of the *Ph2* gene in L16 (nullisomal state of the 3D chromosome) was compensated by the presence of the 3R chromosome of *S. cereale*, and in terms of meiosis stability, L16 did not differ from L17.

Unfortunately, in the literature available to us, we did not find a data about the effect of the 3R chromosome from the triticale cultivar Satu in the genotypes of spring bread wheat on grain productivity and the bread-making quality traits. However, the effect of the 3R chromosome from the line 86-741 (F_6 hexaploid triticale Guangmai 74 (AABBRR)/Fan 6 (bread wheat)) was studied. The authors studied 185 F_8 recombinant inbred lines from crossing bread wheat cultivar Chuanmai 42 with line 86-741. Chromosome 3R was identified by FISH and GISH methods (Wan et al., 2023). The authors found out that the 3R(3D) substitution significantly reduces grain yield, 1,000-grain weight, number of ears per plant, grain weight per ear and has a neutral effect on the number of grains per ear (Wan et al., 2023).

Our studies also noted a decrease in 1,000-grain weight and grain yield, but a neutral effect on grain weight per ear, number of grains per ear, grain content per one spikelet, and a significant increase in ear density. Thus, there is some discrepancy in the effect of 3R(3D) on grain weight per ear. This may be due to differences in the genotype of bread wheat in which the 3R(3D) substitution was studied. In addition, it is necessary to take into account the effect of nullisomy for 3D. Normally, 3D is a carrier of the dominant spherococcoid gene *S-D1a* (McIntosh et al., 2013), respectively, the null state of this chromosome determines the recessive state of this gene – *S-D1b*. It is known that the *S-D1b* gene has a pleiotropic effect, which reduces plant height, ear length, 1,000-grain weight, and increases ear density (Sears, 1947; according to: Salina et al., 2000). All these traits were detected in the L16 line (3R(3D)). Based on this, there is reason to expect that the morpho-biological traits in L16 are formed under the combined influence of the recessive state of the *S-D1b* gene and the direct action of the 3R chromosome.

The absence of 3D and the presence of the 3R chromosome in L16 affected the bread-making quality traits. Basically, these traits worsened in comparison with sibling line L17. The evaluations of SDS sedimentation, dough elasticity and flour strength decreased significantly. An insignificant decrease was noted for other traits: protein content in grain, gluten content and strength (according to the IDK-3M device), elasticity to dough length ratio (P/L), bread volume and porosity.

Conclusion

The spring bread wheat line L16 carries highly effective genes for resistance to leaf and stem rust pathogens, which are attractive for breeding for immunity to common wheat pathogens in the Russian Federation. However, in general, the spring bread wheat line L16 (3R(3D)) requires further work to improve its set of economically valuable traits. This is possible by reducing the amount of alien gene material, i. e. by obtaining recombinations or translocations between bread wheat chromosomes and the 3R chromosome, as well as by selecting a bread wheat genotype that will compensate for the negative impact of the rye chromosome on grain productivity and the bread-making quality traits.

References

Adhikari K.N. Genetic studies of stem rust resistance in oat and triticale: Phd thesis. The University of Sydney, 1996

Adhikari K.N., McIntosh R.A. Inheritance of wheat stem rust resistance in triticale. *Plant Breed.* 1998:117(6):505-513. doi 10.1111/j.1439-0523.1998.tb02199.x

Badaeva E.D., Ruban A.S., Aliyeva-Schnorr L., Municio C., Hesse S., Houben A. *In situ* hybridization to plant chromosomes. In: Liehr T. (Ed.) Fluorescence *In Situ* Hybridization (FISH). Springer Protocols Handbooks. Springer, 2017;477-494. doi 10.1007/978-3-662-52959-1_49

Baranova O., Solyanikova V., Kyrova E., Kon'kova E., Gaponov S., Sergeev V., Shevchenko S., … Tarhov A., Vasilova N., Askhadullin D., Askhadullin D., Sibikeev S. Evaluation of resistance to stem rust and identification of *Sr* genes in Russian spring and winter

wheat cultivars in the Volga region. *Agriculture*. 2023;13(3):635. doi 10.3390/agriculture13030635

Baranova O.A., Adonina I.G., Sibikeev S.N. Molecular cytogenetic characteristics of new spring bread wheat introgressive lines resistant to stem rust. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2024;28(4):377-386. doi 10.18699/vjgb-24-43

Bedbrook R.J., Jones J., O'Dell M., Thompson R.J., Flavell R.B. A molecular description of telomeric heterochromatin in secale species. *Cell*. 1980;19(2):545-560. doi 10.1016/0092-8674(80)90529-2

Dymchenko A.M., Nazarova L.N., Zhemchuzhina A.I. Promising and released varieties of winter wheat differing in resistance to brown rust. *Selektsiya i Semenovodstvo = Breeding and Seed Industry*. 1990;5:16-18 (in Russian)

Faris J.D., Xu S.S., Cai X., Friesen T.L., Jin Y. Molecular and cytogenetic characterization of a durum wheat – *Aegilops speltoides* chromosome translocation conferring resistance to stem rust. *Chromosome Res*. 2008;16(8):1097-1105. doi 10.1007/s10577-008-1261-3

Helguera M., Khan I.A., Kolmer J., Lijavetzky D., Zhong-qi L., Dubcovsky J. PCR assays for the *Lr37-Yr17-Sr38* cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. *Crop Sci*. 2003;43(5):1839-1847. doi 10.2135/cropsci2003.1839

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

Klindworth D.L., Niu Z., Chao S., Friesen T.L., Jin Y., Faris J.D., Cai X., Xu S.S. Introgression and characterization of a goatgrass gene for a high level of resistance to Ug99 stem rust in tetraploid wheat. *G3* (*Bethesda*). 2012;2(6):665-673. doi 10.1534/g3.112.002386

Kon'kova E.A. Characteristics of the virulence of the wheat stem rust pathogen in the conditions of the Saratov region. *Agrarnyy Nauchnyy Zhurnal = The Agrarian Scientific Journal*. 2021;8:23-27. doi 10.28983/asj.y2021i8pp23-27 (in Russian)

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

Lelley T., Larter E.N. Meiotic regulation in triticale: interaction of the rye genotype and specific wheat chromosomes on meiotic pairing in the hybrid. *Can J Jenet Cytol*. 1980;22:1-6

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

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

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

Mago R., Verlin D., Zhang P., Bansal U., Bariana H., Jin Y., Ellis J., Hoxha S., Dundas I. Development of wheat – *Aegilops speltoides* recombinants and simple PCR-based markers for *Sr32* and a new stem rust resistance gene on the 2S#1 chromosome. *Theor Appl Genet*. 2013;126(12):2943-2955. doi 10.1007/s00122-013-2184-8

Martynov S.P. Statistical and Biometric Genetic Analysis in Crop Production and Breeding. AGROS Software Package, version 2.09. Tver, 1999 (in Russian)

McIntosh R.A., Wellings C.R., Park R.F. (Eds) Wheat Rusts. An Atlas of Resistance Genes. CSIRO Australia, 1995. Available: https://bgri.cornell.edu/wp-content/uploads/2021/01/wheat_rust_atlas_full.pdf

McIntosh R.A., Yamazaki Y., Dubcovsky J., Rogers W.J., Morris C., Appels R., Xia X.C. Catalogue of Gene Symbols for Wheat. 12th International Wheat Genetics Symposium. 8-13 September 2013. Yokohama, Japan, 2013. Available: https://shigen.nig.ac.jp/wheat/komugi/genes/macgene/2013/GeneCatalogueIntroduction.pdf

McIntosh R.A., Dubcovsky J., Rogers W.J., Xia X.C., Raupp W.J. Catalogue of Gene Symbols for Wheat: 2018 Supplement. *Annu Wheat Newsl*. 2018;64:73-93

McIntosh R.A., Dubcovsky J., Rogers W.J., Xia X.C., Raupp W.J. Catalogue of Gene Symbols for Wheat: 2022 Supplement. *Annu Wheat Newsl*. 2022;68:68-81

Murray M.G., Thompson W.F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res*. 1980;8(19):4321-4325. doi 10.1093/nar/8.19.4321

Naranjo T., Lacadena J.R., Giraldez R. Interaction between wheat and rye genomes on gomologous and gomoelogous pairing. *Z. Pflanzenzuchtg*. 1979;82:289-305

Orlovskaya O.A., Leonova I.N., Adonina I.G., Salina E.A., Khotyleva L.V., Shumny V.K. Molecular-cytogenetic analysis of triticale and wheat lines with introgressions of the tribe Triticeae species genetic material. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2015;19(5):552-560. doi 10.18699/VJ15.072 (in Russian)

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

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

Rahmatov M., Rouse M.N., Steffenson B.J., Anderson S.C., Wanyera R., Pretorius Z.A., Houben A., Kumarse N., Bhavani S., Johanson E. Sources of stem rust resistance in wheat-alien introgression lines. *Plant Dis*. 2016;100(6):1101-1109. doi 10.1094/PDIS-12-15-1448-RE

Rayburn A.L., Gill B.S. Isolation of a D-genome specific repeated DNA sequence from *Aegilops squarrosa*. *Plant Mol Biol Rep*. 1986;4:104-109. doi 10.1007/BF02732107

Rouse M.N., Nava I.C., Chao S., Anderson J.A., Jin Y. Identification of markers linked to the race Ug99 effective stem rust resistance gene *Sr28* in wheat (*Triticum aestivum* L.). *Theor Appl Genet*. 2012;125(5):877-885. doi 10.1007/s00122-012-1879-6

Salina E., Borner A., Leonova I., Korzun V., Laikova L., Maystrenko O., Roder M.S. Microsatellite mapping of the induced sphaerococcoid mutation genes in *Triticum aestivum*. *Theor Appl Genet*. 2000;100:686-689. doi 10.1007/s001220051340

Salina E.A., Lim Y.K., Badaeva E.D., Scherban A.B., Adonina I.G., Amosova A.V., Samatadze T.E., Vatolina T.Y., Zoshchuk S.A., Leitch A. Phylogenetic reconstruction of *Aegilops* section *Sitopsis* and the evolution of tandem repeats in the diploids and derived wheat polyploids. *Genome*. 2006;49(8):1023-1035. doi 10.1139/g06-050

Schneider A., Linc G., Molnar-Lang M., Graner A.Fluorescence *in situ* hybridization polymorphism using two repetitive DNA clones in different cultivars of wheat. *Plant Breed*. 2003;122(5):396-400. doi 10.1046/j.1439-0523.2003.00891.x

Schubert I., Shi F., Fuchs J., Endo T.R. An efficient screening for terminal deletions and translocations of barley chromosomes added to

common wheat. *Plant J.* 1998;14(4):489-495. doi 10.1046/j.1365-313X.1998.00125.x

Sibikeev S.N., Druzhin A.E. Prebreeding research of near-isogenic lines of spring bread wheat with a combination of translocations from *Agropyron elongatum* (Host.) P.B. and *Aegilops ventricosa* Tausch. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2015;19(3):310-315. doi 10.18699/VJ15.040 (in Russian)

Sibikeev S.N., Krupnov V.A. Evolution of leaf rust and protection from it in the Volga region. *Vestnik Saratovskogo Gosuniversiteta im. N.I. Vavilova = The Bulletin Saratov State Agrarian University in Honor of N.I. Vavilov*. 2007;S:92-94 (in Russian)

Sibikeev S.N., Druzhin A.E., Gultyaeva E.I., Andreeva L.V. Analysing effects of 4AS.4AL-7S#2S translocations upon yields and grain quality of spring milling wheat. *Uspekhi Sovremennogo Estestvoznaniya = Advances in Current Natural Sciences*. 2019;8:34-38. doi 10.17513/use.37179 (in Russian)

Singh S.J., McIntosh R.A. Allelism of two genes for stem rust resistance in triticale. *Euphytica*. 1988;38:185-189. doi 10.1007/BF00040190

Stakman E.C., Stewart D.M., Loegering W.Q. Identification of Physiologic Races of *Puccinia graminis* var. *tritici*. Washington, DC, USA: United States Department of Agriculture – Agricultural Research Service, 1962. Available: https://www.ars.usda.gov/ARSUserFiles/50620500/Cerealrusts/Pgt/Stakman_code_Pgt.pdf

Tsilo T.J., Jin Y., Anderson J.A. Diagnostic microsatellite markers for detection of stem rust resistance gene *Sr36* in diverse genetic backgrounds of wheat. *Crop Sci*. 2008;48(1):253-261. doi 10.2135/cropsci2007.04.0204

Wan H., Yang M., Li J., Wang Q., Liu Z., Zhang J., Li S., Yang N., Yang W. Cytological and genetic effects of rye chromosomes 1RS and 3R on the wheat-breeding founder parent Chuanmai 42 from southwestern China. *Mol Breed*. 2023;43(5):40. doi 10.1007s11032-023-01386-0

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

Conflict of interest. The authors declare no conflict of interest.
Received January 30, 2025. Revised April 10, 2025. Accepted May 19, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):789-797

doi 10.18699/vjgb-25-86

Variability of the mineral composition of durum wheat grain (*Triticum durum* Desf.) under different environmental conditions

I.N. Leonova [1] ✉, P.N. Malchikov [1, 2], N.A. Vinichenko [1], V.V. Piskarev [1, 3], M.G. Myasnikova [2], V.A. Aparina [1, 3], T.V. Chaheeva [2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Samara Federal Research Scientific Center of the Russian Academy of Sciences, Samara Scientific Research Agriculture Institute named after N.M. Tulaykov, Bezenchuk, Samara region, Russia
[3] Siberian Research Institute of Plant Production and Breeding – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Krasnoobsk, Novosibirsk region, Russia

✉ leonova@bionet.nsc.ru

Abstract. The composition of wheat grain plays a key role in determining its nutritional value. In this work, a collection of 133 durum wheat varieties and breeding lines was assessed for the content of macroelements (Ca, Mg, K), microelements (Cu, Mn, Zn, Fe, Na) and toxic metals (Pb, Cd, and Cr) in grain under the environmental conditions of Samara and Novosibirsk regions in 2023. The results showed a wide range of variations in the concentration of all elements depending on genotypic differences between the samples as well as the growing region. Ca and Mg contents in the varieties grown in Samara region showed a significant excess of 3.1- and 1.5-fold, respectively. Zn, Pb, and Cr content in the varieties cultivated in Novosibirsk turned out to be two times as high. Statistical analysis of element concentrations in the varieties of different origin indicates that Russian breeding lines significantly outperform Russian cultivars in Mg content, while being inferior in K, Cu, and Mn. Compared to Russian cultivars and breeding lines, foreign varieties demonstrated higher contents of K and heavy metals Cd and Cr. Correlation analysis using mean values of indicators for two environments showed highly significant ($p < 0.001$) positive relationships between the content of microelements Fe/Mn ($r^2 = 0.69$), Fe/Zn ($r^2 = 0.49$), and Zn/Mn ($r^2 = 0.46$), which suggests a feasibility of selecting genotypes for several elements at once. Multivariate statistics divided the durum wheat collection into two groups, one of them including Russian cultivars and breeding lines as well as some foreign genotypes. A separate cluster included seven Russian breeding lines placed at a distance from the other varieties, which suggested their potential differences at the genetic level. Comparing these lines with respect to mineral composition showed that they were, on average, characterized by higher Mg, K, Zn, and Fe contents. The data obtained in this study can be used for genetic research and breeding to improve the grain mineral composition of the modern durum wheat varieties.

Key words: durum wheat; macroelements; microelements; heavy metals

For citation: Leonova I.N., Malchikov P.N., Vinichenko N.A., Piskarev V.V., Myasnikova M.G., Aparina V.A., Chaheeva T.V. Variability of the mineral composition of durum wheat grain (*Triticum durum* Desf.) under different environmental conditions. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):789-797. doi 10.18699/vjgb-25-86

Funding. The study was supported by the Russian Science Foundation, project No. 23-16-00041 (https://rscf.ru/project/23-16-00041).

Acknowledgements. Multiplication of the varieties from Russian originators was carried out at the Collective Center for Plant Reproduction of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences and supported by the budget project FWNR-2022-0017.

Вариабельность минерального состава зерна твердой пшеницы (*Triticum durum* Desf.) в различных экологических условиях

И.Н. Леонова [1] ✉, П.Н. Мальчиков [1, 2], Н.А. Виниченко [1], В.В. Пискарев [1, 3], М.Г. Мясникова [2], В.А. Апарина [1, 3], Т.В. Чахеева [2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Самарский федеральный исследовательский центр Российской академии наук, Самарский научно-исследовательский институт сельского хозяйства им. Н.М. Тулайкова, п. г. т. Безенчук, Самарская область, Россия
[3] Сибирский научно-исследовательский институт растениеводства и селекции – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, р. п. Краснообск, Новосибирская область, Россия

✉ leonova@bionet.nsc.ru

Аннотация. Минеральный состав зерна пшеницы играет ключевую роль в определении его питательной ценности. В данной работе проведена оценка коллекции из 133 сортов и селекционных линий яровой твердой пшеницы отечественного и иностранного происхождения по содержанию в зерне макроэлементов (Ca, Mg, K), микроэлементов (Cu, Mn, Zn, Fe, Na) и токсичных металлов (Pb, Cd, Cr) при выращивании в экологических условиях Самарской и Новосибирской областей. Результаты показали широкий размах варьирования концентрации всех элементов в зависимости от генотипических различий между образцами и региона выращивания. Значительное превышение концентрации Ca и Mg в 3.1 и 1.5 раза соответственно отмечено у сортообразцов, выращенных в Самарской области. Культивирование образцов в условиях Новосибирской области сопровождалось превышением содержания Zn, Pb и Cr в зерне более чем в два раза. Статистический анализ содержания минеральных элементов у образцов различного происхождения свидетельствует о том, что российские селекционные линии достоверно превышают отечественные сорта по содержанию Mg, но при этом уступают по концентрации K, Cu и Mn. Сортообразцы иностранной селекции отличались от российских сортов и линий повышенным содержанием K и тяжелых металлов Cd и Cr. Анализ корреляций, проведенный на основании средних значений показателей по двум регионам, указывает на наличие высокодостоверных ($p < 0.001$) положительных взаимосвязей между содержанием микроэлементов Fe/Mn ($r^2 = 0.69$), Fe/Zn ($r^2 = 0.49$) и Zn/Mn ($r^2 = 0.46$), что предполагает возможность проведения отбора генотипов по нескольким элементам одновременно. Многомерный статистический анализ разделил образцы на две группы, одна из которых включала российские сорта и селекционные линии, а также часть иностранных образцов. Отдельный кластер состоял из семи российских селекционных линий, расположенных дистантно от остальных образцов, что предполагает их различия на генетическом уровне. Сравнение этих линий по минеральному составу показало, что линии в среднем характеризуются более высокими концентрациями Mg, K, Zn и Fe. Данные по содержанию микро- и макроэлементов в изученных сортообразцах яровой твердой пшеницы могут быть использованы для генетических исследований и практической селекции для улучшения существующих сортов по минеральному составу.

Introduction

Mineral micro- and macronutrients play a significant role in maintaining human body functions and have a substantial impact on human health. The functional role of most mineral elements is diverse. In particular, they act as cofactors in various enzymatic reactions and are also involved in redox reactions during electron transfer, oxygen binding and transport in tissues, interaction of molecules with a cellular receptor, and regulation of gene expression (Sigel et al., 2013; Jomova et al., 2022; Islam et al., 2023).

Despite the importance of mineral nutrients, their optimal doses are also critical for the normal functioning of the body, as both excess and deficiency of certain minerals can lead to various physiological disorders. Deficiency of macroelements such as calcium, magnesium, and potassium causes muscular system disorders, changes in hormonal status, and may lead to malignant growths (Zoroddu et al., 2019; Ali, 2023). Iron deficiency is among the causes of anemia, cardiovascular diseases, and immune system disorders due to iron being incorporated into hemoproteins, such as hemoglobin and enzymes for xenobiotic degradation (Camaschella, 2019; Dixit et al., 2020). Insufficient zinc intake leads to growth and sexual development delays, reduced immunity, and various mental disorders (Hambidge, 2000). Symptoms of copper deficiency may include various joint lesions and pigmentation disorders in skin and hair (Olivares, Uauy, 1996). On the other hand, excessive intake of zinc, iron, and copper can lead to liver fibrosis and cirrhosis, neurodegenerative diseases, impaired immune and cognitive functions, and severe forms of anemia (Wessling-Resnick, 2017; Schoofs et al., 2024).

In addition to vital micro- and macroelements required in optimal concentrations, humans are also exposed to a number of toxic metals (lead, cadmium, mercury, chromium, aluminum), which can enter the body through food and have negative effects. The main toxicity mechanisms of heavy metals include mitochondrial apoptosis, interference with various signaling pathways and oxidative stress, changes in gene activity regulation due to various types of DNA damage, all of which can lead to the development of chronic diseases and the emergence of malignancies (Kiran et al., 2022; Jomova et al., 2025).

Food products are the main sources of minerals for the human body. For most of the world's population, food products made from bread and durum wheat are the main sources of protein, vitamins, and minerals. Bread wheat (*Triticum aestivum* L., $2n = 42$, AABBDD genome) is one of the most valuable food grain crops, ranking among the highest-rated crops in most countries. Daily consumption of bread wheat products provides up to 20 % of the required calories and up to 10–15 % of iron and zinc (Tadesse et al., 2019; Aghalari et al., 2022).

Unlike bread wheat, durum wheat (*Triticum durum* Desf., $2n = 28$, AABB genome) is the only raw material for the production of high-quality macaroni products with a characteristic amber color and excellent taste. According to available data, a significant number of cultivated durum wheat varieties outperform bread wheat varieties in the content of zinc, iron, calcium, magnesium, and other minerals (Cakmak et al., 2010; Del Coco et al., 2019; Saini et al., 2023). The main producers and consumers of durum wheat products are the Mediterranean countries (Italy, Turkey, Greece, Tunisia, France), which account for more than 50 % of the cultivated area. Other major suppliers of durum wheat include Canada, the United States, Mexico, India, and Kazakhstan. Until the mid-20th century, Russia was one of the major producers of durum wheat grain and ranked first in the world in terms of sown area, which reached 20 million hectares (Martínez-Moreno et al., 2022; Malchikov, Myasnikova, 2023). A significant reduction in

sowing and grain harvesting occurred in the early 1990s after the collapse of the USSR, and until recently, the area under durum wheat was estimated at ~0.7 million hectares, which is no higher than 1.7 % of the global area (Goncharov, Kurashov, 2018).

In recent years, considerable attention has been paid to improving the mineral composition of wheat grain and increasing the concentration of essential nutrients. Biofortification breeding programs were used as vehicles to introduce the newly developed bread wheat samples with genetically increased zinc and iron contents (Khokhar et al., 2018; Virk et al., 2021; Tanin et al., 2024). Efforts on developing biofortified breeding lines of bread wheat and searching for donors with high protein, mineral, and antioxidant content are made in the Russian Federation as well (Morgounov et al., 2022; Potapova et al., 2023; Gordeeva et al., 2024). The data from long-term trials of wheat landraces and modern varieties as well as breeding and introgression lines focusing on grain quality and mineral composition traits have been systematized, which made it possible to identify the genotypes with target traits for breeding schemes (Shepelev et al., 2022; Orlovskaya et al., 2023; Leonova et al., 2024; Shamanin et al., 2024). However, when it comes to durum wheat, similar studies on the genetic diversity of Russian varieties and breeding lines in terms of grain mineral composition are critically lacking (Pototskaya et al., 2023; Sochalova et al., 2023).

The goal of this study was to investigate the genetic variability in contents of microelements (zinc, iron, copper, and manganese), macroelements (calcium, magnesium, and potassium), and toxic metals (lead, cadmium, and chromium) in the collection of spring durum wheat varieties and breeding lines grown under the environmental conditions of the Samara and Novosibirsk regions.

Materials and methods

Plant material and field trial conditions. Plant material consisted of 133 durum wheat samples, including 35 Russian cultivars, 68 Russian breeding lines, 29 foreign accessions, and Turanian wheat Khorasan (Table S1)[1]. Among the Russian breeding lines, 39 genotypes were developed at the Samara Scientific Research Agriculture Institute, whereas the remaining ones originated from other Russian breeding centers. The samples were grown under field trial conditions on the premises of research institutes in the Novosibirsk and Samara regions in 2023. In the Novosibirsk region, the sowing was carried out in the field of the Siberian Research Institute of Plant Production and Breeding – a branch of the Institute of Cytology and Genetics SB RAS (54°54′51.4″N, 82°58′37.1″E). The soils of the experimental site are mainly leached chernozem of medium depth. The humus content in the topsoil is 5.7–6.9 %. The soils are highly supplied with mobile soil phosphates and potassium. The contents per 100 g of soil are as follows: P_2O_5 – 42 mg, K_2O – 35 mg. The total nitrogen content in the soil before sowing was 0.31 %. Sowing was done manually in furrows at a depth of 5–7 cm. The plot size was 0.4 m^2, in two replicates.

[1] Tables S1–S4 are available in Supplementary Materials at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Leonova_Engl_29_6.xlsx

The experimental fields of the Samara Scientific Research Agriculture Institute – a branch of the Samara Federal Research Scientific Center of the Russian Academy of Sciences – are located in the urban-type settlement of Bezenchuk (52°05′85.5″N, 49°02′55.9″E). The main soil type is common chernozem with medium to heavy loam texture. The average humus content is 4.8 %, nitrogen – 5.9 mg/kg, phosphorus – 279 mg/kg, potassium — 203 mg/kg. The variety samples were sown in a randomized layout on plots of 7.4 m^2 in two replicates. The variety Bezenchukskaya 210 was used as a standard in both fields.

The comparison of the climatic conditions during the growing season between the regions, as well as the comparison of the results with long-term average values showed that in 2023, Novosibirsk region experienced elevated air temperatures throughout the growing season, with the exception of the second ten-day periods of May and August, when temperatures were below the long-term average (–1.3 and –2.0 °C against the long-term average, respectively) (Table S2). The most significant increase (+8.4 ºC) above the long-term average temperature was recorded in the first ten days of June, while precipitation during this period was only 27.8 % of the long-term average. Overall, the first half of the growing season (May–June) was characterized by a severe lack of precipitation (14.9 and 47.5 %, respectively), while in July precipitation fell within the normal range (102.1 %), and in August it was significantly above the long-term average (167.6 %). In the Samara region, there was a general precipitation deficit throughout the growing season against a background of moderate temperatures. From seedling stage to wax ripeness, the total precipitation was 89.3 mm. The most favorable conditions for durum wheat cultivation during the growing season occurred in the first and third ten-day periods of June, during the booting and heading stages, respectively.

Content evaluation of micro- and macronutrients and heavy metals. The grains of durum wheat samples were analyzed with respect to contents of eight micro- and macronutrients (zinc, iron, copper, manganese, sodium, calcium, magnesium, and potassium) and three toxic metals (lead, cadmium, and chromium). Here, a 300 mg of grain was treated with 1 ml of hydrogen peroxide (60 %) and 5 ml of concentrated nitric acid. Sample mineralization was carried out in a microwave oven for 40 minutes. The sample volume was then brought to 50 ml with deionized water and diluted 50 times for element quantification. The chemical composition was analyzed using an atomic absorption spectrometry (AAS) device ContrAA 800 D (Analytik Jena, Germany). Each sample was analyzed in duplicate.

Statistical analysis. Statistical analysis was performed using the Statistica v. 10 software (StatSoft, Inc., USA). The significance of differences between mean trait values was assessed using the Mann–Whitney test and Student's *t*-test. Trait values are presented as means (M) and standard deviations (± SD). To evaluate the effects of genotype and environmental factors, two-way analysis of variance (ANOVA) was used. The relationship between the contents of various elements was assessed using Spearman's correlation coefficient. Principal component analysis (PCA) and dendrogram plotting were

performed using the PAST v. 4.03 software (Hammer et al., 2001). The UPGMA (unweighted pair group method with arithmetic mean) method was used for dendrogram plotting, and statistical significance of clustering was evaluated using a permutation test with 1,000 iterations.

Results

The assessment of mineral element content in the durum wheat grain grown under the environmental conditions of two regions (Samara and Novosibirsk regions) revealed a wide range of variation in nutrient concentration, as well as differences depending on the growing region (Table 1; Table S1). The most significant regional differences were observed for elements such as Ca and Mg, the content of which in field conditions of the Samara region was 3.1 and 1.5 times higher, respectively. For samples grown in the Novosibirsk region, higher concentrations of Zn, Pb, and Cr were recorded: specifically 2.5, 2.3, and 2.2 times higher, respectively. All regional differences in element contents were statistically significant ($p < 0.0001$), except for Na (Table 1), the concentration of which was the same in both regions. The distribution of most mineral contents in genotypes grown in both fields was approximately normal, with the exception of lead concentration in the field of the Samara Scientific Research Agriculture Institute, which was significantly shifted toward lower values.

ANOVA used to assess the effects of genotype and environmental factors on the phenotypic expression of traits showed that field conditions most affected the contents of Ca and Zn, with low and non-significant contribution of genotype observed for Ca concentration (Table S3). In contrast, Na content was mainly influenced by genotypic differences. A significant genotypic effect was found for the concentrations of Mg, K, Cu, Fe, and Cd, which substantially exceeded the influence of environmental factors.

Since the studied collection consisted of several groups of varieties of different origin, it was of interest to determine whether these groups differed in element concentrations. It can be seen from Table 2 that Russian breeding lines significantly exceeded Russian cultivars in Mg content, but had lower levels of K, Cu, and Mn. Foreign varieties differed from Russian cultivars and lines by higher potassium content. The concentrations of toxic elements Cd and Cr were significantly higher on average in foreign samples compared to Russian ones, specifically 1.9 and 1.8 times higher, respectively. No significant differences between groups were found for Zn and Fe concentrations.

According to the averaged data from both regions, the highest Ca content (>700 mg/kg) was observed in the Russian varieties Pamyati Chekhovicha, Zhemchuzhina Sibiri, Annushka, Krasnokutka 13, and in the Russian breeding lines L73, L75, and L76 developed at the Altai Research Institute of Agriculture. In terms of Mg content, the breeding lines L21, L23, L24, L25, and L26 from the Samara Scientific Research Agriculture Institute stood out, with element concentrations exceeding 2,000 mg/kg. The Russian varieties Annushka, Bezenchukskaya Stepnaya, Volnodonskaya, and the foreign varieties Hyperno and Tamaroi were characterized by high zinc and iron content (>54 and 51 mg/kg, respectively). High

Table 1. Contents of micro- and macronutrients and toxic metals (mg/kg) in grains of the durum wheat varieties grown under the environmental conditions of the Samara and Novosibirsk regions in 2023

Element	Samara Region	Novosibirsk Region
Macronutrients (M ± SD)		
Ca	920.4 ± 170.7****	297.4 ± 116.9
Mg	1,327.0 ± 178.2***	891.4 ± 276.8
K	3,977.6 ± 402.9***	4,356.7 ± 577.9
Micronutrients (M ± SD)		
Cu	3.2 ± 0.9*	3.7 ± 1.1
Mn	37.1 ± 7.9**	27.0 ± 6.5
Zn	25.8 ± 8.1****	64.0 ± 12.7
Fe	47.2 ± 6.0***	42.2 ± 8.8
Na	27.2 ± 6.3^{ns}	27.3 ± 12.1
Toxic metals (M ± SD))		
Pb	0.23 ± 0.05***	0.52 ± 0.02
Cd	0.05 ± 0.004***	0.03 ± 0.002
Cr	0.82 ± 0.05***	1.79 ± 0.11

Note. The data are presented as mean value (M) ± standard deviation (SD); * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns – differences are non-significant.

Cd concentrations, which in some cases exceeded permissible limits, were detected in the foreign varieties Tessadur, Achille, Fuego, and the breeding lines L51 and L56 (Table S1). The high levels of elements such as Mg (1,598.2 mg/kg), K (4,536.5 mg/kg), and Zn (56.1 mg/kg) observed in the ancient wheat Khorasan are also worth noting, confirming the data obtained by other authors and suggesting the potential of this species for improving the nutritional value of modern durum wheat varieties (Bordoni et al., 2017).

Spearman correlation test based on the mean values from both regions indicated highly significant positive correlations between Fe and Mn ($r^2 = 0.69$), Fe/Zn ($r^2 = 0.49$), Zn/Mn ($r^2 = 0.46$), Zn/Pb ($r^2 = 0.41$), and Cr/K ($r^2 = 0.41$) (Fig. 1; Table S4). Weak negative correlations were observed in the pairs as follows: Cu/Ca, Cu/Mg, Cu/Na, and Pb/Cr ($r^2 = -0.22$, -0.24, -0.27, and -0.34, respectively).

The results of mineral content assessment in the two regions were used to identify potential clustering of durum wheat varieties. Principal component analysis (PCA) was used to analyze the relationship between element concentrations and the affiliation of genotypes to the groups of different origin: Russian varieties, breeding lines from the Samara Scientific Research Agriculture Institute, breeding lines from other Russian breeding centers, and foreign accessions (Fig. 2).

Table 2. Contents of micro- and macronutrients and toxic metals in grains of varieties and breeding lines of Russian and foreign origin (Samara and Novosibirsk regions, 2023)

Element (mg/kg)	Russian		Foreign varieties[b]
	varieties	breeding lines[a]	
Ca	625.1 ± 398.3	626.2 ± 315.9	548.3 ± 336.3
Mg	1,178.9 ± 330.8**[b]	1,225.1 ± 644.5**[b]	1,017.6 ± 317.4
K	4,222.4 ± 718.2*[a]	4,026.5 ± 382.2****[b]	4,428.4 ± 456.1
Cu	3.73 ± 0.9***[a]	3.15 ± 0.9***[b]	3.76 ± 1.2
Mn	34.1 ± 8.2*[a]	30.5 ± 9.1	33.2 ± 8.1
Zn	48.8 ± 19.4	43.9 ± 23.2	42.5 ± 21.6
Fe	43.9 ± 8.4	44.9 ± 7.2	44.9 ± 8.9
Na	28.5 ± 10.1****[b]	29.1 ± 13.2****[b]	21.6 ± 5.5
Pb	0.49 ± 0.19***[ab]	0.35 ± 0.11	0.29 ± 0.01
Cd	0.035 ± 0.008****[b]	0.030 ± 0.002****[b]	0.064 ± 0.005
Cr	1.0 ± 0.4****[b]	1.2 ± 0.6****[b]	1.9 ± 0.6

Note. The data are presented as mean value ± standard deviation. Letters ([a, b]) indicate statistically significant differences between groups; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

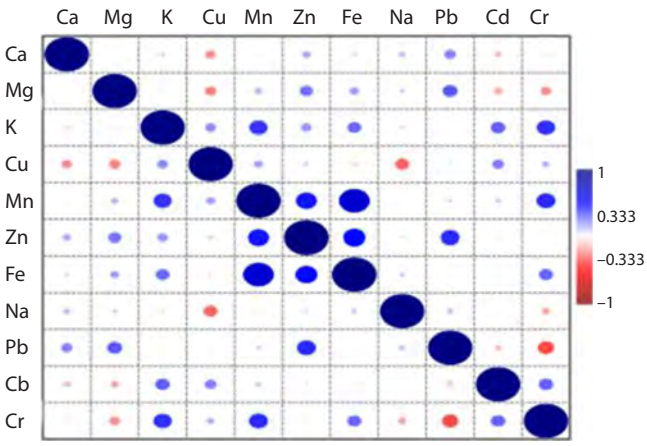


Fig. 1. Genotypic correlations between the concentrations of micro- and macronutrients and toxic metals in grains of durum wheat cultivars and breeding lines.

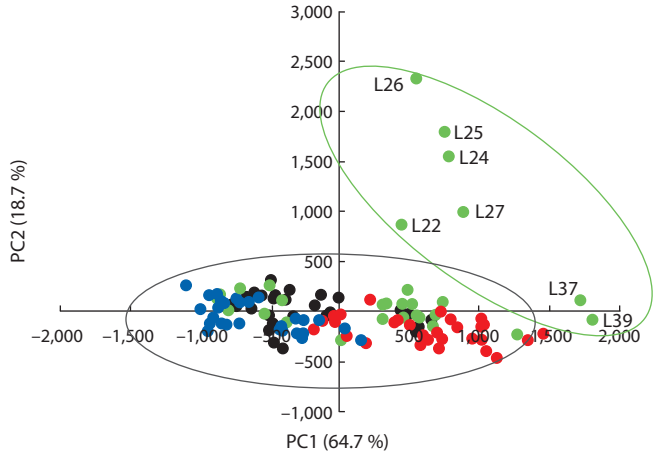


Fig. 2. PCA plot illustrating the distribution of durum wheat genotypes in the space of the first two principal components.

Durum wheat samples are color-coded as follows: black – Russian varieties; blue – breeding lines from Russian institutions; green – breeding lines from the Samara Research Institute; red – foreign varieties.

Considering the first two principal components, which account for 64.7 and 18.7 % of the genetic variation, respectively, all studied varieties are divided into two groups. Most of the genotypes are clustered into one large group, with Russian durum wheat cultivars, breeding lines from Russian centers, and partially lines from the Samara Scientific Research Agriculture Institute grouped on the left side of this cluster. Foreign varieties are primarily located in the lower right part of the cluster. Seven lines (L22, L24, L25, L26, L27, L37, and L39) form a separate group and are positioned at a considerable distance from the other samples, which may indicate genetic-level differences. Comparing the mineral compositions of these lines showed that, on average, they were characterized by higher concentrations of Mg, K, Zn, and Fe (Table S1).

Clustering of the samples using the UPGMA method confirmed the presence of two main groups (Fig. 3). Cluster 1 includes seven breeding lines (L22, L24, L25, L26, L27, L37, and L39) from the Samara Scientific Research Agriculture Institute and is clearly separated from the other samples. The second cluster is divided into two subclusters (labeled 2 and 3 in the Figure), one of which mainly includes Russian varieties and breeding lines from Russian breeding centers. Foreign varieties are predominantly grouped in the right part of subcluster 3.

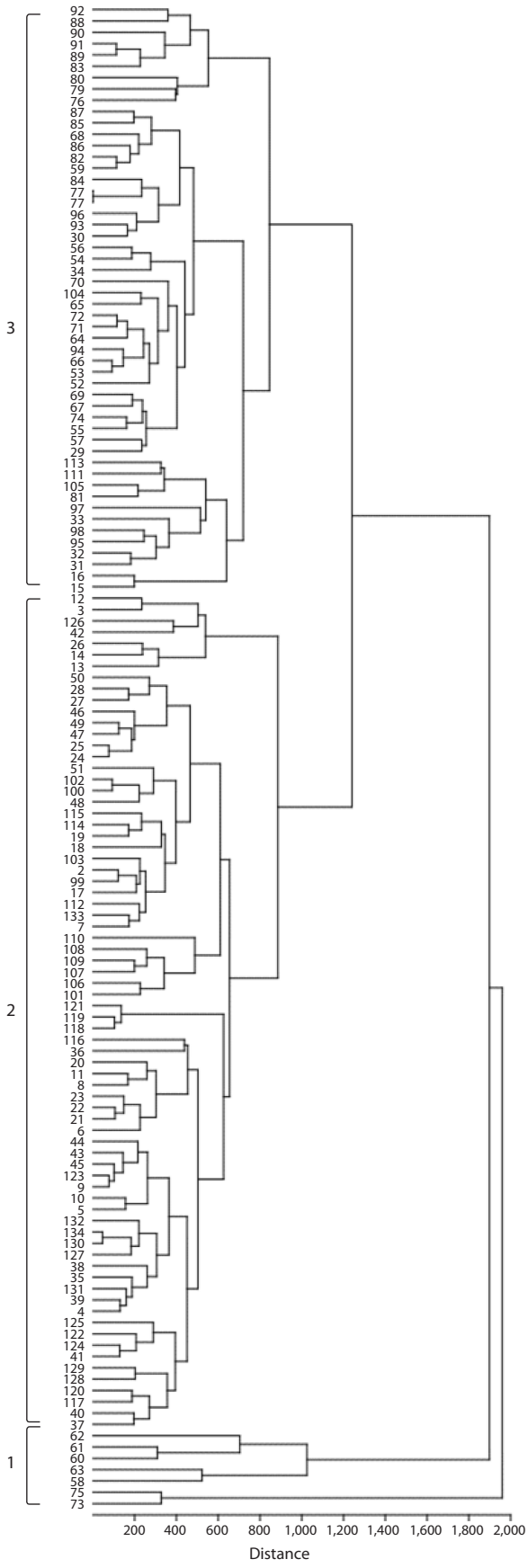


Fig. 3. UPGMA dendrogram illustrating the clustering of durum wheat genotypes based on the assessment of concentrations of 11 micro- and macronutrients and toxic metals in grains.

Discussion

In this paper, we investigated the grain concentrations of 11 micro- and macronutrients and toxic metals in durum wheat genotypes of both Russian and foreign origin, grown under the environmental conditions of the Samara and Novosibirsk regions. A wide range of variation was observed for all elements depending on both genotype and environmental factors. Significant regional differences were identified in the concentrations of essential macro- and micronutrients (Ca, K, and Zn) and toxic metals (Pb, Cd, and Cr) (Table S1). It is worth noting that the variation range for K, Mg, Cu, Mn, and Fe in both regions was comparable to the results reported by I.V. Pototskaya et al. (2023) obtained from the cultivation of 20 Russian and Kazakh durum wheat varieties in the Omsk region. An exception was observed for Ca and Zn contents, with the former being significantly higher and the latter lower in the Omsk fields compared to Novosibirsk.

An important aspect is the grain concentration of toxic metals, since exceeding the maximum permissible concentrations (MPCs) of essential trace nutrients such as Zn and Cu in food products can lead to harmful effects. Most varieties grown in the Novosibirsk region showed zinc concentrations in grains above the MPC (50 mg/kg), whereas the varieties from the Samara region showed contents falling within the acceptable range (Table S1). These differences may be related to zinc content in soils of the respective regions. Similarly, an increased lead content in grains was observed in nearly half of the genotypes grown in both regions.

According to a 30-year monitoring study of the mineral composition of soils in the chernozems of Western Siberia, no exceedance of the MPCs for mobile forms of heavy metals such as Cu, Zn, Pb, Cd, and Cr was detected (Krasnitsky et al., 2024). Soil analysis in the Samara region also indicates that the content of mobile forms of lead, nickel, chromium, copper, and zinc in all soil types and subtypes with varying composition and humus content does not exceed the MPCs (Obushchenko, Gnedenko, 2014). However, no analysis of toxic metals in the soils of the experimental fields was conducted, and this issue requires further investigation.

Changes in heavy metal concentrations in soil may be influenced by factors such as soil preparation systems, application of organomineral fertilizers, use of pesticides and plant protection products, atmospheric precipitation, and various anthropogenic impacts from industrial enterprises, transportation, and others. Several studies have shown that systematic introduction of mineral and organic fertilizers contributes to the accumulation of heavy metals in soil and may affect the mineral composition and concentration in wheat grains (Ryan et al., 2004; Pandino et al., 2020; Wysocka et al., 2025). Root and foliar treatments with nitrogen fertilizers at various concentrations can increase the uptake and accumulation of Cd in durum wheat grains even at low soil Cd levels (Özkutlu, 2024). The introduction of organomineral fertilizers to barley crops in the soils of the Samara region, for example, led to an increase in the total content and mobility of heavy metals (Cu, Cd, Pb, Zn) and enhanced Pb migration into spring wheat plants (Trots V.B. et al., 2015; Trots N.M., Bokova, 2024).

Heavy metal contamination may also occur in soils located near industrial facilities (Prosyannikov, 2014). The mechanisms of heavy metal transfer and accumulation in soil–plant systems are highly complex, and the intensity of accumulation depends on the type of toxic element, weather conditions, regional soil structure, as well as the species and genotype of the plant. However, despite the increase in heavy metal concentrations in agricultural soils, most authors report that this does not significantly affect the accumulation of toxic substances in final agricultural products (Protasova, 2014; Wang et al., 2017; Ugulu et al., 2021).

The data obtained on the concentrations of micro- and macronutrients and heavy metals were used to study the grouping of genotypes in principal component space and to plot a phylogenetic tree. The results showed no clear separation into contrasting groups, except for a small number of lines developed at the Samara Scientific Research Agriculture Institute and foreign varieties (Fig. 2 and 3). Russian lines from different breeding centers formed a single cluster with durum wheat cultivars introduced into production in different years, which may indicate insufficient genetic diversity among both varieties and breeding lines. The modern global durum wheat pool is characterized by moderate genetic diversity, as confirmed by the results obtained from a wide range of genotypes from different countries (Zhao et al., 2009; Hakki et al., 2014; Hocaoğlu et al., 2020; Naseri et al., 2024).

The genetic diversity of Russian spring durum wheat varieties developed between 1929 and 2004 was previously assessed using pedigree data (Martynov et al., 2004). The analysis revealed that about 20 % of the pool of original Russian durum wheat varieties previously used in hybridization had been lost. A similar conclusion can be drawn for foreign durum wheat varieties, as breeding efforts over a long period focused primarily on developing high-yielding genotypes with resistance to diseases and lodging, which led to a narrowing of the gene pool (Hernandez-Espinosa et al., 2020; Xynias et al., 2020). To date, no comprehensive studies of the genetic pool of Russian spring durum wheat in terms of mineral composition have been conducted, so it is not yet possible to draw a conclusion regarding the landrace superiority.

The search for sources of genetic diversity draws attention to the inclusion of wheat relatives in hybridization in an attempt to increase the content of essential micro- and macronutrients in grains. It is known that tetraploid species such as *T. dicoccum*, *T. dicoccoides*, *T. timopheevii*, and hybrid lines derived from them are characterized by significantly higher levels of zinc and iron in grains (Cakmak et al., 2010; Del Coco et al., 2019; Tekin et al., 2022; Leonova et al., 2024). The data on the origin of the seven breeding lines from the Samara Scientific Research Agriculture Institute that formed a separate cluster indicate that tetraploid species *T. dicoccum* and *T. timopheevii* were used in their development. A previous microsatellite analysis of early generations of these lines revealed the presence of alien translocations from *T. timopheevii* in chromosome 6B (Malchikov et al., 2015). However, further research is needed to establish a link between the presence of alien insertions in the genome and their effect on the content of specific elements.

Conclusion

The assessment of micro- and macronutrient and toxic metal content in grains of spring durum wheat varieties conducted in this study revealed significant variation in most elements depending on genotype and environmental conditions. The presence of significant positive correlations Fe/Mn, Fe/Zn, and Zn/Mn indicates the possibility of selecting genotypes based on multiple micronutrients simultaneously. The analysis of the results suggests that Russian varieties and breeding lines exhibit moderate genetic diversity; however, the observed range of trait variation allows for the identification of samples that can be used to improve the mineral composition of grain. It was shown that Cd content in grains of all studied samples, except for five foreign varieties, does not exceed the maximum permissible concentrations (MPCs). To determine the cause of the elevated Zn, Pb, and Cr contents exceeding the MPCs in some samples, additional data on the concentration of these metals in the soil of the experimental plots are required.

References

Aghalari Z., Dahms H.U., Sillanpää M. Evaluation of nutrients in bread: a systematic review. *J Heal Popul Nutr.* 2022;41(1):50. doi 10.1186/s41043-022-00329-3

Ali A.A.H. Overview of the vital roles of macro minerals in the human body. *J Trace Elem Miner.* 2023;4:100076. doi 10.1016/j.jtemin.2023.100076

Bordoni A., Danesi F., Di Nunzio M., Taccari A., Valli V. Ancient wheat and health: a legend or the reality? A review on KAMUT khorasan wheat. *Intern J Food Sci Nutrition*. 2017;68(3):278-286. doi 10.1080/09637486.2016.1247434

Cakmak I., Pfeiffer W.H., McClafferty B. Biofortification of durum wheat with zinc and iron. *Cereal Chem.* 2010;87(1):10-20. doi 10.1094/CCHEM-87-1-0010

Camaschella C. Iron deficiency. *Blood*. 2019;133(1):30-39. doi 10.1182/blood-2018-05-815944

Del Coco L., Laddomada B., Migoni D., Mita G., Simeone R., Fanizzi F.P. Variability and site dependence of grain mineral contents in tetraploid wheats. *Sustainability.* 2019;11(3):736. doi 10.3390/su11030736

Dixit S.P., Rajan L., Palaniswamy D., Mohankumar S.K. Importance of iron absorption in human health: an overview. *Curr Nutr Food Sci.* 2020;17(3):293-301. doi 10.2174/1573401316999200801021752

Goncharov S.V., Kurashov M.Yu. Prospects for the development of the Russian durum wheat market. Vestnik of Voronezh State Agrarian University. 2018;2(57):66-75. doi 10.17238/issn2071-2243.2018.2.66 (in Russian)

Gordeeva E.I., Shamanin V.P., Khlestkina E.K., Shoeva O.Y. On peculiarities of breeding purple-grained wheat based on varieties with anthocyanin pigmentation of coleoptiles and stems. *Agric Biol*. 2024;59(3):507-524. doi 10.15389/agrobiology.2024.3.507eng

Hakki E.E., Dograr N., Pandey A., Khan M.K., Hamurcu M., Kayis S.A., Gezgin S., Ölmez F., Akkaya M.S. Molecular and elemental characterization of selected Turkish durum wheat varieties. *Not Bot Horti Agrobot Cluj-Napoca*. 2014;42(2):431-439. doi 10.15835/nbha4229621

Hambidge M. Human zinc deficiency. *J Nutr.* 2000;130(5):1344S-1349S. doi 10.1093/jn/130.5.1344S

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

Hernandez-Espinosa N., Laddomada B., Payne T., Huerta-Espino J., Govindan V., Ammar K., Ibba M.I., Pasqualone A., Guzman C. Nu-

tritional quality characterization of a set of durum wheat landraces from Iran and Mexico. *LWT*. 2020;124:109198. doi 10.1016/j.lwt.2020.109198

Hocaoğlu O., Akçura M., Kaplan M. Changes in the grain element contents of durum wheat varieties of Turkey registered between 1967–2010. *Commun Soil Sci Plant Anal*. 2020;51(4):431-439. doi 10.1080/00103624.2019.1709487

Islam M.R., Akash S., Jony M.H., Alam M.N., Nowrin F.T., Rahman M.M., Rauf A., Thiruvengadam M. Exploring the potential function of trace elements in human health: a therapeutic perspective. *Mol Cell Biochem*. 2023;478(10):2141-2171. doi 10.1007/s11010-022-04638-3

Jomova K., Makova M., Alomar S.Y., Alwasel S.H., Nepovimova E., Kuca K., Rhodes C.J., Valko M. Essential metals in health and disease. *Chem Biol Interact*. 2022;367:110173. doi 10.1016/j.cbi.2022.110173

Jomova K., Alomar S.Y., Nepovimova E., Kuca K., Valko M. Heavy metals: toxicity and human health effects. *Arch Toxicol*. 2025;99:153-209. doi 10.1007/s00204-024-03903-2

Khokhar J.S., Sareen S., Tyagi B.S., Singh G., Wilson L., King I.P., Young S.D., Broadley M.R. Variation in grain Zn concentration, and the grain ionome, in field-grown Indian wheat. *PLoS One*. 2018;13(1):e0192026. doi 10.1371/journal.pone.0192026

Kiran, Bharti R., Sharma R. Effect of heavy metals: an overview. *Mater Today Proc*. 2022;51:880-885. doi 10.1016/j.matpr.2021.06.278

Krasnitsky V.M., Bobrenko I.A., Schmidt A.G., Bobrenko E.G. Monitoring of heavy metals and arsenic in chernozems and plants of the south of Western Siberia. *Plodorodie*. 2024;6(141):99-103. doi 10.25680/S19948603.2024.141.23 (in Russian)

Leonova I.N., Kiseleva A.A., Salina E.A. Identification of genomic regions conferring enhanced Zn and Fe concentration in wheat varieties and introgression lines derived from wild relatives. *Int J Mol Sci*. 2024;25(19):10556. doi 10.3390/ijms251910556

Malchikov P.N., Myasnikova M.G. Development, results and prospects of the spring durum wheat breeding in Russia (post-Soviet states). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2023;27(6):591-608. doi 10.18699/VJGB-23-71

Malchikov P.N., Maysnikova M.G., Leonova I.N., Salina E.A. Introgression of stability to powdery mildew (*Blumeria graminis* DC. f. *tritici*) from *Triticum timofeevii* Zhuk. and *Triticum dicoccum* Shuebl. in genome *Triticum durum* Desf. *Zernovoe Khozyajstvo Rossii = Grain Economy of Russia*. 2015;2:63-67 (in Russian)

Martínez-Moreno F., Ammar K., Solís I. Global changes in cultivated area and breeding activities of durum wheat from 1800 to date: a historical review. *Agronomy*. 2022;12(5):1135. doi 10.3390/agronomy12051135

Martynov S.P., Dobrotvorskaya T.V., Pukhalskiy V.A. Analysis of genetic diversity of spring durum wheat (*Triticum durum* Desf.) cultivars released in Russia in 1929–2004. *Russ J Genet*. 2005;41(10):1113-1122. doi 10.1007/s11177-005-0208-4

Morgounov A., Li H., Shepelev S., Ali M., Flis P., Koksel H., Savin T., Shamanin V. Genetic characterization of spring wheat germplasm for macro-, microelements and trace metals. *Plants*. 2022;11(16):2173. doi 10.3390/plants11162173

Naseri R., Cheghamirza K., Mohammadi R., Zarei L., Beheshti A.A. Evaluation of grain quality and its relationship with agro-physiological traits in durum wheat. *Cereal Res Commun*. 2024;52(2):813-823. doi 10.1007/s42976-023-00430-1

Obuschenko S.V., Gnedenko V.V. Monitoring of micronutrient and heavy metal content in soil of Samara region. *Mezdunarodnyj Zhurnal Prikladnykh i Fundamentalnykh Issledovanij = International Journal of Applied and Fundamental Research*. 2014;7:30-34 (in Russian)

Olivares M., Uauy R. Copper as an essential nutrient. *Am J Clin Nutr*. 1996;63(5):791S-796S. doi 10.1093/ajcn/63.5.791

Orlovskaya O.A., Vakula S.I., Khotyleva L.V., Kilchevsky A.V. Mineral composition of bread wheat lines with introgressions of alien genetic material. *Proc Appl Bot Genet Breed*. 2023;184(1):42-52. doi 10.30901/2227-8834-2023-1-42-52

Özkutlu F. Effects of applying different N sources on Cd accumulation, mineral micronutrients, and grain yield of durum wheat. *J Soil Sci Plant Nutr*. 2024;24:4261-4268. doi 10.1007/s42729-024-01831-9

Pandino G., Mattiolo E., Lombardo S., Lombardo G.M., Mauromicale G. Organic cropping system affects grain chemical composition, rheological and agronomic performance of durum wheat. *Agriculture*. 2020;10(2):46. doi 10.3390/agriculture10020046

Potapova N.A., Timoshchuk A.N., Tiys E.S., Vinichenko N.A., Leonova I.N., Salina E.A., Tsepilov Y.A. Multivariate genome-wide association study of concentrations of seven elements in seeds reveals four new loci in Russian wheat lines. *Plants*. 2023;12(17):3019. doi 10.3390/plants12173019

Pototskaya I.V., Koshkin M.N., Shpigel A.L., Shamanin V.P. Variability of the of macro- and microelements content in the grain of durum wheat under conditions of the southern forest-steppe of Western Siberia. *Vestnik Omskogo GAU = Vestnik of Omsk SAU*. 2023;2(50):58-67 (in Russian)

Prosyannikov V.I. Ecological-agrochemical characterization of arable soils in the south-eastern region of Western Siberia in terms of heavy metal contents. *Plodorodie*. 2014;5:41-43 (in Russian)

Protasova N.A. Heavy metals in chernozems and cultivated plants of the Voronezh region. *Agrokhimia*. 2005;2:80-86 (in Russian)

Ryan M.H., Derrick J.W., Dann P.R. Grain mineral concentrations and yield of wheat grown under organic and conventional management. *J Sci Food Agric*. 2004;84(3):207-216. doi 10.1002/jsfa.1634

Saini P., Kaur H., Tyagi V., Saini P., Ahmed N., Dhaliwal H.S., Sheikh I. Nutritional value and end-use quality of durum wheat. *Cereal Res Commun*. 2023;51:283-294. doi 10.1007/s42976-022-00305-x

Schoofs H., Schmit J., Rink L. Zinc toxicity: understanding the limits. *Molecules*. 2024;29(13):3130. doi 10.3390/molecules29133130

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. *Agric Biol*. 2024;59(3):492-506. doi 10.15389/agrobiology.2024.3.492eng

Shepelev S., Morgounov A., Flis P., Koksel H., Li H., Savin T., Sharma R., Wang J., Shamanin V. Variation of macro- and microelements, and trace metals in spring wheat genetic resources in Siberia. *Plants*. 2022;11(2):149. doi 10.3390/plants11020149

Sigel A., Sigel H., Sigel R. (Eds) Interrelations between Essential Metal Ions and Human Diseases. Metal Ions in Life Sciences. Vol. 13. Springer, 2013. doi 10.1007/978-94-007-7500-8

Sochalova L.P., Aparina V.A., Boyko N.I., Zuev E.V., Morozova E.V., Musinov K.K., Vinichenko N.A., Leonova I.N., Piskarev V.V. Studying a collection of common-wheat varieties for leaf rust resistance, crop yield and grain quality in the environmental conditions of Novosibirsk region. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2023;27(8):988-999. doi 10.18699/ VJGB-23-114

Tadesse W., Sanchez-Garcia M., Assefa S.G., Amri A., Bishaw Z., Ogbonnaya F.C., Baum M. Genetic gains in wheat breeding and its role in feeding the world. *Crop Breed Genet Genom*. 2019;1:e190005. doi 10.20900/cbgg20190005

Tanin M.J., Saini D.K., Kumar P., Gudi S., Sharma H., Kaur J.P., Abassy O., Bromand F., Sharma A. Iron biofortification in wheat: past, present, and future. *Curr Plant Biol*. 2024;38:100328. doi 10.1016/j.cpb.2024.100328

Tekin M., Emiralioğlu O., Yeken M.Z., Nadeem M.A., Çiftçi V., Baloch F.S. Wild relatives and their contributions to wheat breeding. In: Ancient Wheats. Springer International Publ., 2022;197-233. doi 10.1007/978-3-031-07285-7_9

Trots N.M., Bokova A.A. Influence of organomineral fertilizers on the accumulation of heavy metals in chernozem soils under conditions of the Middle Volga region. *Bulletin of Samara State Agricultural Academy.* 2024;9(1):81-88. doi 10.55170/1997-3225-2024-9-1-81-88 (in Russian)

Trots V.B., Akhmatov D.A., Trots N.M. Influence of fertilizers on accumulation of heavy metal in soil and phytomass of grain crops. *Zernovoe Khozyajstvo Rossii = Grain Econ Russ.* 2015;1:95-104 (in Russian)

Ugulu I., Ahmad K., Khan Z.I., Munir M., Wajid K., Bashir H. Effects of organic and chemical fertilizers on the growth, heavy metal/metalloid accumulation, and human health risk of wheat (*Triticum aestivum* L.). *Environ Sci Pollut Res Int.* 2021;28(10):12533-12545. doi 10.1007/s11356-020-11271-4

Virk P.S., Andersson M.S., Arcos J., Govindaraj M., Pfeiffer W.H. Transition from targeted breeding to mainstreaming of biofortification traits in crop improvement programs. *Front Plant Sci.* 2021;12:703990. doi 10.3389/fpls.2021.703990

Wang S., Wu W., Liu F., Liao R., Hu Y. Accumulation of heavy metals in soil-crop systems: a review for wheat and corn. *Environ Sci Pollut Res.* 2017;24(18):15209-15225. doi 10.1007/s11356-017-8909-5

Wessling-Resnick M. Excess iron: considerations related to development and early growth. *Am J Clin Nutr*. 2017;106:1600-1605. doi 10.3945/ajcn.117.155879

Wysocka K., Cacak-Pietrzak G., Sosulski T. Mineral concentration in spring wheat grain under organic, integrated, and conventional farming systems and their alterations during processing. *Plants*. 2025;14(7):1003. doi 10.3390/plants14071003

Xynias I.N., Mylonas I., Korpetis E.G., Ninou E., Tsaballa A., Avdikos I.D., Mavromatis A.G. Durum wheat breeding in the Mediterranean region: current status and future prospects. *Agronomy*. 2020;10(3):432. doi 10.3390/agronomy10030432

Zhao F.J., Su Y.H., Dunham S.J., Rakszegi M., Bedo Z., McGrath S.P., Shewry P.R. Variation in mineral micronutrient concentrations in grain of wheat lines of diverse origin. *J Cereal Sci.* 2009;49(2):290-295. doi 10.1016/j.jcs.2008.11.007

Zoroddu M.A., Aaseth J., Crisponi G., Medici S., Peana M., Nurchi V.M. The essential metals for humans: a brief overview. *J Inorg Biochem.* 2019;195:120-129. doi 10.1016/j.jinorgbio.2019.03.013

Conflict of interest. The authors declare no conflict of interest.
Received May 30, 2025. Revised June 25, 2025. Accepted June 30, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):798-811

doi 10.18699/vjgb-25-87

On the genetic structure and origin of the little ground squirrel *Spermophilus pygmaeus* (Pallas, 1778) in the North Caucasus

F.A. Tembotova, A.Kh. Amshokova, M.S. Gudova

Tembotov Institute of Ecology of Mountain Territories of the Russian Academy of Sciences, Nalchik, Russia
mpapieva@inbox.ru

Abstract. Little ground squirrel *Spermophilus pygmaeus* (Pallas, 1778) is a polytypic species of significant interest for the study of taxonomic diversity, genetic structure, gene flow and genetic diversity. Despite the long history of study, the taxonomy of representatives of the genus *Spermophilus* in the North Caucasus remains poorly developed. Among the unresolved issues are the phylogenetic relationships between the "mountain" and "plain" ground squirrels of the North Caucasus. An equally important aspect of the work is the study of the genetic diversity of little ground squirrel, given that the species is considered an integral component of steppe and desert ecosystems, providing their most important biocenotic functions. Based on the analysis of the 840 bp mtDNA *cytochrome b* gene fragment, new data on the genetic variability of *S. pygmaeus* from the eastern extremity of the Western Caucasus were obtained. Unlike previous studies that showed the so-called mountain ground squirrel to inhabit the Caucasus Mountains, this work identified two haplogroups of *S. pygmaeus* in the studied areas at an altitude of 1,400–1,700 m above sea level, one of which is close to the lowland (East Caucasian) and the other to the mountain (Central Caucasian) groups of the little ground squirrel. The genetic distance between the two haplogroups was 1.54 %. The different evolutionary ages of the three identified groups of *S. pygmaeus* in the North Caucasus (A1, A2, and B) are most likely associated with the multi-stage settlement of the studied area by the little ground squirrel. The results of molecular dating suggest that the western haplogroup penetrated as a continuous strip into the Central, Eastern Caucasus and the eastern extremity of the Western Caucasus through the Stavropol Upland and the Caspian Lowland less than 400 thousand years ago. As a result of the first wave of dispersal of the ground squirrel from the Russian Plain, the species became established in the eastern extremity of the Western Caucasus in the area of the village of Khasaut, and in the Eastern Caucasus – in the north of the Nogai Steppe (Sukhokumsk) and in the southern outskirts of the Caspian Lowland (Kar-Kar 1 Valley). The younger age of haplogroup A2 (less than 300 thousand years), also originating from the Eastern Caucasus (Khumtop, Zelenomorsk, Lvovsky 13, Kar-Kar 2), is most likely due to the re-colonization of the Caspian lowland by the ground squirrel, which was regularly flooded by the Caspian Sea in historic times. The absence of a continuous forest belt in the Central Caucasus, in particular in the Kabardino-Balkarian Republic, allowed *S. pygmaeus* to penetrate into the mountains later, less than 200 thousand years ago, through three gorges: Cherek, Baksan and Malkinsky. It is more likely that the species penetrated into the subalpics of the Western Caucasus (Khurzuk and Uchkulan) from the Central Caucasus, as evidenced by the same evolutionary age of animals of the Western (Uchkulan, Khurzuk) and Central Caucasus. Regarding the taxonomic status of the Caucasian mountain ground squirrel, we consider it premature to draw any conclusions, since not all areas of the Caucasus were covered by research.

Key words: *Spermophilus pygmaeus*; *cytochrome b* (*cyt b*); genetic diversity; Western Caucasus

For citation: Tembotova F.A., Amshokova A.Kh., Gudova M.S. On the genetic structure and origin of the little ground squirrel *Spermophilus pygmaeus* (Pallas, 1778) in the North Caucasus. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):798-811. doi 10.18699/vjgb-25-87

Funding. The work was carried out under the program of the State Assignment of the Tembotov Institute of Ecology of Mountain Territories of the Russian Academy of Sciences (FMEU-2023-0001 "Diversity (genetic, morphological, taxonomic) of vertebrates of the North Caucasus, the structure of their distribution in the region as a basis for long-term monitoring of natural and anthropogenic ecosystems").

К генетической структуре и происхождению малого суслика *Spermophilus pygmaeus* (Pallas, 1778) на Северном Кавказе

Ф.А. Темботова, А.Х. Амшокова, М.С. Гудова

Институт экологии горных территорий им. А.К. Темботова Российской академии наук, Нальчик, Россия
mpapieva@inbox.ru

Аннотация. Малый суслик *Spermophilus pygmaeus* (Pallas, 1778) – политипический вид, представляющий существенный интерес для изучения таксономического разнообразия, генетической структуры, потока генов и генетического разнообразия. Несмотря на длительную историю изучения, систематика представителей рода

Spermophilus на Северном Кавказе остается слабо разработанной. Неразрешенным вопросом остаются и филогенетические отношения между «горными» и «равнинными» сусликами Северного Кавказа. Не менее важным аспектом работы является оценка генетического разнообразия малого суслика, поскольку вид считается неотъемлемым компонентом степных и пустынных экосистем, обеспечивая их важнейшие биоценотические функции. На основании анализа фрагмента гена *cyt b* мтДНК длиной 840 п.н. получены новые данные по генетической изменчивости *Spermophilus pygmaeus* восточной оконечности Западного Кавказа. В отличие от ранее проведенных работ, показавших обитание в горах Кавказа так называемого горного суслика, в настоящей работе на исследованных территориях на высоте 1400–1700 м над ур. моря выявлены две гаплогруппы *S. pygmaeus*, одна из которых близка к равнинным (восточно-кавказским), а вторая – к горным (центрально-кавказским) группировкам малого суслика. Генетическая дистанция между двумя гаплогруппами составила 1.54 %. Разный эволюционный возраст трех выявленных группировок *S. pygmaeus* на Северном Кавказе (A1, A2 и B), скорее всего, связан с многоэтапным заселением исследуемой территории малым сусликом. Данные молекулярного датирования позволяют предположить, что западная гаплогруппа проникла сплошной полосой на Центральный, Восточный Кавказ и восточную оконечность Западного Кавказа через Ставропольскую возвышенность и Прикаспийскую низменность менее 400 тыс. лет назад. В процессе первой волны расселения суслика с Русской равнины вид закрепился в восточной оконечности Западного Кавказа в районе с. Хасаут, а на Восточном Кавказе – на севере Ногайской степи (Сухокумск) и в южных окраинах Прикаспийской низменности (долина Кар-Кар 1). Более молодой возраст гаплогруппы A2 (менее 300 тыс. лет), также происходящей с Восточного Кавказа (Хумтоп, Зеленоморск, Львовский 13, Кар-Кар 2), вероятнее всего, обусловлен повторным заселением сусликом Прикаспийской низменности, регулярно затапливаемой водами Каспия в историческое время. Отсутствие сплошного пояса лесов на Центральном Кавказе, в частности в Кабардино-Балкарии, позволило позже, менее 200 тыс. лет назад, проникнуть *S. pygmaeus* в горы по трем ущельям: Черекскому, Баксанскому и Малкинскому. Более вероятно, что в субальпику Западного Кавказа (Хурзука и Учкулана) вид проник уже с Центрального Кавказа, о чем свидетельствует один и тот же эволюционный возраст животных Западного (Учкулан, Хурзук) и Центрального Кавказа. Касательно таксономического статуса кавказского горного суслика считаем преждевременным делать какие-либо выводы, так как не все территории Кавказа были охвачены исследованиями.

Ключевые слова: *Spermophilus pygmaeus*; *cyt b*; генетическое разнообразие; Западный Кавказ

Introduction

The little ground squirrel *Spermophilus pygmaeus* (Pallas, 1778) is a polytypic species of considerable interest for the study of taxonomic diversity, genetic structure, gene flow, and genetic diversity. Despite the long history of study, the taxonomy of representatives of the genus *Spermophilus* in the North Caucasus remains poorly developed. The mountain ground squirrel was first collected on the northern slope of Elbrus in the subalpine meadow belt and was described by E. Menetries as an independent species (Menetries, 1832). Many researchers distinguished the mountain ground squirrel as an independent species (Brandt, 1843; Sviridenko, 1937; Vinogradov, Argiropulo, 1941; Mammalian Fauna…, 1963; Fauna of the USSR, 1965; Vorontsov, Lyapunova, 1969; Gromov, Baranova, 1981; Korablev, 1983; Harrison et al., 1993; Hoffmann, 1993; Gromov, Erbaeva, 1995; Tsvirka et al., 2003; Tsvirka, Korablev, 2014), while other authors believe that the Caucasian mountain ground squirrel is a subspecies of the little one (Satunin, 1907; Obolenskii, 1927; Ognev, 1947; Vereshchagin, 1959; Orlov et al., 1969; Ivanov, 1976; Ermakov et al., 2006; Nikol'skii et al., 2007). I.Ya. Pavlinov and A.A. Lisovsky (2012) distinguish *Spermophilus (pygmaeus) pygmaeus* (left-bank little ground squirrel relative to the Volga River), *Spermophilus (pygmaeus) planicola* (right-bank little ground squirrel) and *Spermophilus (pygmaeus) musicus* (Caucasian (mountain) little ground squirrel) as subspecies in the superspecies "*pygmaeus*". This species has attracted the attention of researchers for a long time, but there are very few works based on the analysis of the *cyt b* gene region of the little ground squirrel in the designated area (Harrison et al., 1993; Ermakov et al., 2023; Tembotova et al., 2024). Among the unresolved issues, as rightly noted by O.A. Ermakov and co-authors, are the phylogenetic relationships between the "mountain" and "plain" ground squirrels of the North Caucasus (Ermakov et al., 2018). Also, questions concerning the evolutionary history, the patterns of distribution of the plain and mountain forms of the little ground squirrel are still not fully clarified. The solution of the above-mentioned issues was hampered by the extremely uneven study of the territory with the involvement of insignificant samples in the analysis.

All populations of the little ground squirrel living west of the Volga to the lower Dnieper, as well as in Crimea and the Ciscaucasia, were assigned to the sister species *S. musicus* Ménétries, 1832, since it is the senior synonym applicable to the western lineage of *S. pygmaeus* sensu lato (Simonov et al., 2024). Based on the noted work, *S. musicus* is not only the mountain ground squirrel, but also all little ground squirrels of the right bank of the Volga. This raises the question of whether all right-bank little ground squirrels are genetically homogeneous. The emergence of this question is associated with the results we obtained earlier based on the analysis of the mtDNA *cytochrome b* gene fragment of ground squirrels of the Eastern and Central Caucasus (Tembotova et al., 2024).

The study revealed that two genetically distinct groups of little ground squirrels inhabit the territory of the Eastern Caucasus. In addition, a comparison of the Central Caucasian (mountain) and East Caucasian (plain) groups of little ground squirrels revealed a genetic distance of 1.34 % and an absence of identical haplotypes in the compared groups, which gene-

rally indicates the genetic heterogeneity of *S. pygmaeus* in the studied areas. Distances of the same order (1.29–1.72 %) were obtained between mountain little ground squirrels and ground squirrel populations from the right bank of the Volga River. At first glance, the obtained distances may seem insignificant, but they reach the lower limits of interspecific differences when compared with the distances obtained for representatives of the genus *Spermophilus*: a minimum of 1.4 % (between the species *S. major* and *S. selevini* (=*S. brevicauda*)) and a maximum of 10.7 % (between *S. dauricus* and *S. xanthoprymnus* (Simonov et al., 2024)). It is also not entirely clear whether all populations of the so-called Caucasian mountain ground squirrel inhabiting the mountainous territories of the Caucasus are genetically close to each other and differ equally from the lowland ones.

It should be noted that in most studies attempting to clarify the taxonomic status of the Caucasian mountain ground squirrel, the material was studied mainly only from the Central Caucasus, in particular from the vicinity of Elbrus (Harrison et al., 1993; Ermakov et al., 2006; Nikol'skii et al., 2007; Frisman et al., 2014; Tsvirka, Korablev, 2014), Baksan, Dzhily-Su, Shadzhatmaz gorges (Ermakov et al., 2023). The absence of both literary data obtained on the basis of analysis of the *cytochrome b* gene region of the little ground squirrel of the Western Caucasus and sequences deposited in the GenBank database determines the need to study samples of the little ground squirrel from this territory to determine the status of *S. pygmaeus* inhabiting mountainous areas.

Mitochondrial DNA is one of the most frequently used genetic markers in phylogeographic studies of vertebrates (Avise, 2000; Kholodova, 2009; Lukashov, 2009), which is due to such properties as maternal inheritance, an absence of the recombination process, a high rate of evolution compared to nuclear genes, a large number of copies, etc. *Cytochrome b* has proven itself to be informative and is successfully used in theriological studies at levels from generic to intraspecific (Bannikova, 2004; Abramson, 2007; Kholodova, 2009). In addition, this is the gene for which the most information is available in genetic databases.

The second, no less important, aspect of the work is the study of the genetic diversity of geographic samples of the little ground squirrel. It is known that ground squirrels are an integral component of steppe and desert ecosystems, providing their most important biocenotic functions, but since ancient times they have attracted attention mainly as agricultural pests and carriers of various diseases. Since the 1920s, the fight against ground squirrels as agricultural pests has been going at the state level. Grandiose extermination work against little ground squirrels was carried out in the arid landscapes of the former USSR, natural foci of plague. In the first half of the 20th century, the concept of complete elimination of foci of this infection by extermination of rodents – the carriers of the pathogen – was developed. In the North Caucasus and the North-West Caspian region, almost complete destruction of little ground squirrels was recommended (Kalabukhov, 1933; Pastukhov, 1959). The "recovery" of the natural focus began in accordance with a special program, which reached an unprecedented scale (cited by Shilova, 2011). As a result, the continuous clearing of the territory from the ground squirrel, plowing of the steppes, changes in the intensity of livestock grazing, the development of forest shelterbelts and artificial irrigation led to the destruction of gophers not only in the Caucasus, but also in many regions of Russia. According to S.A. Shilova (2011), a deep depression in the population of the right-bank little ground squirrel began in the south of Russia at the end of the last century and continues to this day. In Kabardino-Balkaria, agricultural development of plain and foothill territories led to an almost complete disappearance of little ground squirrel populations in these territories. In particular, foothill and plain populations of *S. pygmaeus* are listed in the Red Book of the Kabardino-Balkarian Republic (2018) with the status of "on the verge of extinction".

It should be noted that habitat transformation and fragmentation due to human activities resulted in a reduction in the areas suitable for little ground squirrel habitation, which led to an extremely high fragmentation of the species populations. At the same time, it is known that fragmentation and reduction of ranges often affect the genetic structure of wild animal populations, complicating the exchange of genes between different parts of the range, reducing the effective population size and leading to an increase in the level of inbreeding.

In connection with the above, the aim of the work was to study the genetic structure and genetic diversity of *S. pygmaeus* from the eastern end of the Western Caucasus based on the analysis of the mtDNA *cytochrome b* gene fragment. The results of this study were compared with the ones previously obtained (Ermakov et al., 2023; Tembotova et al., 2024) in order to assess the taxonomic diversity of *S. pygmaeus* in the North Caucasus.

Materials and methods

In this work, we used muscle tissue samples of *S. pygmaeus* from different geographical locations of the eastern end of the Western Caucasus: Karachay-Cherkess Republic (KCR): upper reaches of the Kuban River, in the vicinity of the villages of Khurzuk and Uchkulan; the Khasaut River tract, a tributary of the Malka River, in the vicinity of the village of Khasaut (Fig. 1). The animals were caught using No. 0 arc traps. The traps were set around the gophers' residential burrows (Karaseva, Telitsina, 1996).

The analyzed sample included 32 sequences of the mitochondrial *cytochrome b* gene (*cyt b*) of *S. pygmaeus* from the KCR, collected in the vicinity of the village of Khasaut and the villages of Khurzuk and Uchkulan (Table 1). In addition, haplotypes of the North Caucasus little ground squirrel from our previously published work (Tembotova et al., 2024) were also used to conduct a comparative analysis.

The remaining sequences of representatives of the genus *Spermophilus*, including the outgroup, were taken from the GenBank database (ncbi.nlm.nih.gov): *S. pygmaeus* – OP588846–OP588904 (Ermakov et al., 2023), AF157907, AF157910 (Harrison et al., 1993); *S. musicus* – AF157900, AF157904 (Harrison et al., 1993), *Spermophilus taurensis* Gündüz et al., 2007 – KY938064, KY938069, KY938073 (Gür, 2017); *Spermophilus citellus* Linnaeus, 1766 – AM691632–AM691640; *Spermophilus xanthoprymnus* Bennett, 1835 –

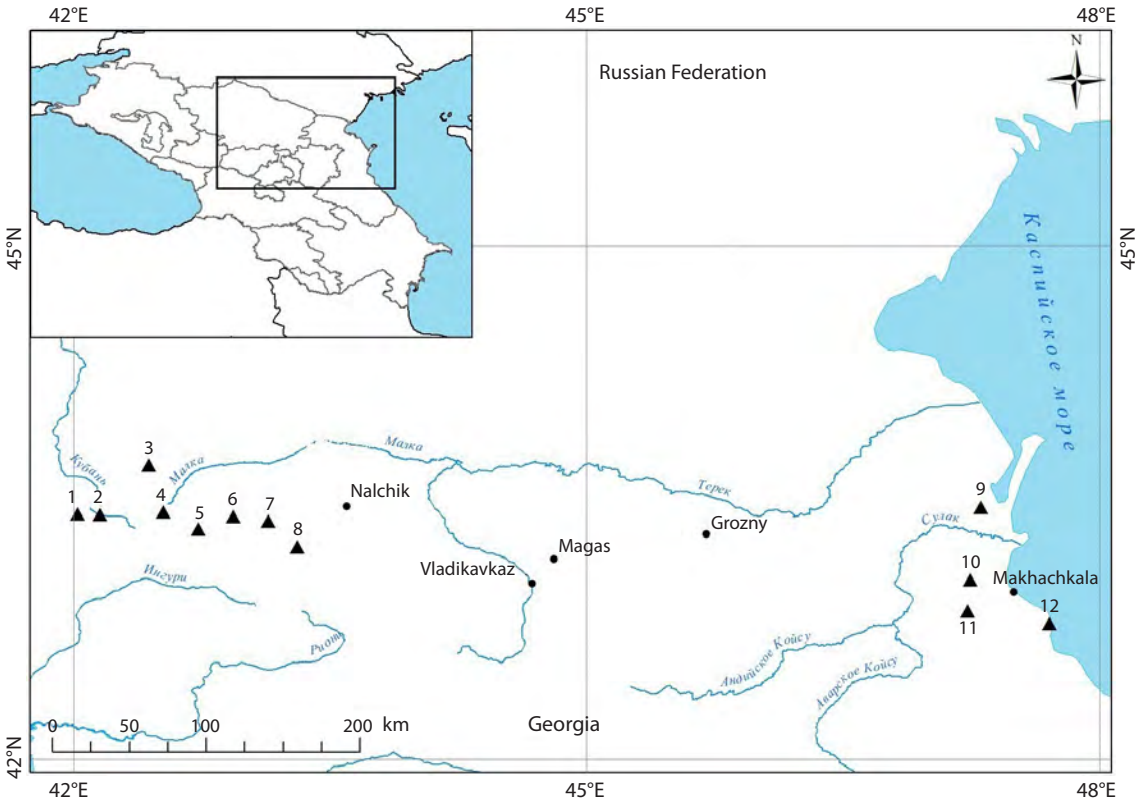


Fig. 1. Material origin card of *S. pygmaeus* in the North Caucasus.

Western Caucasus (KCR): 1 – Uchkulan, 2 – Khurzuk, 3 – Khasaut (new data); Central Caucasus (Kabardino-Balkarian Republic): 4 – Dzhily-Su, 5 – Elbrus, 6 – Tyrnyauz, 7 – Aktoprak, 8 – Bezengi; Eastern Caucasus (Republic of Dagestan): 9 – Lvovsky 13, 10 – Khumtop, 11 – Kar-Kar, 12 – Zelenomorsk (Tembotova et al., 2024).

AM691658–AM691663 (Gündüz et al., 2007) and AF157902, AF157909 (Orlov et al., 1969); *Marmota monax* Linnaeus, 1758 – AF157953 (Harrison et al., 1993).

DNA extraction was performed using the Diatom™ DNA Prep100 kit (Isogene Laboratory, Moscow) according to the manufacturer's protocol. DNA fragments were amplified using the MasterMix X5 kit (Dialat, Moscow). The following primers were used for the polymerase chain reaction: L14725 TGAAAAAYCATCGTTGT (Steppan et al., 1999) H15915 TCTTCATTTYWGGTTTACAAGAC (Harrison et al., 1993), with the PCR cycle parameters recommended in the first work. The resulting PCR products were purified by reprecipitation in 0.15 M sodium acetate solution, in 90 % ethanol, followed by washing with 70 % ethanol. The quality of the obtained PCR products was assessed by electrophoresis in 1.5 % agarose gel in the presence of ethidium bromide. Sequencing of nucleotide sequences was performed in both directions at Syntol (Moscow). Editing and alignment of the obtained sequences were performed using the BioEdit 7.0.9.0 program (Hall, 1999) using the Clustal W algorithm and edited manually.

Statistical data processing, including calculation of the number of polymorphic sites, the number of haplotypes, nucleotide and haplotype diversity, as well as the neutrality tests of Tajima (Tajima, 1989) and Fu (Fu, 1996), was performed in the Arlequin v.3.5 program (Excoffier, Lischer, 2010). In the same program, an analysis of the distribution of observed and expected values of pairwise nucleotide differences in mtDNA was carried out in accordance with the models of demographic (Rogers, Harpending, 1992) and spatial expansion (Ray et al., 2003). Weighted (net distance) intergroup genetic distances using the Kimura two-parameter model (K2P) (Kimura, 1980) were calculated in the Mega 6 program.

Median haplotype networks were constructed in the Network 4.6.1 program using the Median-Joining method (Bandelt et al., 1999) and then edited using the standard Paint package.

Phylogenetic analysis of nucleotide sequences using the Bayesian MCMC method was performed in MrBayes v3.2.6. (Ronquist, Huelsenbeck, 2003).

Divergence times were estimated in BEAST 1.10.4 (Suchard et al., 2018) using the following calibrations: 10.9 million years for the root divergence node of *Marmota* and other *Spermophilus* species (Yin et al., 2014), 5 million years for the divergence time between *S. xanthoprymnus* and *S. citellus* + *S. taurensis*, and 2.5 million years between *S. citellus* and *S. taurensis* (Gündüz et al., 2007). Data were analyzed using an uncorrelated lognormal relaxed molecular clock model. The most optimal model of nucleotide substitutions (HKY + I) was selected using the MEGA 6 software package. The length of Markov chains (MCMC, Markov Chain Monte Carlo) was set equal to 100 million generations with the selection of every thousandth state and a burn-in value of

Table 1. Characteristics of the studied material on *S. pygmaeus* from the eastern end of the Western Caucasus (KCR)

Number in GenBank (haplotype))	Samples	Collection locality	Collection locality	Identical to haplotypes from the GenBank
PV539552 3730Has	3730Has 137Has 3734Has 138Has	Khasaut	N 43.701667°, E 42.512304°, H = 1,776 m above sea level	
PV539556 3727Has	3727Has			
PV539557 3733Has	3733Has			
PV539558 3729Has	3729Has			
PV539559 3728Has	3728Has			
PV539560 141Has	141Has			
PV539561 135Has	135Has			
PV539579 3707Hurz	3707Hurz 3710Hurz	Khurzuk	N 43.415534°, E 42.162560°, H = 1,484 m above sea level	
PV539581 3653Hurz	3653Hurz			
PV539562 3709Hurz	3709Hurz 3706Hurz 3652Hurz 3651Hurz 3656Hurz 3650Hurz 3654Hurz 3659Hurz			
	3713Uchk 3712Uchk 3715Uchk 3711Uchk 3716Uchk 3714Uchk 3661Uchk 3663Uchk 3665Uchk	Uchkulan	N 43.455450°, E 42.090520°, H = 1,367 m above sea level.	
PV539582 3660Uchk	3660Uchk			
PV539583 3705Uchk	3705Uchk			AF157904, OP588865–OP588868, OP588903, OP588904

10 %. The convergence of the parameters was assessed based on achieving ESS (effective sample size) values >200 using Tracer 1.7 software (Rambaut et al., 2018). The divergence time of the *Spermophilus* dendrogram nodes was calculated for six variants of nucleotide substitutions per million years: 0.5, 0.9, 1.2, 2.4, 3.2 and 6.7 %.

Results

Based on the analysis of tissue samples of *S. pygmaeus* from the eastern end of the Western Caucasus, 32 nucleotide sequences of the *cytochrome b* gene fragment with a length of 840 bp were obtained. All of them were uploaded to the GenBank database under the numbers PV539552–PV539583. The analyzed sequences contained 24 variable sites, of which 14 were parsimony-informative.

The results of the phylogenetic analysis showed the same tree topology (Fig. 2) as in previous studies performed on this group and revealing the division of *S. pygmaeus* into western and eastern groups (Ermakov et al., 2023).

Previously, we demonstrated the isolated position of the Central Caucasian (mountain) and East Caucasian (plain) samples on the phylogenetic tree (Tembotova et al., 2024). The

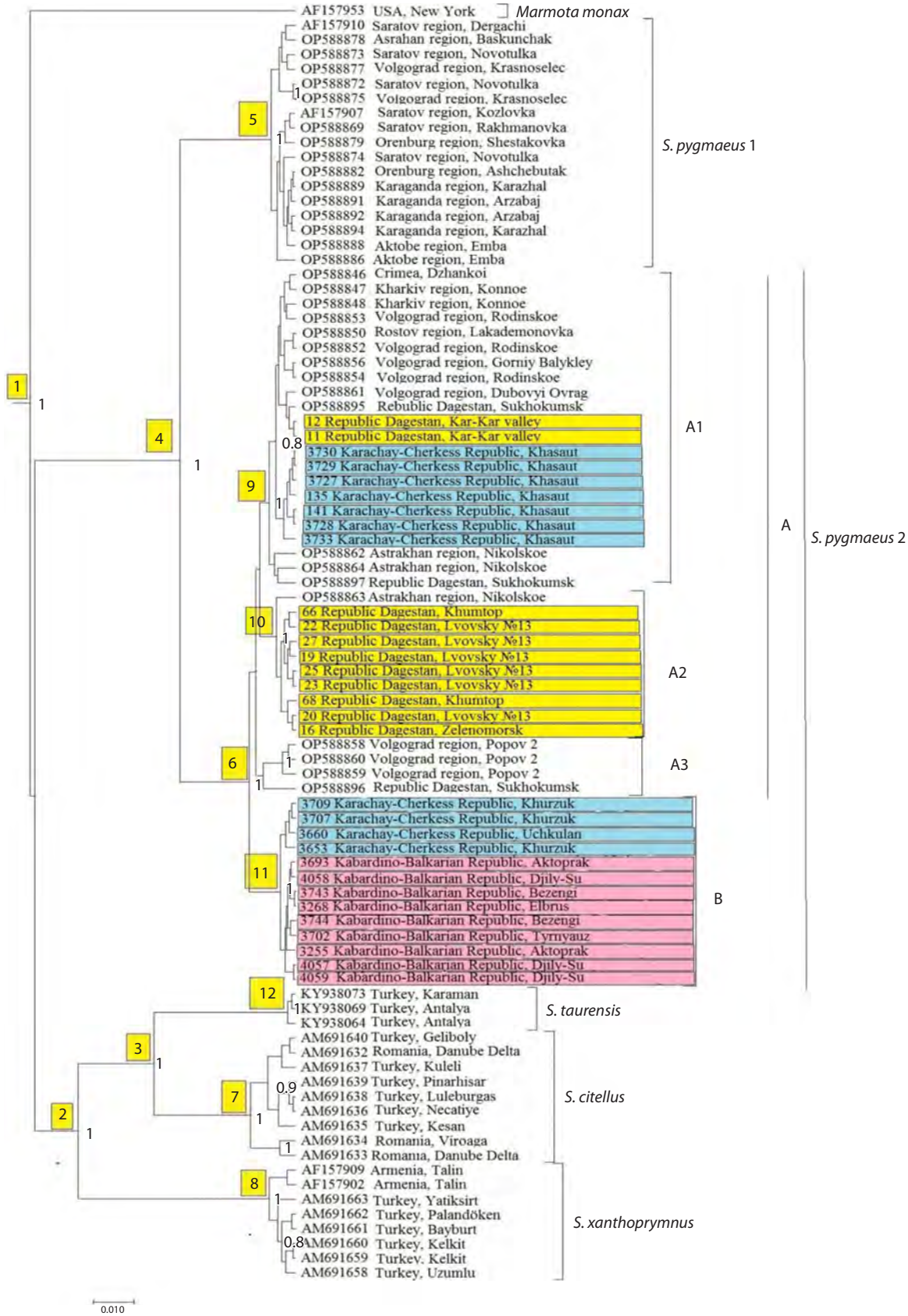


Fig. 2. Bayesian phylogenetic tree of representatives of the genus *Spermophilus* based on the analysis of the mtDNA *cyt b* gene (840 bp) (after (Tembotova et al., 2024) with additions).

The numbers in the branching nodes are the values of posterior probabilities (greater than 0.70), the numbers in the squares are the node numbers. The haplotypes of the little ground squirrel of the Republic of Dagestan are highlighted in yellow, those of the Karachay-Cherkess Republic are highlighted in blue, and those of the Kabardino-Balkarian Republic are highlighted in pink.

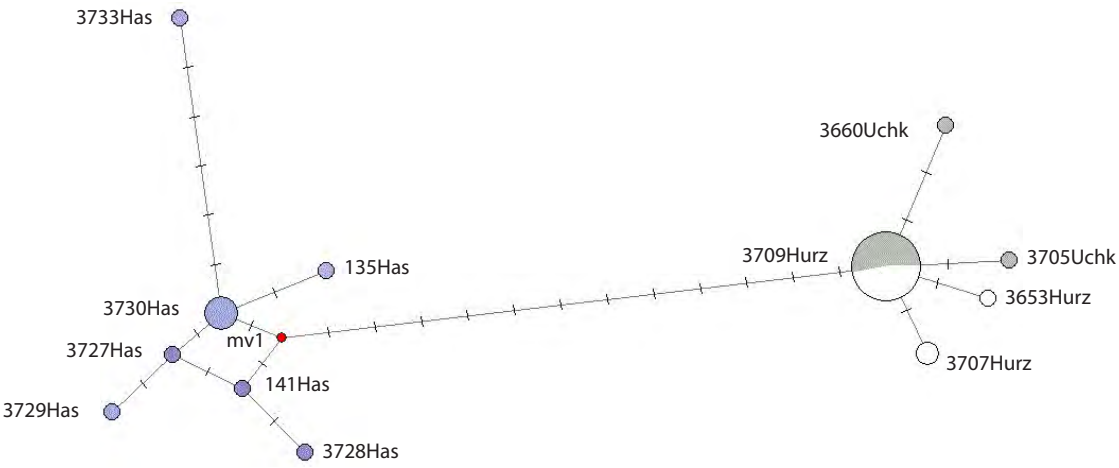


Fig. 3. Median network of mtDNA haplotypes of *S. pygmaeus*, constructed based on the analysis of the *cyt b* region (algorithm – median-joining, Network 4.6.1).

The size of the circle is proportional to the number of identical haplotypes. Haplotypes identified in the territory of the Karachay-Cherkess Republic: Hurz (vicinity of the village Khurzuk) – in white, Uchk (vicinity of the village Uchkulan) – in gray, Has (vicinity of the village Khasaut) – in lilac. The number of cross bars on the branches indicates the number of nucleotide substitutions; the "mv" mark denotes hypothetical haplotypes.

new haplotypes from the Karachay-Cherkess Republic analyzed in this work were split into two haplogroups. Thus, out of the three analyzed samples, the mitotypes of two samples (Uchkulan and Khurzuk) fell into haplogroup (B), formed by the haplotypes of the Central Caucasian animals. Unlike other mountain samples, little ground squirrels from the vicinity of the village of Khasaut are closer to the plain haplotypes and form one haplogroup A1 with them.

The median network of haplotypes also demonstrates the existing distribution of the haplotypes found in the territory of the Karachay-Cherkess Republic into two main haplogroups (Fig. 3). It is noteworthy that almost all mitotypes from the vicinity of the village of Khasaut form a separate compact haplogroup. The second haplogroup included mitotypes of the little ground squirrel from Uchkulan and Khurzuk. The genetic distance between the two haplogroups was 1.54 %.

The obtained 32 sequences from the KCR formed 12 haplotypes, of which nine were unique, and three were described in 2–17 individuals. As can be seen from Table 1 and Figure 3, the maximum number of unique haplotypes (60 %) is noted in the sample from the vicinity of the village of Khasaut. The most frequently occurring mitotype 3709Hurz was noted in eight individuals from the vicinity of the village of Khurzuk and nine individuals from the vicinity of the village of Uchkulan. For the 22 studied individuals from Uchkulan and Khurzuk, only five haplotypes were described, of which three are unique. It is possible that the loss of haplotypes is associated with a general decrease in numbers.

To clarify the clustering of the analyzed haplotypes, an additional median network was constructed (Fig. 4) with the inclusion in the analysis of sequences of the little ground squirrel previously obtained by us (Tembotova et al., 2024) and O.A. Ermakov et al. (2023).

Analysis of the median network demonstrates the division of the little ground squirrel haplotypes into two groups: the plain (A), which in turn is subdivided into three haplogroups (A1, A2, A3) and the mountain (B). As shown earlier, the sample from the vicinity of the village of Khasaut clusters together with mitotypes from haplogroup A1. Two haplotypes, 11Kar and 12Kar, from the Eastern Caucasus (Republic of Dagestan, Kar-Kar 1 Valley) also fell here. The genetic distance between the animals from the Kar-Kar 1 Valley and those from the vicinity of the village of Khasaut was only 0.36 %. The haplotypes of the animals from Uchkulan and Khurzuk, together with the Central Caucasian samples (Bezengi village, Aktopraksky pass, Tyrnyauz area, Irikchat gorge, Dzhily-Su tract), formed a separate haplogroup B. The most common haplotype 3709Uch from the Karachay-Cherkessia Republic differs from the haplotype 3743Bez, described in 39 individuals from different geographical points of the Central and Western Caucasus, by a single substitution. In addition, it is important to note that the haplotype 3705Uch turned out to be identical to the Central Caucasian haplotype 3743Bez. Haplogroup A2 is formed by the haplotypes of only the East Caucasian animals.

Analysis of genetic variability showed that little ground squirrel samples from the area near the villages of Uchkulan and Khurzuk are characterized by low values of haplotype (h) and nucleotide (π) diversity (Table 2). In the sample from the vicinity of the village of Khasaut, on the contrary, relatively high values of nucleotide and haplotype diversity are observed. Thus, the nucleotide diversity in this sample was 0.0028, and the haplotype diversity was 0.867. Since the sample from the vicinity of the village of Khasaut is clustered separately from the other samples from the Karachay-Cherkess Republic, the parameters of genetic variability of the populations were calculated only for two combined samples – Uchkulan and Khurzuk. As a result, for the combined sample (n = 22), the haplotype diversity was 0.407 ± 0.128, the nucleotide diversity was 0.0006 ± 0.0006.

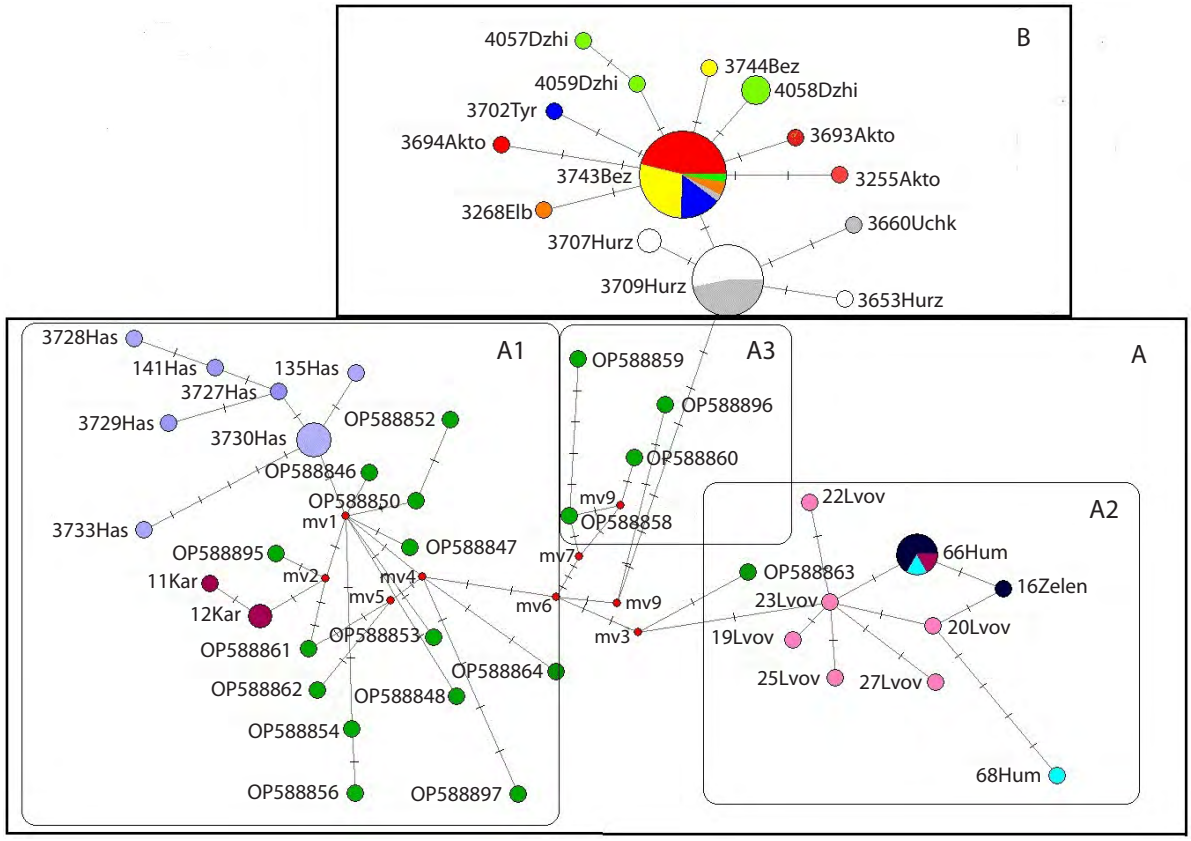


Fig. 4. Median network of mtDNA haplotypes of *S. pygmaeus*, constructed on the basis of the analysis of the *cyt b* region (algorithm – median-joining, Network 4.6.1).

The size of the circle is proportional to the number of identical haplotypes. Haplotypes identified in the Republic of Dagestan: Kar (Kar-Kar Valley) – are shown in brown, Zelen (vicinity of the village Zelenomorsk) – in black, Hum (vicinity of the village Khumtop) – in blue, Lvov (vicinity of the village Lvovsky 13) – in pink; in Kabardino-Balkarian Republic: Tyr (vicinity of the town Tyrnyauz) – in blue, Bez (vicinity of the village Bezengi) – in yellow, Akto (Aktopraksky Pass) – in red, Elb (vicinity of the village Elbrus) – in orange, Dzhi (Dzhily-Su tract) – in light green; in Karachay-Cherkess Republic: Hurz (vicinity of the village Khurzuk) – in white, Uchk (vicinity of the village Uchkulan) – in gray, Has (vicinity of the village Khasaut) – in lilac. The number of crossbars on the branches indicates the number of nucleotide substitutions; the "mv" mark indicates hypothetical haplotypes.

Table 2. Indicators of haplotype (h) and nucleotide (π) diversity and the values of the Tajima's and Fu's tests of *S. pygmaeus* from Karachay-Cherkess Republic

Geographic groups (sample size)	*N*	$\pi \pm$ S.E.	h ± S.E.	Tajima's D	Fu's F
Khasaut (*n* = 10)	7	0.0028 ± 0.0019	0.867 ± 0.107	**–1.765**	**–2.756**
Khurzuk (*n* = 11)	3	0.0006 ± 0.0006	0.473 ± 0.162	–0.778	–0.659
Uchkulan (*n* = 11)	3	0.0004 ± 0.0005	0.346 ± 0.172	**–1.430**	**–1.246**
For *Spermophilus* KCR without Khasaut (*n* = 22) overall	5	0.0006 ± 0.0006	0.407 ± 0.128	**–1.667**	**–2.662**

Note. *N* – number of haplotypes; S.E. – standard error. Statistically significant test values are highlighted in bold.

The values of the Tajima's D and Fu's Fs tests in all three samples were negative (Table 2), while the Tajima's and Fu's tests were significant for two samples – Uchkulan and Khasaut. In the combined sample (Uchkulan+Khurzuk), the values of the Tajima's and Fu's tests were also negative and statistically significant.

Of the three noted little ground squirrel haplogroups (A1, A2 and B), A1 and B are the most genetically distinct, with a genetic distance of 1.46 %. Nearly the same distance (1.41 %) was obtained between haplogroups A2 and B. And finally, the minimum distance was obtained when comparing groups A1 and A2 (0.74 %). Regarding the genetic distances obtained be-

Table 3. Genetic distances between geographical samples of *S. pygmaeus* from the North Caucasus (mtDNA *cytochrome b* gene region)

Samples	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Aktoprak		0.004	0.000	0.000	0.001	0.001	0.000	0.004	0.004	0.004	0.004	0.004	0.000
2. Khasaut	0.015		0.004	0.004	0.004	0.004	0.004	0.003	0.003	0.003	0.002	0.003	0.004
3. Bezengi	0.000	0.017		0.000	0.001	0.001	0.000	0.004	0.004	0.004	0.005	0.004	0.000
4. Elbrus	0.000	0.017	0.000		0.001	0.001	0.000	0.004	0.004	0.004	0.005	0.004	0.000
5. Khurzuk	0.001	0.015	0.001	0.001		0.000	0.001	0.004	0.004	0.004	0.004	0.004	0.001
6. Uchkulan	0.001	0.015	0.001	0.001	0.000		0.001	0.004	0.004	0.004	0.004	0.004	0.001
7. Djily-Su	0.000	0.017	0.000	0.000	0.002	0.001		0.004	0.004	0.004	0.005	0.004	0.000
8. Khumtop	0.012	0.009	0.013	0.013	0.012	0.012	0.014		0.000	0.000	0.004	0.000	0.004
9. Lvovsky13	0.014	0.008	0.015	0.015	0.014	0.014	0.015	0.000		0.001	0.004	0.001	0.004
10. Zelenomorsk	0.015	0.010	0.017	0.017	0.015	0.015	0.017	0.000	0.001		0.004	0.000	0.004
11. Karkar 1	0.017	0.004	0.018	0.018	0.017	0.017	0.019	0.010	0.010	0.011		0.004	0.005
12. Karkar 2	0.016	0.010	0.017	0.017	0.016	0.016	0.017	0.000	0.001	0.000	0.011		0.004
13. Tyrnyauz	0.000	0.017	0.000	0.000	0.001	0.001	0.000	0.013	0.015	0.017	0.018	0.017	

Note. Below the diagonal are the values of intergroup distances, above the diagonal are the corresponding values of the standard error.

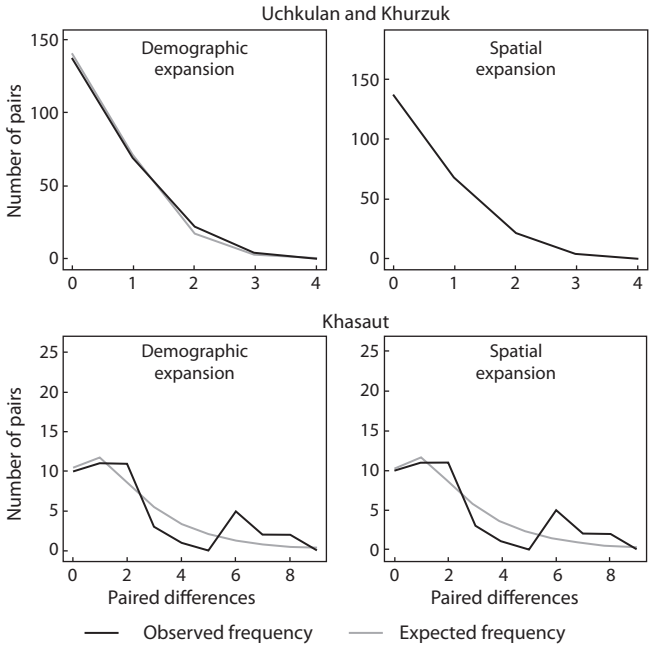


Fig. 5. Histograms of the distribution of paired differences (mismatch distribution) of *S. pygmaeus*: vicinity of the villages of Uchkulan+Khurzuk and vicinity of the village of Khasaut (demographic expansion and spatial expansion).

tween the three studied samples from the Karachay-Cherkess Republic, it should be noted that when comparing little ground squirrels from Uchkulan and Khurzuk, the obtained distance was zero. And the sample from Khasaut was equally different from the *S. pygmaeus* groups from Uchkulan and Khurzuk with a distance of 1.53 %.

Table 3 shows the genetic distances between the geographic samples of *S. pygmaeus* from the North Caucasus. As can be seen, the sample from the vicinity of the village of Khasaut differs from all other Central Caucasian samples (Aktoprak, Bezengi, Irikchat, Tyrnyauz, Dzhily-Su) with distances of 1.5–1.7 %, and from the East Caucasian samples, with distances of 0.4–1 %. When comparing two West Caucasian samples (Uchkulan, Khurzuk) with the Central Caucasian samples, the distances were only 0–0.2 %, and when they were similarly compared with the East Caucasian samples, the minimum distance was 1.2 %, and the maximum was 1.7 %.

The analysis of the frequency distribution of paired nucleotide differences between haplotypes (Fig. 5) was also carried out for two samples: from the vicinity of the village of Khasaut and the combined Uchkulan+Khurzuk sample. The combined sample shows a unimodal distribution pattern, close to expectations for a growing population, which may probably indicate a recent demographic expansion or spatial expansion (less than 200 thousand years ago) after a decline in population. The analysis of the distribution of the number of nucleotide substitutions in the sample from the vicinity of the village of Khasaut revealed multimodality, which probably indicates the presence of two or more subpopulations.

As noted earlier (Tembotova et al., 2024), the results of molecular dating were obtained based on three calibration points: 10.9 million years for the root node of divergence of *Marmota* and other *Spermophilus* species (Yin et al., 2014), 5 million years for the divergence time between *S. xanthoprymnus* and *S. citellus* + *S. taurensis*, and 2.5 million years between *S. citellus* and *S. taurensis* (Gündüz et al., 2007). The

Table 4. Divergence times (in million years) of *Spermophilus* taxa and individual clusters with six evolutionary rate variants

Node number on the tree	Evolution rate					
	0.5 %	0.9 %	1.2 %	2.4 %	3.1 %	6.7 %
1. *Marmota / Spermophilus*	8.479	7.625	7.351	6.491	6.244	5.532
2. *S. xanthoprymnus / S. taurensis+S. citellus*	4.708	4.270	4.135	3.676	3.536	3.254
3. *S. taurensis / S. citellus*	2.624	2.410	2.340	2.109	2.037	1.829
4. *S. pygmaeus* 1("eastern" group) / *S. pygmaeus* 2 ("western" group)	2.287	2.014	1.923	1.638	1.545	1.224
5. *S. pygmaeus* 1 ("eastern" group)	0.349	0.309	0.295	0.252	0.240	0.194
6. *S. pygmaeus* 2 ("western" group)	0.735	0.646	0.617	0.526	0.498	0.400
7. *S. citellus*	0.740	0.657	0.629	0.546	0.516	0.421
8. *S. xanthoprymnus*	0.415	0.366	0.351	0.302	0.286	0.234
9. *S. pygmaeus* (haplogroup A1)	0.369	0.325	0.310	0.266	0.252	0.203
10. *S. pygmaeus* (haplogroup A2)	0.262	0.230	0.221	0.189	0.179	0.146
11. *S. pygmaeus* (haplogroup B)	0.182	0.161	0.155	0.133	0.127	0.103
12. *S. taurensis*	0.108	0.096	0.092	0.081	0.077	0.063

inclusion of additional samples from the Karachay-Cherkess Republic in the analysis did not significantly affect the previously obtained results (Tembotova et al., 2024), and the age of many nodes remained almost the same (Table 4).

As for the samples from the Karachay-Cherkess Republic, the evolutionary age of the little ground squirrel group from Khasaut together with some haplotypes from Dagestan (11, 12, Kar-Kar valley, OP588895, OP588897, Sukhokumsk), as well as Crimea, Kharkov, Volgograd, Rostov, Astrakhan regions was 369 thousand years (95 % HPD: 0.217–0.538 million years) (node 9) for the model calculated for the mutation rate of 0.5 % per million years. The remaining haplotypes of animals from the Karachay-Cherkess Republic (Uchkulan, Khurzuk) fell into the cluster (B) formed by mitotypes of the Central Caucasian animals. The age of this cluster was 182 thousand years (95 % HPD: 0.080–0.300 million years) (node 11).

Discussion

Based on the obtained 32 sequences, 12 haplotypes distributed between two haplogroups (A and B) were described. Phylogenetic analysis showed that the discovered haplotypes were included in the previously described mountain (Central Caucasian) and lowland (East Caucasian) groups. As can be seen from the median network of haplotypes, the sample from the vicinity of the village of Khasaut is isolated from other mountainous West and Central Caucasian samples and is closest to the East Caucasian lowland samples of the little ground squirrel. Genetic distances obtained by comparing three samples from the Karachay-Cherkess Republic confirm the genetic isolation of the sample from the vicinity of the village of Khasaut. This sample differs from the other two (Uchkulan, Khurzuk) with a distance of 1.54 %. Thus, the results of the conducted study allow us to conclude that in the studied territories (Uchkulan, Khurzuk, Khasaut), the *S. pygmaeus* species is not homogeneous and is represented by two genetically different groups of the little ground squirrel. Moreover, one of the groups is closer to the East Caucasian (plain), and the other to the Central Caucasian (mountain) samples. The village of Khasaut is located on the left bank of the river, just below the confluence of the Bermamyt River, 6 km from the summit of Bolshoy Bermamyt. According to the data of A.I. Dyatlov et al. (1980), in the vicinity of Mount Bermamyt, mountain little ground squirrels penetrated north of the Rocky Ridge in several places (five settlements), which is not observed in other parts of the range. This is the only place in the gap zone where the rivers do not form barriers between the populations in the mountains and on the plains.

As can be seen from Table 2, of the three studied groups of *S. pygmaeus* from the Karachay-Cherkess Republic, the Uchkulan and Khurzuk samples are characterized by low values of the haplotype ($h = 0.346–0.473$) and nucleotide ($\pi = 0.0004–0.0006$) diversity indicators. Similar results were obtained earlier for the Central Caucasian samples (Kabardino-Balkarian Republic) of the little ground squirrel, originating from an altitude of 1,200–1,500 m above sea level (Tembotova et al., 2024).

Low genetic diversity in the Elbrus sample was also found by O.A. Ermakov et al. (2023) (haplotype diversity – 0.333 ± 0.215, and nucleotide diversity – 0.03 %). For comparison, we note that for the little ground squirrel groups from the western and eastern lines, the values of the haplotype diversity index varied from 0.859 to 0.964, and the level of nucleotide diversity π varied from 0.17 to 0.76 %, which is almost six or more times higher than in the Elbrus sample

(Ermakov et al., 2023). Thus, the results of both the present and previous studies show that most mountain samples of the little ground squirrel in the Western and Central Caucasus (with the exception of the high-mountain Dzhily-Su gorge) are characterized by a low level of genetic diversity (Tembotova et al., 2024). Low values of h and π may be the result of a serious decline in numbers over a long period of time (bottleneck effect) (Kholodova, 2006; Abramson, 2007). It is possible that mountain populations of the little ground squirrel have repeatedly experienced a decline in numbers. Low genetic diversity can lead to a decrease in the adaptive capacity of individuals and populations and increase the risk of their extinction (Gitzendanner, Soltis, 2000; Willi et al., 2006). Relatively high values of the noted indicators were revealed in the sample from the vicinity of the village of Khasaut. Thus, the haplotype diversity was almost two times higher, and the nucleotide diversity was five or more times higher than in the other two samples (Uchkulan, Khurzuk) of the little ground squirrel. The sample from the vicinity of the village of Khasaut is closer in genetic diversity to the previously studied plain samples from the southern edge of the Caspian Lowland of the Eastern Caucasus than to the mountain ones. Such a ratio of genetic diversity indices (high h and π) is characteristic not only of populations that have had a high population size for a long time, but also of those formed as a result of the unification of previously isolated and genetically heterogeneous groups (Rogers, Harpending, 1992).

The significantly negative values of the Tajima's test observed in almost all of the studied *S. pygmaeus* samples may indicate a recent population expansion after a decline in numbers (a bottleneck).

The histograms show (Fig. 5) that in the sample from the vicinity of the village of Khasaut, the distribution of nucleotide differences has a multimodal nature, which does not correspond to the expected distribution. Discrepancies between the expected and observed distributions indicate high heterogeneity of the studied sample. In the combined sample (Uchkulan and Khurzuk), the two curves show good agreement and have a unimodal distribution (Fig. 5).

The age of the so-called lowland group (haplogroup A1), which included the mitotypes of the *S. pygmaeus* from Khasaut, is less than 400 thousand years. The age of clade A2, represented by the ground squirrels of the Eastern Caucasus, is less than 300 thousand years. The group of *S. pygmaeus* of the Western (Uchkulan, Khurzuk) and Central Caucasus, which form clade B, is phylogenetically younger. Its age is less than 200 thousand years. The calculated age does not contradict the data and opinions of other authors, who believed that the ancestors of modern mountain ground squirrels penetrated into the highlands from lowland areas at different times (Tsvirka, Korablev, 2014). Considering the statements of many researchers (Sviridenko, 1927; Ioff, 1936; Varshavskii, 1963) that in the new history of the range of the little ground squirrel we have to deal not with the primary, but essentially with the repeated dispersal of this rodent, it is likely that the different evolutionary ages of the three identified haplogroups of *S. pygmaeus* are associated with the multi-stage dispersal of the species across the study area.

The results of molecular dating suggest that the western haplogroup of the little ground squirrel penetrated in a continuous strip into the Central, Eastern Caucasus and the eastern end of the Western Caucasus through the Stavropol Upland and the Caspian Lowland less than 400 thousand years ago.

It can also be assumed that as a result of the first stage of settlement, the little ground squirrel became established at the eastern end of the Western Caucasus in the Khasaut area, as well as on the plain and in the foothills of Kabardino-Balkaria, where stable ground squirrel populations existed until 1990 (Tembotov et al., 1969; Tembotova, Kononenko 2017), which have not been registered in the Kabardino-Balkarian Republic since the end of the 20th century. Less than 200 thousand years ago, as a result of settlement, the species rose to the mountains to an altitude of 2,000 m above sea level and higher along the Baksan, Malkinsky and Chereksky gorges. Apparently, it penetrated the border territories of Karachay-Cherkessia along the subalpine belt, as evidenced by the same evolutionary age of animals of the Western (Uchkulan, Khurzuk) and Central Caucasus.

In the Eastern Caucasus, as a result of the first wave of dispersal of the ground squirrel from the Russian Plain, the species was established in the north of the Nogai Steppe (Sukhokumsk) and in the southern outskirts of the Caspian Lowland (Kar-Kar Valley). It is quite clear that the penetration of the little ground squirrel into the southern regions of the Caspian Lowland occurred through the entire Caspian Lowland, based on which it could be assumed that the mitotypes of the species throughout its territory would be of the same evolutionary age. However, given that the Caspian Sea changed its outlines for a very long time in geological time, the lowland was regularly flooded and then freed from water; it is quite understandable that a stable population did not exist there. From the above, it follows that the Caspian Lowland in the areas of Khumtop, Lvov and Zelenomorsk was re-populated by the ground squirrel after 100 thousand years and it is more likely that the settlement came from the Russian Plain.

Conclusion

The ground squirrel penetrated into the North Caucasus from the western part of a vast range, covering the plain of Eastern Europe, northern Crimea, the Ciscaucasia and the northern parts of Central Asia (Vereshchagin, 1959). The age of the western haplogroup is about 800 thousand years. The spread to the northern parts of Central Asia and the North Caucasus most likely occurred in parallel, since the age of the eastern haplogroup *S. pygmaeus* 1 and the oldest age of the haplogroups that penetrated into the North Caucasus are close – within 350–400 thousand years for *S. pygmaeus* 1 and haplogroup A1 (Fig. 2). At the same time, the ground squirrel penetrated into the North Caucasus in a continuous strip to the Central, Eastern Caucasus and the eastern end of the Western Caucasus through the Stavropol Upland and the Caspian Lowland.

As a result of the first wave of settlement, the little ground squirrel became established in the Western Caucasus in the

Khasaut area, and in the Central Caucasus, on the plain and in the foothills of Kabardino-Balkaria, where stable populations of ground squirrels existed until 1990 (Tembotov et al., 1969; Tembotova, Kononenko 2017), which have not been recorded in the Kabardino-Balkarian Republic since the end of the 20th century. The absence of a continuous forest belt in the Central Caucasus, in particular in the Kabardino-Balkarian Republic, allowed *S. pygmaeus* to penetrate into the mountains later, less than 200 thousand years ago, along three gorges: Cherek, Baksan and Malkinsky. It is more likely that the species penetrated into the subalpics of the Western Caucasus (Khurzuk and Uchkulan) from the Central Caucasus.

The population in Khasaut is probably a genetic isolate, which is confirmed by genetic distances (within 1.54–1.69 %) (Table 3) with animals from neighboring areas of the Karachay-Cherkess Republic (Uchkulan, Khurzuk) and the Kabardino-Balkarian Republic (Bezengi, Aktoprak, Dzhilysu, Irikchat, Tyrnyauz); however, further research is needed.

The first wave of the gopher's settlement in the Eastern Caucasus (in Dagestan) has survived to this day in the north of the Nogai steppe in the Sukhokumsk region, and in the south of the Caspian lowland in the Kar-Kar1 valley, as evidenced by the evolutionary age of haplogroup A1. The younger age of haplogroup A2 (less than 300 thousand years), also originating from the Eastern Caucasus (Khumtop, Zelenomorsk, Lvovsky 13, Kar-Kar 2), is most likely due to the repeated settlement of the Caspian lowland, regularly flooded by the Caspian waters in historical times. This is also evidenced by the genetic distance (0.76–1.1 %, Table 3) between animals from the Kar-Kar 1 valley and the central regions of the Caspian lowland (Khumtop, Zelenomorsk, Lvovsky 13), which gives reason to believe that the settlement apparently originated from the Russian Plain, and that there is a weak gene flow between these populations, which is probably associated with the low mobility of the species, seasonal movements of its young over short distances (maximum 5 km) (Naumov, 2010), and low numbers.

Regarding the taxonomic status of the Caucasian mountain ground squirrel, we consider it premature to draw any conclusions, since not all territories of the Caucasus were covered by the studies. Nevertheless, the results obtained both in the present study and in the previously conducted one (Tembotova et al., 2024) suggest that the genetic distances (1.33–1.67 %) obtained between the lowland and mountain samples of the little ground squirrel of the North Caucasus correspond only to the level of intraspecific differences, according to the gradation given for the genus *Spermophillus* by (Baker, Brandley, 2006). Probably, one can agree with the opinion of N.N. Vorontsov and E.A. Lyapunova (1969) that *S. musicus* is a derivative of *S. pygmaeus*, made on the basis of karyological studies. M.V. Tsvirka and V.P. Korablev (2014) noted that significant transformations of the karyotype of the mountain ground squirrel occurred after it had already settled in mountainous regions; over time, the characteristics that emerged were fixed, leading to the stable isolation of the Caucasian mountain ground squirrel from the lowland populations of the little one. The genetic differentiation and structuring that we observed in the *S. pygmaeus* species in the conditions of the North Caucasus are probably also due to the geographical isolation of the lowland and mountain populations, which caused the emergence of local adaptations to habitat conditions as a result of a reduction in numbers and fragmentation of the range, which is still observed today.

References

Abramson N.I. Phylogeography: results, issues and perspectives. *Informatsionnyy Vestnik VOGiS = The Herald of Vavilov Society for Geneticists and Breeders*. 2007;11(2):307-331 (in Russian)

Avise J.C. Phylogeography: the history and formation of species. Cambridge, Mass: Harvard Univ. Press, 2000. doi 10.2307/j.ctv1nzfgj7

Baker R.J., Brandley R.D. Speciation in mammals and the genetic species concept. *J Mammalogy*. 2006;87(4):643-662. doi 10.1644/06-MAMM-F-038R2.1

Bandelt H.J., Forster P., Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16(1):37-48. doi 10.1093/oxfordjournals.molbev.a026036

Bannikova A.A. Molecular markers and modern phylogenetics of mammals. *Zhurnal Obshchey Biologii = Journal of General Biology*. 2004;65(4):278-305 (in Russian)

Brandt I.F. Observations sur les differentes especes de sousliks de Russie, suivies de remarques sur l'arrangement et la distribution geographique du genre Spermophilus, anse que sur la classification de la familie des ecureuils (Sciurina) en general. *Bulletin Scientifique l'Academie Imperiale des Sciences de Saint-Petersbourg*. 1843;2:357-381

Dyatlov A.I., Petrov P.A., Golubev P.D., Trufanov G.V. On the structure of the range of small gophers (*Citellus pygmaeus* Pall., 1778) in the Elbrus region. *Ekologiya = Ecology*. 1980;5:77-83 (in Russian)

Ermakov O.A., Titov S.V., Savinetskii A.B., Surin V.L., Zborovsky S.S., Lyapunova E.A., Brandler O.V., Formozov N.A. Molecular-genetic and paleoecological arguments for conspecificity of little (*Spermophilus pygmaeus*) and Caucasian mountain (*S. musicus*) ground squirrels. *Zoologičeskij Zhurnal = Zoological Journal* (*Moscow*). 2006;85(12):1474-1483 (in Russian)

Ermakov O.A., Simonov E.P., Surin V.L., Titov S.V. Intraspecific polymorphism of the mitochondrial DNA control region and phylogeography of little ground squirrel (*Spermophilus pygmaeus*, Sciuridae, Rodentia). *Russ J Genet*. 2018;54(11):1332-1341. doi 10.1134/S1022795418110042

Ermakov O.A., Brandler O.V., Ivanov A.Yu., Ivanova A.D., Kesyan A.A., Khalidov A.Kh., Lotiev K.Yu., Lukonina S.A., Tsapko V., Titov S.V. Riverine barriers and geographic variation in little ground squirrel (*Spermophilus pygmaeus*, Sciuridae, Rodentia) based on mitochondrial cytochrome *b* gene sequences. *Russ J Theriol*. 2023:22(1):24-31. doi 10.15298/rusjtheriol.22.1.03

Excoffier L., Lischer H. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resources*. 2010;10(3):564-567. doi 10.1111/j.1755-0998.2010.02847.x

Fauna of the USSR. Mammals. Vol. 3. Ground Squirrels (Marmotinae). Moscow; Leningrad: Nauka Publ., 1965 (in Russian)

Frisman L.V., Korablev V.P., Tsvirka M.V., Brandler O.V., Lyapunova E.A. Expeditions of the 1990s as a contribution to research of genetic differentiation of Palaearctic ground squirrels. *Zoologičeskij Zhurnal = Zoological Journal* (*Moscow*). 2014;93(7):939-950. doi 10.7868/S0044513414070071 (in Russian)

Fu Y. New statistical test of neutrality for DNA sample from a population. *Genetics*. 1996;143(1):557-570. doi 10.1093/genetics/143.1.557

Gitzendanner M.A., Soltis P.S. Patterns of genetic variation in rare and widespread congeners. *Am J Bot*. 2000;87(6):783-792. doi 10.2307/2656886

Gromov I.M., Baranova G.I. Catalogue of Mammals of the USSR (Pliocene–Recent). Leningrad: Nauka Publ., 1981 (in Russian)

Gromov I.M., Erbaeva M.A. Mammals of the Fauna of Russia and Adjacent Territories. Hares and Rodents. St. Petersburg, 1995 (in Russian)

Gündüz İ., Jaarola M., Tez C., Yeniyurt C., Polly P.D., Searle J.B. Multigenic and morphometric differentiation of ground squirrels (*Spermophilus*, Scuiridae, Rodentia) in Turkey, with a description of a new species. *Mol Phylogenet Evol*. 2007;43(3):916-935. doi 10.1016/j.ympev.2007.02.021

Gür H., Perktaş U., Gür M.K. Do climate-driven altitudinal range shifts explain the intraspecific diversification of a narrow ranging montane mammal, Taurus ground squirrels? *Mamm Res*. 2017;63(2):197-211. doi 10.1007/s13364-017-0347-8

Hall T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999;41:95-98

Harrison R.G., Bogdanowicz S.M., Hoffmann R.S., Yensen E., Sherman P.W. Phylogeny and evolutionary history of the ground squirrels (Rodentia: Marmotinae). *J Mamm Evol*. 1993;10(3):249-276. doi 10.1023/B:JOMM.0000015105.96065.f0

Hoffmann R.S. Order Lagomorpha. In: Mammal Species of the World. Washington–London: Smithsonian Institution Press, 1993;807-827

Ioff I.G. On the geographical distribution of gopher fleas in connection with the history of settlement of gophers. In: Parasitological Collection of the Zoological Institute of the USSR Academy of Sciences. Leningrad: AN USSR Publ., 1936;6:313-361 (in Russian)

Ivanov I.V. Small ground squirrel of the North Caucasus. In: Fauna, Ecology and Protection of Animals of the North Caucasus. Iss. 3. Nalchik, 1976;36-88 (in Russian)

Kalabukhov N.I. Ground squirrel population density in plague regions of the North Caucasus region and the possibility of total extermination of ground squirrels. *Works on Plant Protection*. 1933;4(2): 65-86 (in Russian)

Karaseva E.V., Telitsina A.Yu. The Methods of Studying Rodents in the Wild Nature. Moscow: Nauka Publ., 1996 (in Russian)

Kholodova M.V. Use of modern and ancient DNA for studying ecosystem dynamics. In: Dynamics of Modern Ecosystems in the Holocene. Moscow: Sci. Publ. House KMK, 2006;261-266 (in Russian)

Kholodova M.V. Comparative phylogeography: molecular methods, ecological interpretation. *Mol Biol*. 2009;43(5):847-854. doi 10.1134/S002689330905015X

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

Korablev V.P. Cytogenetic differences between the mountain and little ground squirrels. In: Population Variability of Species and Problems of Protecting the Gene Pool of Mammals. Moscow, 1983; 91-92 (in Russian)

Lukashov V.V. Molecular Evolution and Phylogenetic Analysis. Moscow: Binom Publ., 2009 (in Russian)

Mammalian Fauna of the USSR. Moscow; Leningrad: Nauka Publ., 1963 (in Russian)

Menetries E. Catalogue raisonne des objets de zoologie recueillis dans un voyage au Caucase et jusqu'aux frontieres actuelles de la perse Entrepris par ordre de S.M. Lempereur. St. Petersbourg: Academie Impériale des Sciences, 1832

Naumov N.P. Mechanisms of interaction of populations (some mammals and birds taken as an example). *Russian Journal of Ornithology*. 2010;19(576):1003-1024 (in Russian)

Nikol'skii A.A., Ermakov O.V., Titov S.V. Geographical variability of the little ground squirrel (*Spermophilus pygmaeus*): a bioacoustic analysis. *Zoologičeskij Zhurnal = Zoological Journal* (*Moscow*). 2007;86(11):1379-1388 (in Russian)

Obolenskii S.I. A preliminary review of the palaearctic sousliks (*Citellus* and *Spermophilopsis*). *Doklady Akademii Nauk SSSR = Proceeding of the USSR Academy of Sciences*. 1927;188-193 (in Russian)

Ognev S.I. Genus *Citellus* Oken. Animals of the USSR and Adjacent Countries. Vol. 5. Moscow; Leningrad, 1947;11-215 (in Russian)

Orlov V.N., Rodova M.A., Kotenkova E.V. Chromosome differentiation of the ground squirrels subgenus *Citellus*. In: Mammals: Evolution, Karyology, Taxonomy, Fauna. Novosibirsk, 1969;48-49 (in Russian)

Pastukhov B.N. Epizootological state of natural plague foci in the USSR in 1954–1956 and analysis of the measures taken. In: Natural Foci and Epidemiology of Especially Dangerous Infections. Saratov, 1959;5-17 (in Russian)

Pavlinov I.Ya., Lisovsky A.A. The Mammals of Russia: A Taxonomic and Geographic Reference. Moscow: KMK Scientific Press Ltd, 2012 (in Russian)

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 2018:67(5):901-904. doi 10.1093/sysbio/syy032

Ray N., Currat M., Excoffier L. Intra-deme molecular diversity in spatially expanding populations. *Mol Biol Evol*. 2003;20(1):76-86. doi 10.1093/molbev/msg009

Red Book of the Kabardino-Balkarian Republic. Nalchik: Pechatnyi Dvor Publ., 2018 (in Russian)

Rogers A.R., Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol*. 1992;9(3): 552-569. doi 10.1093/oxfordjournals.molbev.a040727

Ronquist F., Huelsenbeck J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19(12):1572-1574. doi 10.1093/bioinformatics/btg180

Satunin K.A. Mammals of northeastern Ciscaucasia according to the collection of the expedition of the Caucasian Museum in the summer of 1906. *Izvestiya Kavkazskogo Muzeya = Proceedings of the Caucasian Museum*. 1907;3(2-3):94-142 (in Russian)

Shilova S.A. Abundance control and conservation of sousliks in Russia (g. *Spermophilus*). *Arid Ecosyst*. 2011;1(4):267-272. doi 10.1134/S2079096111040147

Simonov E., Lopatina N.V., Titov S.V., Ivanova A.D., Brandler O.V., Surin V.L., Matrosova V.A., Dvilis A.E., Oreshkova N.V., Kapustina S.Yu., Golenishchev F.N., Ermakov O.A. Traditional multilocus phylogeny fails to fully resolve Palearctic ground squirrels (*Spermophilus*) relationships but reveals a new species endemic to West Siberia. *Mol Phylogenet Evol*. 2024;195:108057. doi 10.2139/ssrn.4609201

Steppan S.J., Akhverdyan M.R., Lyapunova E.A., Fraser D.G., Vorontsov N.N., Hoffmann R.S., Braun M.J. Molecular phylogeny of the marmots (Rodentia: Sciuridae): tests of evolutionary and biogeographic hypotheses. *Syst Biol*. 1999;48(4):715-734. doi 10.1080/106351599259988

Sviridenko P.A. Prevalence of ground squirrels in the North Caucasus region and some considerations on the origin of the fauna of Ciscaucasian and Kalmyk steppes. *Izvestiya Severo-Kavkazskoy Krayevoy Stantsii Zashchity Rasteniy = Proceedings of the North-Caucasian Plant Protection Unit*. 1927;3:123-171 (in Russian)

Sviridenko P.A. The greater Caucasus ground squirrel *Citellus musicus* Menet. and the origin of the mountain steppe. *Zoologičeskij Zhurnal = Zoological Journal* (*Moscow*). 1937;16(3):448-482 (in Russian)

Suchard M.A., Lemey P., Baele G., Ayres D.L., Drummond A.J., Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. 2018;4(1):vey016. doi 10.1093/ve/vey016

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123(3):585-595. doi 10.1093/genetics/123.3.585

Tembotov A.K., Ivanov V.G., Ivanov I.V., Tembotova E.Zh. On the distribution and geographical variability of the little ground squirrel of the Northern Caucasus. *Byulleten MOIP = Bulletin of Moscow Society of Naturalists. Biological Series*. 1969;74(5):28-41 (in Russian)

Tembotova F.A., Kononenko E.P. Mammals of steppe ecosystems under threat of extinction in the North Caucasus. *Izvestiya Samarskogo Nauchnogo Tsentra RAN = Izvestia of Samara Scientific Center of the Russian Academy of Sciences*. 2017;19(5/2):253-259 (in Russian)

Tembotova F.A., Gudova M.S., Amshokova A.Kh., Khalidov A.Kh. Genetic diversity of the little ground squirrel *Spermophilus pygmaeus* Pallas, 1779 (Sciuridae, Rodentia) in the Northern Caucasus. *Russ J Genet*. 2024;60(7):908-919. doi 10.1134/S1022795424700340

Tsvirka M.V., Korablev V.P. A case of chromosomal speciation in little suslik (*Spermophilus pygmaeus* (Pallas 1832)) and Caucasian mountain suslik (*Spermophilus musicus* (Menetries 1832)) (Rodentia, Sciuridae). *Zoologičeskij Zhurnal = Zoological Journal* (*Moscow*). 2014;93(7):917-925. doi 10.7868/S0044513414070150 (in Russian)

Tsvirka M.V., Korablev V.P., Chelomina G.N. Genetic differentiation of closely related ground squirrel species *Spermophilus musicus*, *S. pygmaeus*, and *S. suslicus* (Rodentia, Sciuridae). In: Systematics, Phylogeny and Paleontology of Small Mammals. St. Petersburg, 2003;228-230 (in Russian)

Varshavskii S.N. The age of little souslik settlements in different landscape zones in connection with the dissemination and history of the species range. *Byulleten MOIP = Bulletin of Moscow Society of Naturalists. Biological Series*. 1963;68(5):3-14 (in Russian)

Vereshchagin N.K. The Mammals of the Caucasus. A History of the Evolution of the Fauna. Moscow; Leningrad, 1959 (in Russian)

Vinogradov B.S., Argiropulo A.I. Guide to Rodents. Fauna of the USSR. Mammals. Moscow: Nauka Publ., 1941 (in Russian)

Vorontsov N.N., Lyapunova E.A. Chromosomes of Palaearctic ground squirrels (*Citellus*, Marmotinae, Sciuridae, Rodentia). In: Mammals: Evolution, Karyology, Faunistics, Systematics. Novosibirsk, 1969;41-47 (in Russian)

Willi Y., Van Buskirk J., Hoffmann A.A. Limits to the adaptive potential of small populations. *Annu Rev Ecol Evol Syst*. 2006;37(1):433-458. doi 10.1146/annurev.ecolsys.37.091305.110145

Yin Y., Jiang W., Zhang Z., Li Y., Twenke B., Turghan M., Yang W., Liu B. The divergence of small mammals in Xinjiang, China, as revealed by phylogentic analyses of COI and *Cytb*. *Anim Biol*. 2014;64(2):163-176. doi 10.1163/15707563-00002435

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

Received March 14, 2025. Revised May 6, 2025. Accepted May 12, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):812-818

doi 10.18699/vjgb-25-88

Metabolic effects of trehalose in mice of the C57BL/6 strain with obesity induced by a high carbohydrate-fat diet

A.B. Pupyshev [1], N.M. Bazhan [2], A.Yu. Kazantseva [2], T.V. Yakovleva [2], V.M. Belichenko [1], N.V. Goncharova [1], T.A. Korolenko [1], M.A. Tikhonova [1]

[1] Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
pupyshevab@neuronm.ru

Abstract. The ability of trehalose to improve metabolic parameters in mice with experimental obesity has been shown to depend on the type of obesity model. In db/db mice, it reduced body weight, insulin, blood glucose, and cholesterol levels. In mice with obesity induced by high-fat dietary intake, it had no effect on body weight but reduced blood insulin levels with compensatory upregulation of insulin signaling gene expression. We studied the effect of trehalose on overweight and metabolic parameters in C57BL/6 inbred mice with obesity induced by a high carbohydrate-fat diet, the "cafeteria diet". The cafeteria diet consisted of free access to water, standard chow, fatty foods (lard), and carbohydrates (biscuits) for 18 weeks. All mice were then randomly divided into four groups for four weeks of treatment: (1) water drinking, (2) drinking 3 % trehalose, (3) cafeteria diet and drinking water, (4) cafeteria diet and drinking 3 % trehalose. Alterations in body mass, food intake, fluid intake, dietary calories, blood biochemical parameters (glucose, triglyceride, cholesterol, HDL, ALT, creatinine levels), expression of carbohydrate metabolism (*Slc2a2*, *Insr*) and autophagy (*Atg8*, *Becn1*, *Park2*) genes in the liver were studied. The cafeteria diet obesity model was accompanied by some signs of metabolic syndrome as it induced an increase in body weight (by 25 %), calorie intake (by 25 %), blood levels of glucose (by 35 %), cholesterol (by 66 %), and triglycerides (by 23 %) in mice. Trehalose had little effect on control mice, causing a decrease in standard food intake and an increase in dietary caloric intake by the number of calories from trehalose itself. In obese mice, trehalose increased total caloric intake and biscuit consumption but had no substantial effect on body weight gain, blood metabolic parameters, or expression of liver genes regulating glucose transport (*Slc2a2*), insulin sensitivity (*Insr*), and autophagy processes (*Atg8*, *Becn1*, *Park2*). Since the cafeteria diet is the most adequate model of alimentary obesity development in humans, our results question the use of trehalose to correct the dietary type of obesity in humans.

Key words: C57BL/6 mice; carbohydrate-fat diet; cafeteria diet; obesity; trehalose; autophagy; qPCR; glucose; triglycerides; cholesterol

For citation: Pupyshev A.B., Bazhan N.M., Kazantseva A.Yu., Yakovleva T.V., Belichenko V.M., Goncharova N.V., Korolenko T.A., Tikhonova M.A. Metabolic effects of trehalose in mice of the C57BL/6 strain with obesity induced by a high carbohydrate-fat diet. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):812-818. doi 10.18699/vjgb-25-88

Funding. The work was supported by the funds of the federal budget for SRINM (theme No. 122042700001-9) and the budget project of ICG SB RAS (No. FWNR-2022-0021).

Метаболические эффекты трегалозы у мышей линии C57BL/6 с ожирением, вызванным диетой с высоким содержанием углеводов и жиров

А.Б. Пупышев [1], Н.М. Бажан [2], А.Ю. Казанцева [2], Т.В. Яковлева [2], В.М. Беличенко [1], Н.В. Гончарова [1], Т.А. Короленко [1], М.А. Тихонова [1]

[1] Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
pupyshevab@neuronm.ru

Аннотация. Показано, что способность трегалозы улучшать метаболические показатели у животных с экспериментальным ожирением зависит от модели ожирения. У мышей линии db/db она снижает вес тела, уровни инсулина, глюкозы и холестерина в крови. У мышей с ожирением, вызванным потреблением высокожировой диеты, она не влияет на вес тела, но снижает уровень инсулина в крови, компенсаторно усиливая экспрессию генов инсулиновой сигнализации. Нами предпринято исследование действия трегалозы на вес и метаболи-

ческие показатели у мышей линии C57BL/6 с избыточным весом, вызванным диетой с повышенным содержанием жиров и углеводов – «диетой кафетерия». Диета кафетерия включала свободный доступ на протяжении 18 недель к воде, стандартному корму, жирной пище (салу) и углеводам (сдобному печенью). Все мыши были рандомно разделены на четыре группы, содержавшиеся в разных условиях в течение 4 недель: 1) питье воды; 2) питье 3 % трегалозы; 3) диета кафетерия и питье воды; 4) диета кафетерия и питье 3 % трегалозы. Исследовали изменения массы тела, потребление корма, жидкости, пищевых калорий, биохимические показатели крови (уровень глюкозы, триглицеридов, холестерина, ЛПВП, АЛТ, креатинина), экспрессию генов углеводного обмена (*Slc2a2*, *Insr*) и аутофагии (*Atg8*, *Becn1*, *Park2*) в печени. Модель ожирения с помощью диеты кафетерия сопровождалась признаками метаболического синдрома, поскольку у этих мышей были повышены: масса тела (на 25 %), количество потребленных калорий (на 20 %), уровни в крови глюкозы (на 35 %), холестерина (на 66 %), триглицеридов (на 23 %). На контрольных мышей трегалоза действовала слабо, вызывая лишь снижение потребления стандартного корма и повышение потребления пищевых калорий на величину калорийности самой трегалозы. У мышей с ожирением трегалоза повышала общее число потребленных калорий и потребление печенья, но существенно не влияла на массу тела, метаболические показатели крови и экспрессию в печени генов, регулирующих транспорт глюкозы (*Slc2a2*), чувствительность к инсулину (*Insr*) и процессы аутофагии (*Atg8*, *Becn1*, *Park2*). Поскольку диета кафетерия является наиболее адекватной моделью формирования ожирения у людей, полученные нами результаты ставят под сомнение возможность использования трегалозы для коррекции моделируемого ожирения у людей.

Ключевые слова: мыши C57BL/6; углеводножировая диета; диета кафетерия; ожирение; трегалоза; аутофагия; ПЦР; глюкоза; триглицериды; холестерин

Introduction

Trehalose (TR) has multiple therapeutic properties, the main of which seem to be the chaperone-like activity and activation of autophagy, which is especially important for neuroprotection (Hosseinpour-Moghaddam et al., 2018; Pupyshev et al., 2022b). Other beneficial TR properties include positive effects on cellular metabolism, carbohydrate and lipid metabolism (Arai et al., 2019; Yaribeygi et al., 2019; Kobayashi et al., 2021), while diabetes, obesity, and neurodegeneration are known to be closely related (Pugazhenthi et al., 2017). Neurodegeneration is considered to be inhibited by TR through the activation of mTOR-independent autophagy (Sarkar, 2013; Tamargo-Gómez, Mariño, 2018).

At the same time, TR significantly affects carbohydrate metabolism, as it can reduce blood glucose levels and insulin resistance (Zhang et al., 2018; Zhang, DeBosch, 2019; Korolenko et al., 2021). TR also has a positive effect on lipid metabolism: it reduces the level of triglycerides in the liver and blood (Stachowicz et al., 2019; Zhang, DeBosch, 2019; Korolenko et al., 2021) and ultimately prevents the development of steatosis dependent on autophagy activity (Zhang et al., 2018; Ren et al., 2019; Su et al., 2025). TR regulates lipid metabolism partly through its effect on the release of adiponectin, which promotes fat burning (Arai et al., 2013; Mizote et al., 2016) through the reduction of the secretion of gastric inhibitory polypeptide GIP, which promotes obesity (Yoshizane et al., 2017) through the effects on the expression of lipoxygenase ALOXE3 and arginase 2, which increase energy consumption (Higgins et al., 2018; Zhang et al., 2019).

In a high-fat diet (HFD)-induced obesity model, TR reduces mesenteric and inguinal fat hypertrophy and brown fat gain (Arai et al., 2019), which is accompanied by an increase in thermogenesis both in C57BL/6 mice and in a genetic model of diabetic obesity, ob/ob mice (Zhang et al., 2018), while in the latter case the result depends on the activity of AMPK, TFEB and UCP1, but not on autophagy (Zhang et al., 2018; Rusmini et al., 2019). In general, the results on the effects of TR on excess body weight are quite contradictory (Arai et al., 2010, 2019; Liu et al., 2013; Sahebkar et al., 2019; Korolenko et al., 2021). Some studies reported a tendency to decrease in the mass of total visceral fat (no more than 5 %) and no significant effect on the body weight in mice kept on HFD after 8-week consumption of 2 % TR (Arai et al., 2013; Liu et al., 2013). According to other data (Korolenko et al., 2021), in db/db mice (a monogenic model of diabetic obesity), three-week treatment with 2 % TR caused a noticeable decrease in the body weight of mice (more than 10 %) and a general therapeutic effect, in particular, a decrease in the levels of cholesterol, triglycerides, and plasma glucose.

It remains unclear to what extent the fat-reducing effect of TR depends on the type of diet or on mutations that cause obesity. In the present study, we considered that the development of alimentary forms of obesity in mice can be induced both by the consumption of HFD food and by the diet with an increased content of both fats and carbohydrates (carbohydrate-fat diet, “cafeteria diet”, DCaf), which is more common in the human population.

In this regard, the study aimed to assess the effect of alimentary consumption of TR on the main metabolic parameters (body weight; food, calorie and water consumption) in C57BL/6 mice with obesity induced by keeping the animals on a carbohydrate-fat diet (DCaf). The biochemical parameters and lipid spectrum of the blood plasma and autophagy, assessed by the expression of autophagy genes in the liver, were also studied.

Materials and methods

Modeling obesity. All animal manipulations performed during the study met the ethical standards approved by the legal acts of the Russian Federation, the principles of the Basel Declaration and the recommendations of the independent bioethics committee of the Institute of Cytology and Genetics SB RAS (protocol number No. 76 dated 04/07/2021). The experiments were performed in male C57BL/6 mice of the conventional vivarium of the Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia).

Three-month-old animals were kept individually in a cage under a 12-hour light : 12-hour dark regimen, at a temperature of 22–24 °C, and with free access to water and pelleted chow (Assortiment-Agro, Novosibirsk, Russia). After two weeks, the animals were either left on standard diet (n = 22) or transferred to a high-fat, high-carbohydrate diet (DCaf) (n = 19) consisting of unsalted pork lard, biscuits, and standard pelleted chow. DCaf causes obesity in C57BL/6 mice within 18 weeks (Makarova et al., 2013); DCaf content is the closest to the current daily diet in humans and it also allows to assess the consumption of certain dietary components.

18 weeks after the start of DCaf intake, the animals were divided into four groups (Fig. 1): 1) standard diet (the consumption of standard chow) and water (n = 11); 2) standard diet (the consumption of standard chow) and 3 % TR solution (the rationale behind the TR dosage adopted in the current study was based on our recent study (Pupyshev et al., 2024)) (n = 11); 3) DCaf and water (n = 9); 4) DCaf and 3 % TR solution (n = 10). Throughout the experiment, mice body weight was measured weekly while food intake was measured three times a week.

Blood biochemistry. After four weeks of trehalose consumption, the animals were sacrificed, blood and liver samples were collected. Blood serum samples were collected as described previously (Goncharova et al., 2016). Metabolic parameters (ALT, creatinine, glucose, triglycerides, cholesterol, and high-density lipoproteins (HDL)) were determined in the blood serum using an AU 680 Biochemical Analyzer (Beckman Coulter, USA).

Analysis of gene expression. Real-time qPCR was used to evaluate the relative expression levels of liver genes involved in the regulation of carbohydrate metabolism (*Insr*, encoding the insulin receptor; *Slc2a2*, encoding the glucose transporter type 2 GLUT2) and autophagy activity (*Atg8*, encoding the autophagy protein LC3-II; *Becn1*, encoding the protein Beclin 1; *Park2*, encoding the protein Parkin), as well as the reference genes *Hprt1*, *B2m*, *Ppia*.

Total RNA was isolated from liver samples using the ExtractRNA kit (Evrogen, Moscow, Russia) according to the manufacturer's instructions. cDNA synthesis was performed using the MMLV RT kit (Evrogen) according to the manufacturer's protocol (https://evrogen.ru/products/cdna/synthesis/mmlv). The resulting cDNA samples were analyzed by qPCR on a LightCycler-480 II thermal cycler (Roche, Switzerland) using the BioMaster HS-qPCR SYBR Blue (2×) reagent kit (Biolabmiks, Novosibirsk, Russia), with forward (F) and reverse (R) primers (150 nM each) for the studied genes. Primers used for the target genes: *Atg8* [F: 5′-AAA GAG TGG AAG ATG TCC GGC-3′ and R: 5′-ACC AGG AAC TTG GTC TTG TCC-3′], *Becn1* [F: 5′-GAA CTC ACA GCT CCA TTA CTT A-3′ and R: 5′-ATC TTC GAG AGA CAC CAT CC-3′], *Insr* [F: 5′-ATC CTC GAA GGT GAG AAG AC-3′ and R: 5′-TGA TAC CAG AGC ATA GGA GC-3′], *Park2* [F: 5′-GGT CCA GTT AAA CCC ACC TAC-3′ and R: 5′-TTA AGA CAT CGT CCC AGC AAG-3′], *Slc2a2* [F: 5′-GGCTAATTTCAGGACTGGTT-3′ and R: 5′-TTTCTTTGCCCTGACTTCCT-3′]. Primers used for the reference genes: *B2m* [F: 5′-GTC TTT CTA TAT CCT GGC TCA-3′ and R: 5′-ATG CTT GAT CAC ATG TCT CG-3′], *Hprt1* [F: 5′-TAC CTA ATC ATT ATG CCG AGG A-3′ and R: 5′-GGT CAG CAA AGA ACT TAT AGC C-3′], *Ppia* [F: 5′-AAA GTT CCA AAG ACA GCA GAA AA-3′ and R: 5′-GCC AGG ACC TGT ATG CTT TAG-3′]. The relative concentration of the tested cDNA was determined using LightCycler 480 software (1.5.1 version) and calibration curves.

Statistics. The data were analyzed using STATISTICA 10.0 (StatSoft, TIBCO Software Inc., Palo Alto, CA, USA). The results were expressed as the mean ± standard error of the mean. A two-tailed Student's t-test was used to compare sample means. A statistically significant level of differences was accepted at $p < 0.05$ (two-tailed).

Results

Modeling diabetic obesity

Before the start of treatment with 3 % TR solution (after 18 weeks of maintenance on DSt or DCaf), mice from comparison groups 1 and 2, as well as obese animals from groups 3 and 4, did not differ in body weight, food consumption, or taste preferences (see the Table, Fig. 2).

Mice that consumed DCaf for 18 weeks developed excess body weight (obesity): body weight increased by 25 % ($p < 0.001$), energy consumption increased by 20 % ($p < 0.001$), and water consumption, on the contrary, was reduced by 40 % ($p < 0.001$) compared to mice in the control group (Fig. 2).

In animals kept on DCaf, the indicator of hyperglycemia (blood glucose level) increased by 35 % (see below), and this differs from the effect of DCaf in other studies, where blood glucose level increased more significantly (Parafati et al., 2015), or in genetic models of obesity, ob/ob and db/db

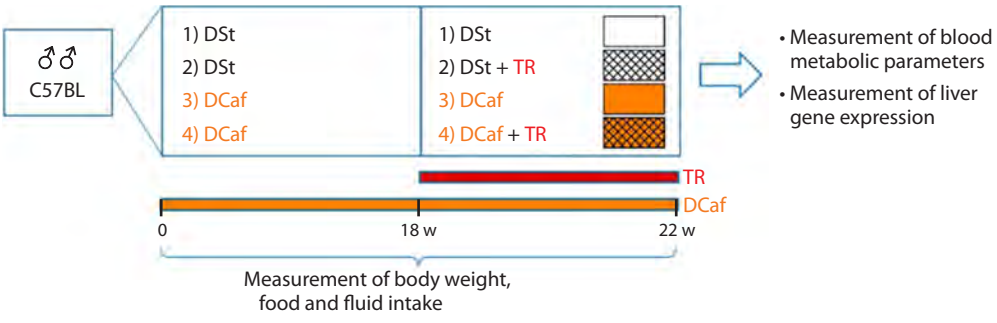


Fig. 1. Scheme of the experiment.

Abbreviations: DSt – standard laboratory diet, DCaf – cafeteria diet, TR – trehalose.

Body weight, water consumption, total energy and dietary component intake in mice in four groups

Index	Group 1 DSt (n = 11)	Group 2 DSt + TR (n = 11)	Group 3 DCaf (n = 9)	Group 4 DCaf + TR (n = 10)
Body weight, g	27.2 ± 0.4	27.3 ± 0.3	33.4 ± 1.0	33.6 ± 1.3
Total feed calories, kcal	13.4 ± 0.4	13.3 ± 0.5	15.8 ± 0.3	15.9 ± 0.2
Standard chow, g	4.5 ± 0.1	4.4 ± 0.2	1.1 ± 0.1	1.3 ± 0.1
Lard, g	–	–	1.0 ± 0.1	0.9 ± 0.1
Biscuits, g	–	–	1.1 ± 0.1	1.1 ± 0.1
Water, ml	5.6 ± 0.2	5.1 ± 0.3	3.3 ± 0.1	3.4 ± 0.1

Note. Mice in groups 1 and 2 maintained on standard diet (DSt) and in groups 3 and 4 maintained on the cafeteria diet (DCaf) for 18 weeks before the start of treatment with 3 % trehalose (TR) solution in groups 2 and 4. Daily component consumption is indicated. Results are expressed as M ± m (n, number of animals).

mice (Pelletier et al., 2020; Korolenko et al., 2021). In our experiment, excess weight increase was not accompanied by the development of high hyperglycemia.

Effects of TR on metabolic parameters in mice kept on standard laboratory diet

Consumption of a 3 % TR solution for four weeks in control mice did not affect the body weight of the animals (Fig. 3). These mice ate less food ($p < 0.05$), while they showed a tendency for an increase in fluid consumption by 10 % (Fig. 4). Taking into account the caloric supplementation from TR in drinking, the total kcal consumption in control mice drinking a 3 % TR solution was more than 30 % higher than in control animals drinking water ($p < 0.001$) (Fig. 3).

Effects of TR on metabolic parameters in mice kept on cafeteria diet

Despite our expectations, the body weight of mice kept on DCaf did not change under the influence of TR (Fig. 3). TR also had a weak effect on their consumption of standard food, liquid, or lard but increased the consumption of the carbohydrate component, biscuits (Fig. 4). TR significantly increased the number of calories consumed, by 18 % (taking into account the unchanged body weight in these animals compared to the group drinking water and the caloric supplementation from TR itself) (Fig. 3).

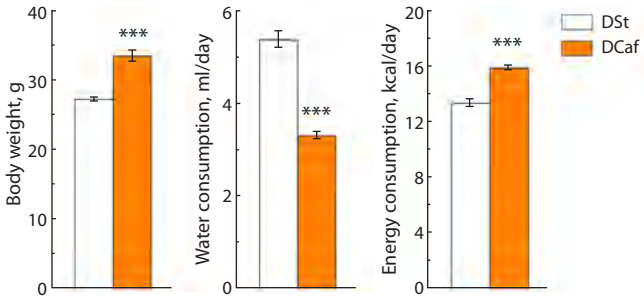


Fig. 2. Body weight, water consumption, and total energy intake of mice maintained on standard diet (n = 22) or cafeteria diet (n = 19) for 18 weeks before they were given 3 % trehalose solution.

*** $p < 0.001$ *vs.* DSt.

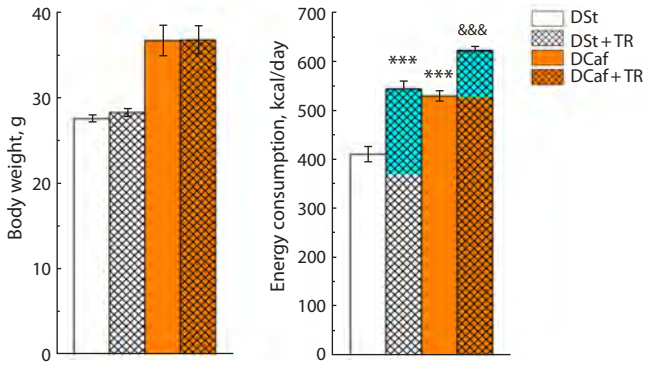


Fig. 3. Body weight and total energy intake in mice maintained on standard diet and cafeteria diet for four weeks during which the animals consumed water or 3 % trehalose solution.

The blue color shows the number of kcal obtained with drinking 3 % TR solution.

*** $p < 0.001$ *vs.* DSt; &&& $p < 0.001$ *vs.* DCaf and water group (*t*-test).

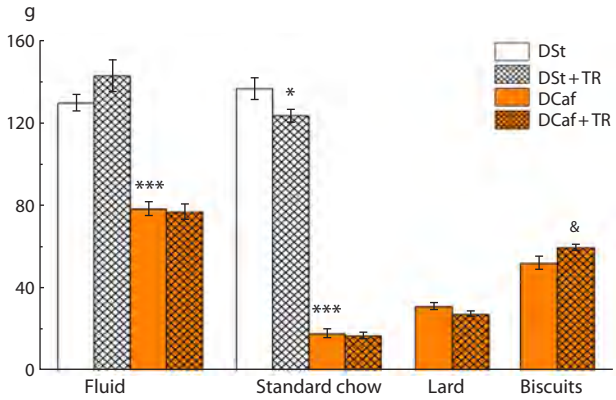


Fig. 4. Total fluid and various food components intake of mice maintained on standard diet or cafeteria diet for four weeks during which the animals consumed water or 3 % trehalose solution.

* $p < 0.05$, *** $p < 0.001$ *vs.* DSt; & $p < 0.05$ *vs.* DCaf and water group (*t*-test).

TR did not produce significant changes in metabolic blood indices in mice maintained on DCaf or in the control group (except for a noticeable trend ($p < 0.07$) in blood glucose growth) (Fig. 5). The maintenance of mice on DCaf per se influenced the overall metabolism to a certain extent increasing the levels of glucose ($p < 0.01$), triglycerides ($p < 0.05$), and especially those of blood cholesterol ($p < 0.001$), i. e. changes in carbohydrate and lipid metabolism were registered.

Effects of TR on transcription of carbohydrate metabolism and autophagy genes in mice kept on cafeteria diet

Neither DCaf nor TR consumption affected the expression of genes regulating glucose uptake from blood (*Slc2a2*, *Insr*) or related to autophagy activity (*Atg8*, *Becn1*, *Park2*) in the liver of mice (Fig. 6).

Discussion

The maintenance of mice on DCaf for 18 weeks was accompanied by an increase in caloric intake (by 20 %) and, as a result, the development of excess body weight (by 25 %). That allowed us to study a normalizing effect of TR on body weight gain found in previous studies in db/db mice (Korolenko et al., 2021). TR as an energy substrate (Sato et al., 1999) increased caloric energy intake both in mice of the control group and in mice given DCaf. Notably, in both groups of mice, an increase in caloric intake mainly associated with TR intake did not augment body mass (Fig. 3). Hence, TR stimulates caloric energy expenditure without affecting body weight. To some extent, this is in line with previous findings that demonstrated the ability of long-term TR treatment to stimulate caloric expenditure in thermogenesis and brown fat burning processes in C57BL/6 mice kept on both standard and high-fat diets (Arai et al., 2013, 2019) and in obese ob/ob mice (Zhang et al., 2018).

Another possible way of regulation by alimentary TR involves the induction of cellular starvation accompanied by activation of cytoprotective autophagy (DeBosch et al., 2016; Mayer et al., 2016; Zhang, DeBosch, 2019). The mechanism is associated with inhibition of the GLUT8 glucose transmembrane transporter, thereby causing energy deficiency in liver cells leading to activation of the adenosine monophosphate-dependent kinase AMPK and its pleiotropic metabolic effect including inhibition of biosynthesis, energy consumption, and activation of autophagy. In the present study, TR appeared to lose its ability to simulate the fasting effect (Zhang, DeBosch, 2019), which leads to autophagy activation. The attenuation of the regulatory properties of TR here is consistent with its lack of effect on the transcription of autophagy genes responding to TR by elevation in our other studies (Pupyshev et al., 2022a). The reason for the weakening of the regulatory properties of TR in C57BL/6 mice remains unclear. Such a switch of the effect of TR does not seem to depend on the diet, as the effects of TR were similar both in control mice and in DCaf-given ones. Perhaps, in the conditions in our study (3 % trehalose, 28 days), there is an escape of TR from the quantitative energy cleavage described earlier (Sato et al., 1999), and then the contradiction between the growth of calories consumed and the lack of weight gain is smoothed out.

The loss of regulatory properties of TR in mice in our experiment is not consistent with the effects of TR revealed in db/db mice with diabetic obesity (Korolenko et al., 2021). In db/db mice, TR reduced the body mass (by more than 10 %),

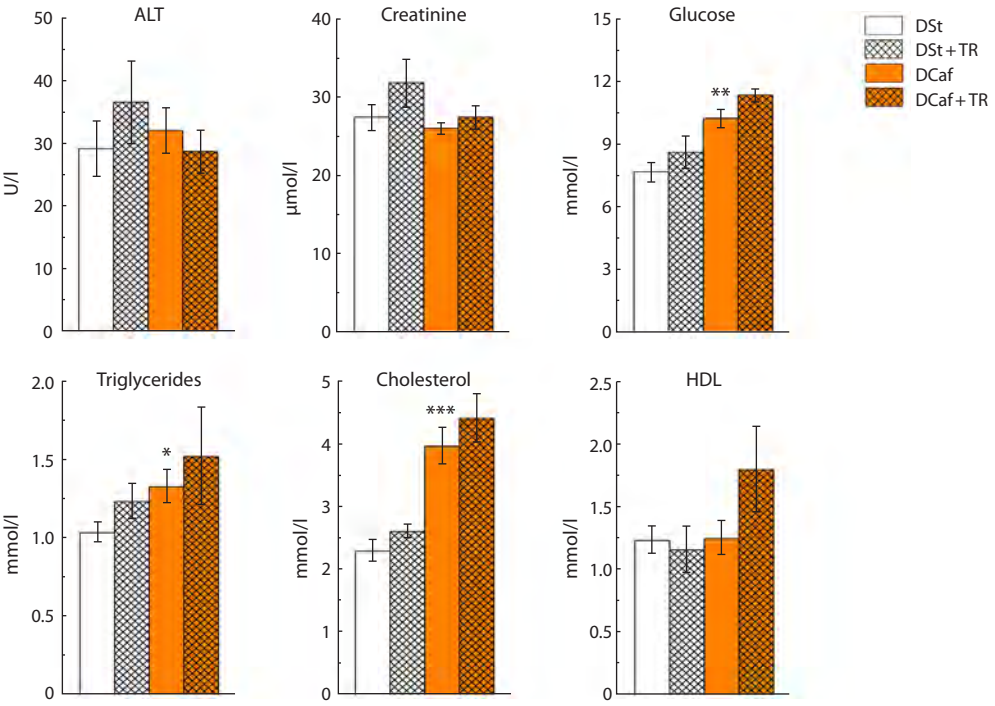


Fig. 5. Blood biochemical parameters in mice maintained on standard diet or cafeteria diet and given water or 3 % trehalose solution for four weeks.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ *vs.* DSt.

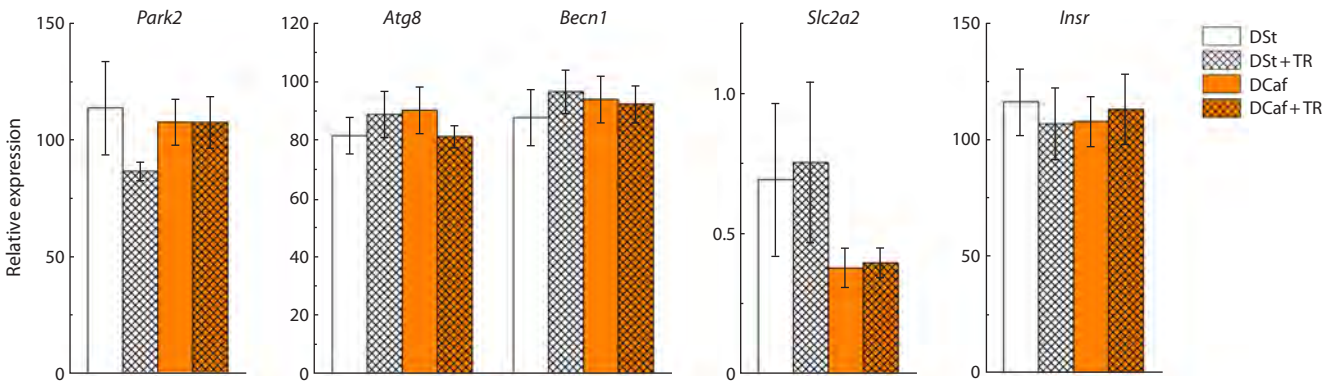


Fig. 6. The mRNA levels of autophagy (*Park2*, *Atg8*, *Becn1*) and carbohydrate metabolism (*Slc2a2*, *Insr*) genes in the liver of mice kept on standard diet or cafeteria diet and given water or 3 % trehalose solution for four weeks.

reduced blood glucose levels, and produced a general recovery effect by lowering blood cholesterol and triglycerides. However, in studies in rabbits, as here, a weak effect of TR on the blood lipid spectrum was found (Sahebkar et al., 2019).

Given the similarity of the model of obesity in DCaf-given mice with the typical development of obesity in humans, the results question the use of TR to correct this most common type of obesity in humans. At the same time, our results do not exclude the approach of treating patients with severe obesity with TR because such treatment was successful for mice with almost 50 % excess weight (Korolenko et al., 2021).

Conclusion

Based on the well-known effect of TR on simulation of starvation (induction of autophagy) and reduction of excess body weight in db/db mice, a study of its effect in C57BL/6 mice given carbohydrate-fat diet (DCaf), a model an alimentary obesity in humans, was performed. In the control and obese (by 25 %) mice, TR (3 % solution drinking, 28 days) augmented significantly the number of calories consumed while this increase in energy consumption was not accompanied by an increase in body weight gain of the mice. The excess in calories consumed might be spent on enhancing the processes of thermogenesis and brown fat burning (Arai et al., 2013, 2019). Trehalose produced only a tendency of increase in blood metabolic parameters (glucose, cholesterol, triglycerides, HDL) and had no effect on the expression of genes regulating carbohydrate metabolism (*Slc2a2*, *Insr*) or autophagy genes (*Atg8*, *Becn1*, *Park2*). In the present study, in mice kept on DCaf, TR did not demonstrate an ability to reduce diabetes and obesity induced by DCaf, nor useful properties for the correction of common dietary type of obesity in humans.

References

Arai C., Arai N., Mizote A., Kohno K., Iwaki K., Hanaya T., Arai S., Ushio S., Fukuda S. Trehalose prevents adipocyte hypertrophy and mitigates insulin resistance. *Nutr Res*. 2010;30(12):840-848. doi 10.1016/j.nutres.2010.10.009

Arai C., Miyake M., Matsumoto Y., Mizote A., Yoshizane C., Hanaya Y., Koide K., Yamada M., Hanaya T., Arai S., Fukuda S. Trehalose prevents adipocyte hypertrophy and mitigates insulin resistance in mice with established obesity. *J Nutr Sci Vitaminol* (*Tokyo*). 2013;59(5):393-401. doi 10.3177/jnsv.59.393

Arai C., Arai N., Arai S., Yoshizane C., Miyata S., Mizote A., Suyama A., Endo S., Ariyasu T., Mitsuzumi H., Ushio S. Continuous intake of trehalose induces white adipose tissue browning and enhances energy metabolism. *Nutr Metab*. 2019;16:45. doi 10.1186/s12986-019-0373-4

DeBosch B.J., Heitmeier M.R., Mayer A.L., Higgins C.B., Crowley J.R., Kraft T.E., Chi M., Newberry E.P., Chen Z., Finck B.N., Davidson N.O., Yarasheski K.E., Hruz P.W., Moley K.H. Trehalose inhibits solute carrier 2A (SLC2A) proteins to induce autophagy and prevent hepatic steatosis. *Sci Signal*. 2016;9(416):ra21. doi 10.1126/scisignal.aac5472

Goncharova N.V., Pupyshev A.B., Filyushina E.E., Loktev K.V., Korolenko E.Ts., Lushnikova E.L., Molodykh O.P., Korolenko T.A., Churin B.V. Depression of macrophages modifies serum lipid profile in hyperlipidemia. *Bull Exp Biol Med*. 2016;160(5):617-621. doi 10.1007/s10517-016-3231-7

Higgins C.B., Zhang Y., Mayer A.L., Fujiwara H., Stothard A.I., Graham M.J., Swarts B.M., DeBosch B.J. Hepatocyte ALOXE3 is induced during adaptive fasting and enhances insulin sensitivity by activating hepatic PPARγ. *JCI Insight*. 2018;3(16):e120794. doi 10.1172/jci.insight.120794

Hosseinpour-Moghaddam K., Caraglia M., Sahebkar A. Autophagy induction by trehalose: molecular mechanisms and therapeutic impacts. *J Cell Physiol*. 2018;233(9):6524-6543. doi 10.1002/jcp.26583

Kobayashi M., Yasukawa H., Arikawa T., Deguchi Y., Mizushima N., Sakurai M., Onishi S., Tagawa R., Sudo Y., Okita N., Higashi K., Higami Y. Trehalose induces SQSTM1/p62 expression and enhances lysosomal activity and antioxidative capacity in adipocytes. *FEBS Open Bio*. 2021;11(1):185-194. doi 10.1002/2211-5463.13055

Korolenko T.A., Dubrovina N.I., Ovsyukova M.V., Bgatova N.P., Tenditnik M.V., Pupyshev A.B., Akopyan A.A., Goncharova N.V., Lin C.L., Zavjalov E.L., Tikhonova M.A., Amstislavskaya T.G. Treatment with autophagy inducer trehalose alleviates memory and behavioral impairments and neuroinflammatory brain processes in db/db mice. *Cells*. 2021;10(10):2557. doi 10.3390/cells10102557

Liu M., Zhang M., Ye H., Lin S., Yang Y., Wang L., Jones G., Trang H. Multiple toxicity studies of trehalose in mice by intragastric administration. *Food Chem*. 2013;136(2):485-490. doi 10.1016/j.foodchem.2012.09.031

Makarova E.N., Chepeleva E.V., Panchenko P.E., Bazhan N.M. Influence of abnormally high leptin levels during pregnancy on metabolic phenotypes in progeny mice. *Am J Physiol Regul Integr Comp Physiol*. 2013;305(11):R1268-R1280. doi 10.1152/ajpregu.00162.2013

Mayer A.L., Higgins C.B., Heitmeier M.R., Kraft T.E., Qian X., Crowley J.R., Hyrc K.L., Beatty W.L., Yarasheski K.E., Hruz P.W., DeBosch B.J. SLC2A8 (GLUT8) is a mammalian trehalose transporter required for trehalose-induced autophagy. *Sci Rep*. 2016;6:38586. doi 10.1038/srep38586

Mizote A., Yamada M., Yoshizane C., Arai N., Maruta K., Arai S., Endo S., Ogawa R., Mitsuzumi H., Ariyasu T., Fukuda S. Daily intake of trehalose is effective in the prevention of lifestyle-related diseases in individuals with risk factors for metabolic syndrome. *J Nutr Sci Vitaminol* (*Tokyo*). 2016;62(6):380-387. doi 10.3177/jnsv.62.380

Parafati M., Lascala A., Morittu V.M., Trimboli F., Rizzuto A., Brunelli E., Coscarelli F., Costa N., Britti D., Ehrlich J., Isidoro C., Mollace V., Janda E. Bergamot polyphenol fraction prevents nonalcoholic fatty liver disease via stimulation of lipophagy in cafeteria diet-induced rat model of metabolic syndrome. *J Nutr Biochem*. 2015;26(9):938-948. doi 10.1016/j.jnutbio.2015.03.008

Pelletier R.M., Layeghkhavidaki H., Vitale M.L. Glucose, insulin, insulin receptor subunits α and β in normal and spontaneously diabetic and obese ob/ob and db/db infertile mouse testis and hypophysis. *Reprod Biol Endocrinol*. 2020;18(1):25. doi 10.1186/s12958-020-00583-2

Pugazhenthi S., Qin L., Reddy P.H. Common neurodegenerative pathways in obesity, diabetes, and Alzheimer's disease. *Biochim Biophys Acta Mol Basis Dis*. 2017;1863(5):1037-1045. doi 10.1016/j.bbadis.2016.04.017

Pupyshev A.B., Belichenko V.M., Tenditnik M.V., Bashirzade A.A., Dubrovina N.I., Ovsyukova M.V., Akopyan A.A., Fedoseeva L.A., Korolenko T.A., Amstislavskaya T.G., Tikhonova M.A. Combined induction of mTOR-dependent and mTOR-independent pathways of autophagy activation as an experimental therapy for Alzheimer's disease-like pathology in a mouse model. *Pharmacol Biochem Behav*. 2022a;217:173406. doi 10.1016/j.pbb.2022.173406

Pupyshev A.B., Klyushnik T.P., Akopyan A.A., Singh S.K., Tikhonova M.A. Disaccharide trehalose in experimental therapies for neurodegenerative disorders: molecular targets and translational potential. *Pharmacol Res*. 2022b;183:106373. doi 10.1016/j.phrs.2022.106373

Pupyshev A.B., Akopyan A.A., Tenditnik M.V., Ovsyukova M.V., Dubrovina N.I., Belichenko V.M., Korolenko T.A., Zozulya S.A., Klyushnik T.P., Tikhonova M.A. Alimentary treatment with trehalose in a pharmacological model of Alzheimer's disease in mice: effects of different dosages and treatment regimens. *Pharmaceutics*. 2024;16(6):813. doi 10.3390/pharmaceutics16060813

Ren H., Wang D., Zhang L., Kang X., Li Y., Zhou X., Yuan G. Catalpol induces autophagy and attenuates liver steatosis in ob/ob and high-fat diet-induced obese mice. *Aging* (*Albany NY*). 2019;11(21):9461-9477. doi 10.18632/aging.102396

Rusmini P., Cortese K., Crippa V., Cristofani R., Cicardi M.E., Ferrari V., Vezzoli G., … Galbiati M., Garrè M., Morelli E., Vaccari T., Poletti A. Trehalose induces autophagy via lysosomal-mediated TFEB activation in models of motoneuron degeneration. *Autophagy*. 2019;15(4):631-651. doi 10.1080/15548627.2018.1535292

Sahebkar A., Hatamipour M., Tabatabaei S.A. Trehalose administration attenuates atherosclerosis in rabbits fed a high-fat diet. *J Cell Biochem*. 2019;120(6):9455-9459. doi 10.1002/jcb.28221

Sarkar S. Regulation of autophagy by mTOR-dependent and mTOR-independent pathways: autophagy dysfunction in neurodegenerative diseases and therapeutic application of autophagy enhancers. *Biochem Soc Trans*. 2013;41(5):1103-1130. doi 10.1042/BST20130134

Sato S., Okamoto K., Minami R., Kohri H., Yamamoto S. Trehalose can be used as a parenteral saccharide source in rabbits. *J Nutrition*. 1999;129(1):158-164. doi 10.1093/jn/129.1.158

Stachowicz A., Wiśniewska A., Kuś K., Kiepura A., Gębska A., Gajda M., Białas M., Totoń-Żurańska J., Stachyra K., Suski M., Jawień J., Korbut R., Olszanecki R. The influence of trehalose on atherosclerosis and hepatic steatosis in apolipoprotein E knockout mice. *Int J Mol Sci*. 2019;20(7):1552. doi 10.3390/ijms20071552

Su S., Liu X., Zhu M., Liu W., Liu J., Yuan Y., Fu F., Rao Z., Liu J., Lu Y., Chen Y. Trehalose ameliorates nonalcoholic fatty liver disease by regulating IRE1α-TFEB signaling pathway. *J Agric Food Chem*. 2025;73(1):521-540. doi 10.1021/acs.jafc.4c08669

Tamargo-Gómez I., Mariño G. AMPK: regulation of metabolic dynamics in the context of autophagy. *Int J Mol Sci*. 2018;19(12):3812. doi 10.3390/ijms19123812

Yaribeygi H., Yaribeygi A., Sathyapalan T., Sahebkar A. Molecular mechanisms of trehalose in modulating glucose homeostasis in diabetes. *Diabetes Metab Syndr*. 2019;13(3):2214-2218. doi 10.1016/j.dsx.2019.05.023

Yoshizane C., Mizote A., Yamada M., Arai N., Arai S., Maruta K., Mitsuzumi H., Ariyasu T., Ushio S., Fukuda S. Glycemic, insulinemic and incretin responses after oral trehalose ingestion in healthy subjects. *Nutr J*. 2017;16(1):9. doi 10.1186/s12937-017-0233-x

Zhang Y., DeBosch B.J. Using trehalose to prevent and treat metabolic function: effectiveness and mechanisms. *Curr Opin Clin Nutr Metab Care*. 2019;22(4):303-310. doi 10.1097/MCO.0000000000000568

Zhang Y., Sowers J.R., Ren J. Targeting autophagy in obesity: from pathophysiology to management. *Nat Rev Endocrinol*. 2018;14(6):356-376. doi 10.1038/s41574-018-0009-1

Zhang Y., Higgins C.B., Fortune H.M., Chen P., Stothard A.I., Mayer A.L., Swarts B.M., DeBosch B.J. Hepatic arginase 2 (Arg2) is sufficient to convey the therapeutic metabolic effects of fasting. *Nat Commun*. 2019;10(1):1587. doi 10.1038/s41467-019-09642-8

Conflict of interest. The authors declare no conflict of interest.
Received April 16, 2025. Revised June 20, 2025. Accepted June 25, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):819-827

doi 10.18699/vjgb-25-89

Association of the rs823144 variant of the *RAB29* gene with the activity of lysosomal hydrolases in blood cells and risk of Parkinson's disease

K.S. Basharova [1, 2], A.I. Bezrukova [1, 2], K.A. Senkevich [3], G.V. Baydakova [4], A.V. Rybakov[5], I.V. Miliukhina [1, 5], A.A. Timofeeva [2], E.Yu. Zakharova[1, 4], S.N. Pchelina [1, 2], T.S. Usenko [1, 2]

[1] Petersburg Nuclear Physics Institute named by B.P. Konstantinov of National Research Centre "Kurchatov Institute", Gatchina, Russia
[2] Pavlov First Saint Petersburg State Medical University, St. Petersburg, Russia
[3] Montreal Neurological Institute, McGill University, Montreal, Canada
[4] Bochkov Research Centre for Medical Genetics, Moscow, Russia
[5] N.P. Bechtereva Institute of the Human Brain of the Russian Academy of Sciences, St. Petersburg, Russia
kbasharova@yandex.ru; basharova_ks@pnpi.nrcki.ru; usenko_ts@pnpi.nrcki.ru

Abstract. Recent genome-wide association studies have identified a link between the *RAB29* gene and Parkinson's disease (PD). The Rab29 protein encoded by *RAB29* regulates leucine-rich repeat kinase 2 (LRRK2). Mutations in the *LRRK2* gene increase its kinase activity and contribute to autosomal dominant forms of PD. Previous research has shown that altered LRRK2 kinase activity may correlate with the activity of lysosomal hydrolases and the concentration of sphingolipids. This study aimed to assess the association of the rs823144 variant in the promoter region of the *RAB29* gene with PD risk, and to evaluate *RAB29* expression, lysosomal hydrolase activity, and sphingolipid concentrations in the blood of PD patients. We screened the rs823144 variant of the *RAB29* gene in a cohort of PD patients ($N = 903$) and controls ($N = 618$) using next-generation sequencing (NGS) and polymerase chain reaction (PCR) followed by restriction fragment length polymorphism analysis. The expression of the *RAB29* gene was measured in peripheral blood mononuclear cells (PBMCs) using qPCR. We assessed the activities of lysosomal hydrolases (glucocerebrosidase (GCase), alpha-galactosidase (GLA), acid sphingomyelinase (ASMase), and galactosylcerebrosidase (GALC)) and the concentrations of sphingolipids (globotriaosylsphingosine (LysoGb3), sphingomyelin (LysoSM), and hexosylsphingosine (HexSph)) in blood using high-performance liquid chromatography with tandem mass spectrometry (HPLC-MS/MS). The *RAB29* rs823144 C allele was associated with a reduced risk of PD in the Northwestern Russian population (OR = 0.7806, 95 % CI: 0.6578–0.9263, $p = 0.0046$), which is consistent with global data. However, no significant association was observed between the rs823144 C allele and *RAB29* mRNA expression in PBMCs. Notably, the C allele was associated with increased GLA activity and decreased concentrations of LysoGb3 and LysoSM in the blood of PD patients. In conclusion, we demonstrate for the first time an association between the *RAB29* rs823144 C allele and a reduced risk of PD in the Northwestern Russian population. Moreover, the *RAB29* rs823144 C allele is associated with altered lysosomal enzyme activity and sphingolipid profiles, suggesting a potential role of *RAB29* in sphingolipid metabolism relevant to PD pathogenesis.

Key words: Parkinson's disease; *RAB29*; lysosomal hydrolases; lysosphingolipids; LRRK2

For citation: Basharova K.S., Bezrukova A.I., Senkevich K.A., Baydakova G.V., Rybakov A.V., Miliukhina I.V., Timofeeva A.A., Zakharova E.Yu., Pchelina S.N., Usenko T.S. Association of the rs823144 variant of the *RAB29* gene with the activity of lysosomal hydrolases in blood cells and risk of Parkinson's disease. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):819-827. doi 10.18699/vjgb-25-89

Funding. This work was carried out within the framework of the state assignment of the Ministry of Science and Higher Education of the Russian Federation (Project No. 1023031500037-7-1.6.8;1.6.1; 1.6.2;1.6.3: Study of molecular and cellular components of the pathogenesis of socially significant diseases for the development of methods for early diagnosis and treatment) and the state assignment of the Ministry of Health of the Russian Federation (Project No. 123030200067-6: Development of genetic panels for the diagnosis of synucleinopathies based on genomic research).

Acknowledgements. We express our sincere gratitude to all the patients who participated in this study.

Ассоциация варианта rs823144 гена *RAB29* с активностью лизосомных гидролаз в клетках крови и риском болезни Паркинсона

К.С. Башарова [1, 2] ✉, А.И. Безрукова [1, 2], К.А. Сенкевич [3], Г.В. Байдакова [4], А.В. Рыбаков[5], И.В. Милюхина [1, 5], А.А. Тимофеева [2], Е.Ю. Захарова[1, 4], С.Н. Пчелина [1, 2], Т.С. Усенко [1, 2] ✉

[1] Петербургский институт ядерной физики им. Б.П. Константинова Национального исследовательского центра «Курчатовский институт», Гатчина, Россия
[2] Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, Санкт-Петербург, Россия
[3] Монреальский неврологический институт при Университете Макгилла, Монреаль, Канада
[4] Медико-генетический научный центр им. акад. Н.П. Бочкова, Москва, Россия
[5] Институт мозга человека им. Н.П. Бехтеревой Российской академии наук, Санкт-Петербург, Россия
✉ kbasharova@yandex.ru; basharova_ks@pnpi.nrcki.ru; usenko_ts@pnpi.nrcki.ru

Аннотация. Геномные исследования последних лет выявили ассоциацию гена *RAB29* с болезнью Паркинсона (БП). Белок Rab29, кодируемый геном *RAB29*, – один из регуляторов богатой лейциновыми повторами киназы 2 (LRRK2). Мутации в гене *LRRK2* ассоциированы с увеличением киназной активности LRRK2 и приводят к развитию аутосомно-доминантных форм БП. Недавно показано, что изменение киназной активности LRRK2 может быть связано с изменением активности лизосомных гидролаз и концентрации лизосфинголипидов. Цель данного исследования заключалась в оценке ассоциации rs823144 в промоторе гена *RAB29* с БП с экспрессией гена *RAB29*, активностью лизосомных гидролаз и концентрацией лизосфинголипидов в крови при БП. В ходе исследования проведены скрининг варианта rs823144 гена *RAB29* в группе пациентов с БП (N = 903) и в контроле (N = 618) с использованием методов массового параллельного секвенирования и полимеразная цепная реакция (ПЦР) с последующим рестрикционным анализом. Экспрессия гена *RAB29* оценивалась в мононуклеарах периферической крови методом ПЦР в режиме реального времени. Активности лизосомных гидролаз (глюкоцереброзидаза (GCase), альфа-галактозидаза (GLA), кислая сфингомиелиназа (ASMase), галактозилцереброзидаза (GALC)) и концентрации лизосфинголипидов (глоботриаозилсфингозин (LysoGb3), сфингомиелин (LysoSM), гексозилсфингозин (HexSph)) оценивались в крови методом высокоэффективной жидкостной хроматографии с тандемной масс-спектрометрией (ВЭЖХ-МС/МС). Аллель С rs823144 гена *RAB29* ассоциирован с пониженным риском БП в северо-западной популяции Российской Федерации (ОШ: 0.7806, 95 % ДИ: 0.6578–0.9263, p = 0.0046), что соответствует мировым данным. Однако в ходе работы не выявлено ассоциации аллеля С rs823144 гена *RAB29* с уровнем мРНК гена *RAB29* в мононуклеарах периферической крови. В то же время носительство аллеля С rs823144 было ассоциировано с повышенной активностью GLA и сниженной концентрацией LysoGb3 в крови при БП. Таким образом, нами впервые показана ассоциация аллеля С rs823144 гена *RAB29* с пониженным риском БП в северо-западной популяции Российской Федерации. Аллель С rs823144 ассоциирован с повышенной активностью GLA и сниженной концентрацией LysoGb3 в крови при БП. Полученные результаты позволяют предположить ассоциацию гена *RAB29* с метаболизмом сфинголипидов.

Ключевые слова: болезнь Паркинсона; *RAB29*; лизосомные гидролазы; лизосфинголипиды; LRRK2

Introduction

Parkinson's disease (PD) is a common, slowly progressive neurodegenerative disorder characterized by the loss of dopaminergic neurons in the substantia nigra (SN) of the brain (Lill, 2016). A central pathological mechanism in PD pathogenesis is the accumulation and aggregation of the alpha-synuclein protein in the SN. Although PD is primarily sporadic, approximately 15 % of patients report a positive family history. The molecular mechanisms underlying PD remain largely unclear; however, increasing evidence implicates lysosomal dysfunction as a key contributor to disease pathogenesis (Nechushtai et al., 2023). In particular, our group and others have demonstrated reduced lysosomal hydrolase activity and altered sphingolipid levels in the peripheral fluids of patients with idiopathic PD (Alcalay et al., 2015; Galper et al., 2022; Usenko et al., 2022). Additionally, changes in lysosomal enzyme activity and sphingolipid concentrations have been observed in PD cases associated with mutations in the *LRRK2* gene, one of the most common monogenic forms of the disease (Alcalay et al., 2015; Usenko et al., 2023, 2024).

The *LRRK2* gene encodes leucine-rich repeat kinase 2 (LRRK2), a multidomain protein implicated in Parkinson's disease pathogenesis (Zimprich et al., 2004). A key group of LRRK2 substrates comprises small Rab GTPases, which are critical regulators of vesicular trafficking, particularly within the endolysosomal system (Steger et al., 2016; Wang et al., 2014). Dysregulated LRRK2 kinase activity disrupts the trafficking of lysosomal hydrolases to their proper destinations, thereby impairing lysosomal function (MacLeod et al., 2013; Ysselstein et al., 2019; Rivero-Ríos et al., 2020; Kedariti et al., 2022).

Among the LRRK2 substrates, Rab29 – encoded by the *RAB29* gene – has attracted particular interest (Steger et al., 2016). Rab29 has been identified as a key upstream regulator of LRRK2, responsible for its activation (Liu et al., 2018; Madero-Pérez et al., 2018; Purlyte et al., 2018; Kuwahara,

Iwatsubo, 2020). Rab29 localizes to the membranes of lysosomes and the Golgi apparatus, where it recruits inactive cytoplasmic LRRK2 monomers and promotes their oligomerization into active dimers or tetramers (Purlyte et al., 2018; Zhu et al., 2023). The *RAB29* gene is located within the PARK16 locus, which has previously been associated with reduced PD risk (Satake et al., 2009; Pihlstrøm et al., 2015; Nalls et al., 2019). A recent multi-trait analysis of genome-wide association studies (MTAG) further confirmed the association between *RAB29* and PD at both the transcriptomic and proteomic levels (Shi et al., 2024).

Several studies have identified variants in the promoter region of *RAB29* that are associated with a reduced risk of Parkinson's disease. These variants are thought to influence *RAB29* gene expression levels (Gan-Or et al., 2012; Khaligh et al., 2017; Sun et al., 2021), potentially modulating the activation of LRRK2 and thereby affecting lysosomal hydrolase activity in PD.

The aim of this study was to investigate the association of the rs823144 single nucleotide polymorphism (SNP), located in the promoter region of *RAB29*, with PD risk, *RAB29* gene expression, lysosomal hydrolase activity – including glucocerebrosidase (GCase), α-galactosidase (GLA), galactocerebrosidase (GALC), and acid sphingomyelinase (ASMase) – and the concentrations of lysosphingolipids in the blood. The lysosphingolipids analyzed included hexosylsphingosine (HexSph), a mixture of glucosylsphingosine (GlcSph) and galactosylsphingosine (GalSph); lysosphingomyelin (LysoSM); and lysoglobotriaosylsphingosine (LysoGb3). These parameters were evaluated in both PD patients and healthy control subjects.

Materials and methods

Characteristics of the study groups. The study included 903 patients with sporadic PD and 618 control individuals matched for age and gender. All patients were recruited from the clinic of the N.P. Bechtereva Institute of the Human Brain of the Russian Academy of Sciences. The control group comprised individuals seen at the consultative and diagnostic center of the First St. Petersburg State Medical University named after academician I.P. Pavlov. To exclude PD and other neurodegenerative disorders, all control participants underwent neurological examination. Clinical and demographic characteristics of the study groups are presented in Table 1. No significant differences in age or gender distribution were observed between the groups ($p > 0.05$).

All procedures involving human participants were conducted in accordance with the ethical standards of the National Research Ethics Committee and the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Written informed consent was obtained from each participant prior to inclusion in the study. The study protocol was approved by the Ethics Committee of the First St. Petersburg State Medical University named after academician I.P. Pavlov (Protocol No. 275, dated September 04, 2023).

Genetic analysis. Two methods were used to genotype the rs823144 variant in the *RAB29* gene: massively parallel sequencing (next-generation sequencing, NGS) and polymerase chain reaction (PCR) followed by restriction fragment length polymorphism (RFLP) analysis. Peripheral blood samples were collected from all study participants, and genomic DNA was extracted using phenol-chloroform extraction, as previously described (Maniatis et al., 1994).

NGS genotyping of the rs823144 variant in the *RAB29* gene was performed in a subset of 521 PD patients and 420 control individuals using molecular inversion probes, as previously described (Rudakou et al., 2021). Sequencing was conducted on the Illumina NovaSeq 6000 SP PE100 platform. Sequence alignment was performed using the Burrows–Wheeler Aligner (BWA) with the hg19 human genome reference (Li, Durbin, 2009). Variant calling and post-alignment quality control were carried out using the Genome Analysis Toolkit (GATK, v3.8) (McKenna et al., 2010). Variants were filtered based on coverage depth and quality metrics; only those with a minimum read depth >30 and a quality score >20 were included in the analysis.

An additional 473 PD patients and 384 controls were screened for rs823144 using PCR followed by restriction analysis. Primer sequences were designed using Primer3 v. 0.4.0 (http://bioinfo.ut.ee/primer3-0.4.0/) (FOR: 5′-CCCTGCACGTGACGCTTG-3′, REV: 5′-GAATCCCAGTCAGCTCCTTACA-3′). Restriction enzyme selection was performed using the NEBCutter tool (Vincze et al., 2003), and BstAC I was chosen for RFLP analysis (Fig. 1).

A total of 91 PD patients and 186 control individuals were genotyped for the rs823144 variant in the *RAB29* gene using both PCR with restriction analysis and NGS. These overlaps were accounted for in subsequent statistical analyses.

To validate the results obtained from both methods, a subset of samples was confirmed by Sanger sequencing using a Nanofor-05 genetic analyzer (Synthol, Russia). The Sanger sequencing data were visualized and analyzed using Tracy software (Rausch et al., 2020) (Fig. 2).

Evaluation of the relative expression level of the *RAB29* gene in peripheral blood mononuclear cells of PD patients and controls. Peripheral venous blood samples were collected from PD patients (N = 30) and control individuals (N = 43). Peripheral blood mononuclear cells (PBMCs) were

Table 1. Clinical and demographic characteristics of the study groups

Group	Sex (male:female)	Age, years	Age at onset, years	Duration of the disease, years
PD (N = 903)	378:525	65 (25–90)	59 (20–88)	3 (1–36)
Controls (N = 618)	228:390	64 (40–96)	–	–

Note. PD – Parkinson's disease; data are presented as median (min–max).

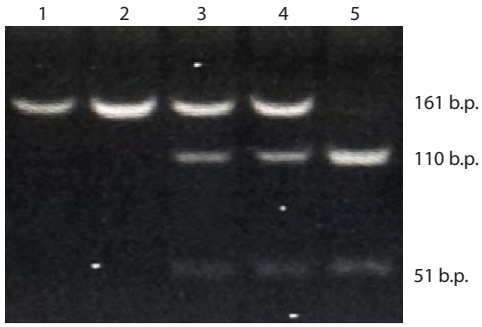


Fig. 1. Electropherogram showing the results of genotyping of the rs823144 variant in the *RAB29* gene.

1, 2 – homozygote for the A allele (genotype AA), 3, 4 – heterozygote (genotype AC), 5 – homozygote for the C allele (genotype CC).

isolated by density gradient centrifugation using Ficoll-Paque PLUS (GE Healthcare) at 400*g* for 40 minutes, as described by Böyum (1968). The resulting mononuclear fraction was washed twice with phosphate-buffered saline (PBS; Biolot, St. Petersburg) and centrifuged at 3,000 rpm for 10 minutes. Total RNA was extracted from PBMCs using the RNeasy Mini Kit (Qiagen, 74104, USA). Complementary DNA (cDNA) was synthesized via reverse transcription using the RevertAid First Strand cDNA Synthesis Kit (K1622, Thermo Scientific, Lithuania).

The relative expression level of the *RAB29* gene in PBMCs of PD patients (N = 30) and control individuals (N = 43) was quantified by real-time PCR using SYBR Green I as the intercalating dye. The housekeeping genes *RPLP0* and *GAPDH*, which are constitutively expressed in PBMCs, were used as internal reference genes for normalization. Primer sequences were designed using the Primer3 v. 0.4.0 program (https://bioinfo.ut.ee/primer3-0.4.0) (FOR: 5′-CGGTTTCACAGGTTGGACAG-3′, REV: 5′-CCCTTGGGTGGACAAAGACA-3′). The relative mRNA level for each gene was calculated by comparing the threshold amplification levels ΔΔCt (Livak, Schmittgen, 2001).

Evaluation of lysosomal hydrolase activities and lysosphingolipid concentrations in peripheral blood of PD patients and controls. Peripheral venous blood samples were collected from PD patients and controls into EDTA tubes. To obtain dried blood spots, 40 µl of whole blood was applied to each spot on a filter paper test blank, after which the spots were allowed to air dry at room temperature for 2 hours and were then stored at +4 °C until extraction. Activities of four lysosomal hydrolases: glucocerebrosidase (GCase), α-galactosidase (GLA), galactocerebrosidase (GALC), and sphingomyelinase (ASMase), as well as the concentration of three lysosphingolipids: hexazyl sphingosine (HexSph) (a mixture of glycosyl sphingosine (GlcSph) and galactosyl sphingosine (GalSph)), lysosphingomyelin (LysoSM) and lysoglobotriaosyl sphingosine (LysoGb3) were estimated by high-performance liquid chromatography coupled with tandem mass spectrometry (HPLC-MS/MS) according to our previously published protocol (Pchelina et al., 2018).

Statistical data processing. Statistical analyses were performed using the R programming environment (version 4.0.5). Odds ratios (OR) with 95 % confidence intervals (CIs) were calculated using logistic regression models adjusted for age and gender. The nonparametric Mann–Whitney U-test was applied to compare the relative expression levels of the *RAB29* gene, lysosomal hydrolase activities, and lysosphingolipid concentrations between study groups. To evaluate the association between the rs823144 variant of the *RAB29* gene and lysosomal hydrolase activity, multiple linear regression analysis was conducted, adjusting for age, gender, and disease duration. Statistical significance was set at $p < 0.05$. Data are presented as median (min–max).

Results

Association between rs823144 of the *RAB29* gene and PD risk

Genotyping of the rs823144 variant in the *RAB29* gene among PD patients and controls from the Northwestern Russian population revealed that the major allele is A. Hardy–Weinberg equilibrium (HWE) analysis confirmed that the genotype

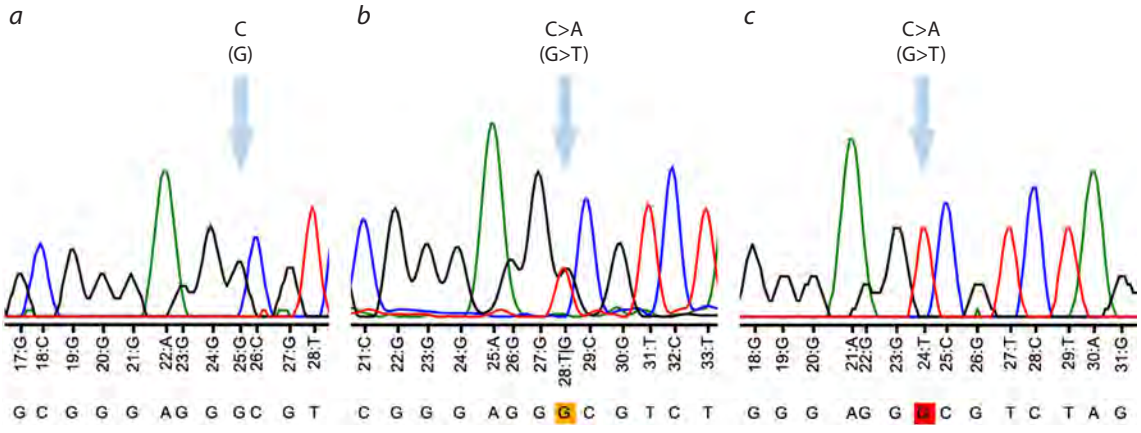


Fig. 2. Graphical representation of Sanger sequencing results (electropherogram displaying the results of genotyping of the rs823144 variant of the *RAB29* gene).

a – homozygote for the C allele (genotype CC); *b* – heterozygote (genotype AC); *c* – homozygote for the A allele (genotype AA).

distributions in both groups were in equilibrium ($p > 0.05$). The genotype frequencies are summarized in Table 2. Logistic regression analysis demonstrated that the rs823144 variant is significantly associated with a reduced risk of Parkinson's disease in this population (OR: 0.7806, 95 % CI: 0.6578–0.9263, $p = 0.0046$).

Evaluation of the relative level of *RAB29* gene mRNA in the mononuclear fraction of peripheral blood cells of PD patients and controls

This study assessed for the first time the association between the rs823144 variant of the *RAB29* gene and the relative expression level of *RAB29* mRNA in peripheral blood mononuclear cells (PBMCs) from PD patients and controls. The median relative expression of *RAB29* mRNA in PBMCs was 1.00 (0.22–1.75) in PD patients and 0.96 (0.13–1.79) in controls, with no statistically significant difference between groups ($p > 0.05$). When stratified by rs823144 genotype, the relative *RAB29* mRNA levels were as follows: in PD patients, AA genotype carriers had a median expression of 1.09 (0.–1.68), while carriers of AC and CC genotypes combined showed 1.13 (0.36–1.75); in controls, AA carriers had 1.05 (0.58–1.28) and AC + CC carriers had 0.97 (0.55–1.33). No significant association was found between the rs823144 genotype and *RAB29* mRNA levels in either group ($p > 0.05$).

Association of the rs823144 variant of the *RAB29* gene with the activity of lysosomal hydrolases and the concentration of lysosphingolipids in the peripheral blood of patients with PD and controls

This study is the first to investigate the association of the rs823144 variant of the *RAB29* gene with lysosomal hydrolase activities and lysosphingolipid concentrations in the blood of PD patients and controls. Patients with PD exhibited significantly increased GALC activity and decreased LysoSM concentrations compared to controls ($p = 0.008$ and $p = 0.01$, respectively) (Table 3). Stratification by rs823144 genotype revealed that carriage of the C allele of the rs823144 variant in the *RAB29* gene was associated with increased GLA activity and decreased levels of LysoGb3 – a substrate of GLA – in PD patients ($p = 0.038$ and $p = 0.022$, respectively) (Table 3). These associations were confirmed by multiple linear regression adjusted for gender, age, and disease duration (GLA: $\beta = 1.11$, $p = 0.024$; LysoGb3: $\beta = –0.23$, $p = 0.015$) (Table 4). No significant differences were observed in GCase, GLA and ASMase activities, or HexSph and LysoGb3 concentrations when analyzing the combined groups of PD patients and controls, regardless of rs823144 genotype. Similarly, no genotype-dependent differences were detected in GCase, ASMase, GALC activities, or HexSph concentrations.

Notably, PD patients carrying the C allele exhibited decreased blood concentrations of LysoGb3 and LysoSM compared to the combined control group ($p = 0.045$ and $p = 0.015$, respectively), whereas PD patients with the AA genotype showed increased LysoSM levels relative to controls ($p = 0.022$). The increase in GLA activity and corresponding decrease in its substrate LysoGb3 were specific to carriers of C allele of the rs823144 variant in the *RAB29* gene within the PD group and were not observed in controls.

Discussion

The *RAB29* gene is one of five genes located within the PARK16 locus on chromosome 1q32, previously implicated in Parkinson's disease risk (Simón-Sánchez et al., 2009; Tucci et al., 2010). Multiple association studies have identified the minor allele C of rs823144, positioned in the promoter region of *RAB29*, as protective against PD across various populations (Gan-Or et al., 2012; Xia et al., 2015; Khaligh et al., 2017; Sun et al., 2021). It has been hypothesized that this allele modulates transcription factor binding (Gan-Or et al., 2012; Khaligh et al., 2017).

In silico analyses suggest that the C allele of the rs823144 variant in the *RAB29* gene abolishes the binding site for c-Ets-1 and introduces binding sites for p300, GATA-1, and Sp1, potentially enhancing *RAB29* expression (Gan-Or et al., 2012). However, data from the GTEx database (https://www.gtexportal.org/) indicate that the CC genotype of rs823144 correlates with reduced *RAB29* expression in whole blood and brain tissues ($p < 0.0001$). In turn, increased expression of the *RAB29* gene, encoding one of the main regulators of LRRK2 kinase, may potentially lead to impaired activation of this kinase. LRRK2, by phosphorylating Rab family proteins, regulates endolysosomal transport (Reczek et al., 2007; Wei et al., 2023). Dysregulation of LRRK2 activation can impair lysosomal hydrolase transport and function, contributing to PD pathogenesis.

Table 2. Frequencies of rs823144 genotypes and alleles in the study groups

Genotypes and alleles	PD, % (*N* = 903)	Controls, % (*N* = 618)	OR (95 % CI), *p*-value
AA	63.3 (*N* = 572)	57 (*N* = 352)	AC+CC vs AA: 0.7658 (0.6213–0.9438), *p* = 0.0123; CC vs AA+CC: 0.6671 (0.4392–1.0132), *p* = 0.0576
AC	31.5 (*N* = 284)	35.4 (*N* = 219)	
CC	5.2 (*N* = 47)	7.6 (*N* = 47)	
AC+CC	36.7 (*N* = 331)	43 (*N* = 266)	
AA+AC	94.8 (*N* = 856)	92.4 (*N* = 571)	
A	79.1 (*N* = 1428)	74.7 (*N* = 923)	C vs A: 0.7806 (0.6578–0.9263), *p* = 0.0046
C	20.9 (*N* = 378)	25.3 (*N* = 313)	

Note. PD – Parkinson's disease; OR – odds ratio.

Table 3. Lysosomal hydrolase activity and lysosphingolipid levels in the peripheral blood of PD patients and controls

Genotypes	Lysosomal hydrolase activity, mM/L/hour				Lysosphingolipid concentration, ng/ml		
	GCase	GLA	ASMase	GALC	HexSph	LysoGb3	LysoSM
Combined groups							
PD ($N = 211$)	6.78 (2.07–23.08)	4.8 (1.33–36.39)	4.54 (1.53–13.25)	2.13 (0.21–12.68) *$p = 0.008$	2.63 (0.49–13.23)	0.8 (0.04–40.77)	3.62 (0.72–16.08) *$p = 0.01$
Controls ($N = 179$)	6.29 (1.55–32.13)	4.18 (1.03–14.81)	4.14 (1.4–12.39)	1.86 (0.24–9.35)	2.97 (0.57–15.36)	0.78 (0.03–2.31)	3.98 (0.59–11.6)
Groups divided by genotypes of the rs823144 variant of the *RAB29* gene							
PD AA ($N = 91$)	5.67 (2.07–19.52)	4.05 (1.2–13.93)	4.28 (1.67–11.83)	2.17 (0.12–12.68)	2.55 (0.5–13.81)	0.93 (0.02–2.49)	3.84 (1.65–16.08) *$p = 0.022$
PD AC + CC ($N = 55$)	5.78 (2.42–23.08)	5.39 (1.61–13.72) **$p = 0.038$	4.76 (1.53–9.31)	2.2 (0.52–7.28)	2.89 (0.87–13.08)	0.61 (0.04–3.73) *$p = 0.045$ **$p = 0.022$	3.45 (0.72–15.34) *$p = 0.015$
Controls AA ($N = 52$)	6.92 (1.55–32.13)	4.23 (1.86–12.94)	4.35 (1.5–10.99)	2.19 (0.58–9.35)	2.37 (0.57–12.11)	0.77 (0.16–2.27)	4.37 (0.59–14.87)
Controls AC + CC ($N = 40$)	8.11 (3.9–17.23)	4.42 (2.17–12.6)	4.4 (1.82–12.03)	2.19 (0.96–8.37)	2.54 (0.69–9.87)	0.86 (0.03–2.31)	4.44 (2.03–11.75)

Note. PD – Parkinson's disease; GCase – glucocerebrosidase; GLA – α-galactosidase; GALC – galactocerebrosidase; ASMase – sphingomyelinase; HexSph – hexasylsphingosine; LysoSM – lysosphingomyelin; LysoGb3 – lysoglobotriaosylsphingosine.

* Compared with the combined control group; ** compared with PD patients with the AA genotype. Data are presented as median (min–max).

In our study of the Northwestern Russian population, we confirmed the association of the C allele of rs823144 with a reduced risk of PD, which is consistent with global findings (Gan-Or et al., 2012; Khaligh et al., 2017; Sun et al., 2021). Notably, we did not observe any association between the C allele of rs823144 genotype and *RAB29* mRNA levels in peripheral blood mononuclear cells from either PD patients or controls.

Lysosomal dysfunction is widely recognized as a central mechanism in PD. Previous work by our group and others has demonstrated altered lysosomal hydrolase activity and sphingolipid metabolism in peripheral fluids and postmortem brain regions of PD patients (Alcalay et al., 2018; Nelson et al., 2018; Huebecker et al., 2019; Chang et al., 2022; Usenko et al., 2022, 2024). Specifically, we reported increased GALC activity and decreased LysoSM – a substrate of acid ASMase – concentration in PD peripheral blood (Usenko et al., 2022, 2024). In turn, disruption of lysosphingolipid metabolism, including through altered lysosomal hydrolase activity, may contribute to the aggregation of α-synuclein (Mazzulli et al., 2011; Marie et al., 2015).

Accumulating lysosphingolipids in neurons can stabilize neurotoxic alpha-synuclein oligomers (Battis et al., 2023). Prior brain autopsy studies revealed correlations between LysoGb3 isoform concentrations and pathological phosphorylated alpha-synuclein as well as an negative correlation of GLA activity with levels of α-synuclein phosphorylated at serine 129 – the pathological form of the protein that predominates in aggregates in PD (Nelson et al., 2018). Our novel findings linking the C allele of rs823144 to increased GLA activity and decreased LysoGb3 concentration, alongside elevated LysoSM levels in AA genotype carriers, underscore a potential role for *RAB29* in sphingolipid metabolism in PD.

It should be noted that a marked reduction in GLA activity due to mutations in the *GLA* gene leads to a rare lysosomal storage disorder, Fabry disease. Notably, elevated LysoGb3 concentration is a risk factor for white matter lesions in Fabry disease. (Rombach et al., 2010). We have also observed LysoGb3 accumulation in neuronopathic mucopolysaccharidoses (Baydakova et al., 2020), suggesting that LysoGb3 elevation may not be exclusive to Fabry disease. Therefore, the association of the *RAB29* rs823144 C allele with increased GLA activity and decreased LysoGb3 in PD patients may influence disease progression and clinical phenotype.

The present study has several limitations. The sample sizes of PD patients and the control group included in the experiment assessing the association of the *RAB29* rs823144 variant with *RAB29* mRNA levels were limited. The compared groups were population-wise heterogeneous. Furthermore, we did not directly assess the impact of rs823144 on LRRK2 kinase activity, which warrants investigation in future studies.

Conclusion

This study demonstrates for the first time worldwide that the C allele of the rs823144 variant in the *RAB29* gene, previously identified as protective against PD in other populations,

Table 4. Lysosomal hydrolase activities and lysosphingolipid concentrations in the peripheral blood of patients with PD (regression analysis)

Lysosomal hydrolases/lysosphingolipids	Genotype/ other parameters	PD			Controls		
		β	95 % CI	*p*-value	β	95 % CI	*p*-value
GCase	AC+CC	–0.95	–2.32–0.43	0.17	1.44	–0.25–3.14	0.09
	Sex (male:female)	0.55	–0.81–1.90	0.43	0.12	–1.61–1.85	0.89
	Age at onset, years	–0.03	–0.094–0.028	0.29	–0.06	–0.17–0.05	0.30
	Duration of the disease, years	0.004	–0.042–0.049	0.87	–	–	–
GLA	AC+CC	**1.11**	**0.15–2.07**	**0.02**	–0.29	–1.52–0.95	0.64
	Sex (male:female)	0.03	–0.93–0.98	0.96	0.93	–0.33–2.18	0.14
	Age at onset, years	–0.03	–0.07–0.01	0.16	–0.00001	–0.080–0.080	1
	Duration of the disease, years	–0.002	–0.03–0.03	0.93	–	–	–
ASMase	AC+CC	0.23	–0.53–0.98	0.55	0.31	–0.85–1.46	0.60
	Sex (male:female)	0.05	–0.69–0.80	0.87	–0.20	–1.38–0.97	0.73
	Age at onset, years	**0.06**	**0.03–0.09**	**0.0005**	–0.021	–0.096–0.054	0.58
	Duration of the disease, years	**0.025**	**0.0005–0.050**	**0.046**	–	–	–
GALC	AC+CC	0.080	–0.57–0.72	0.81	0.07	–0.64–0.77	0.85
	Sex (male:female)	0.16	–0.48–0.80	0.62	–0.15	–0.87–0.57	0.67
	Age at onset, years	–0.017	–0.046–0.012	0.25	–0.011	–0.057–0.035	0.63
	Duration of the disease, years	–0.0051	–0.027–0.016	0.64	–	–	–
HexSph	AC+CC	–0.014	–0.92–0.90	0.97	0.12	–0.80–1.04	0.80
	Sex (male:female)	0.35	–0.55–1.24	0.45	0.49	–0.44–1.43	0.30
	Age at onset, years	–0.028	–0.069–0.013	0.18	0.011	–0.048–0.071	0.71
	Duration of the disease, years	–0.028	–0.058–0.0022	0.069	–	–	–
LysoGb3	AC+CC	–0.23	**–0.41–0.044**	**0.015**	0.039	–0.16–0.24	0.70
	Sex (male:female)	–0.08	–0.26–0.10	0.39	–0.043	–0.25–0.16	0.67
	Age at onset, years	–0.007	–0.015–0.0016	0.11	–0.0039	–0.017–0.009	0.55
	Duration of the disease, years	–0.001	–0.007–0.004	0.66	–	–	–
LysoSM	AC+CC	–0.17	–1.18–0.84	0.74	–0.25	–1.51–1.014	0.70
	Sex (male:female)	–0.82	–1.82–0.17	0.10	–0.63	–1.91–0.65	0.33
	Age at onset, years	–0.064	–0.11–0.018	**0.006**	0.070	–0.011–0.15	0.089
	Duration of the disease, years	–0.028	–0.061–0.0056	0.10	–	–	–

Note. PD – Parkinson's disease; GCase – glucocerebrosidase; GLA – α-galactosidase; GALC – galactocerebrosidase; ASMase – sphingomyelinase; HexSph – hexasylsphingosine; LysoSM – lysosphingomyelin; LysoGb3 – lysoglobotriaosylsphingosine.

is associated with a reduced risk of PD in the North-West region of Russia. Additionally, we report for the first time that carriage of the C allele correlates with increased GLA activity and decreased LysoGb3 concentration in the blood of PD patients.

These findings suggest a potential role for *RAB29* in lysosphingolipid metabolism and imply that the rs823144 variant may influence the clinical course of PD. Further research is warranted to elucidate the relationship between the PARK16 locus, *RAB29* gene, lysosphingolipid metabolism, and PD progression.

References

Alcalay R.N., Levy O.A., Waters C.C., Fahn S., Ford B., Kuo S.H., Mazzoni P., … Wolf P., Oliva P., Keutzer J., Marder K., Zhang X. Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. *Brain.* 2015;138(9):2648-2658. doi 10.1093/brain/awv179

Alcalay R.N., Wolf P., Levy O.A., Kang U.J., Waters C., Fahn S., Ford B., … Chung W.K., Oliva P., Keutzer J., Marder K., Zhang X.K. Alpha galactosidase A activity in Parkinson's disease. *Neurobiol Dis.* 2018;112:85-90. doi 10.1016/j.nbd.2018.01.012

Battis K., Xiang W., Winkler J. The bidirectional interplay of α-synuclein with lipids in the central nervous system and its implications

for the pathogenesis of Parkinson's disease. *Int J Mol Sci*. 2023; 24(17):13270. doi 10.3390/ijms241713270

Baydakova G., Ilyushkina A., Gaffke L., Pierzynowska K., Bychkov I., Ługowska A., Wegrzyn G., Tylki-Szymanska A., Zakharova E. Elevated LysoGb3 concentration in the neuronopathic forms of mucopolysaccharidoses. *Diagnostics* (*Basel*). 2020;10(3):155. doi 10.3390/diagnostics10030155

Böyum A. Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. *Scand J Clin Lab Invest Suppl*. 1968;97:77-89

Chang K.H., Cheng M.L., Tang H.Y., Huang C.Y., Wu H.C., Chen C.M. Alterations of sphingolipid and phospholipid pathways and ornithine level in the plasma as biomarkers of Parkinson's disease. *Cells*. 2022;11(3):395. doi 10.3390/cells11030395

Galper J., Dean N.J., Pickford R., Lewis S.J.G., Halliday G.M., Kim W.S., Dzamko N. Lipid pathway dysfunction is prevalent in patients with Parkinson's disease. *Brain*. 2022;145(10):3472-3487. doi 10.1093/brain/awac176

Gan-Or Z., Bar-Shira A., Dahary D., Mirelman A., Kedmi M., Gurevich T., Giladi N., Orr-Urtreger A. Association of sequence alterations in the putative promoter of *RAB7L1* with a reduced Parkinson disease risk. *Arch Neurol*. 2012;69(1):105-110. doi 10.1001/archneurol.2011.924

Huebecker M., Moloney E.B., Van Der Spoel A.C., Priestman D.A., Isacson O., Hallett P.J., Platt F.M. Reduced sphingolipid hydrolase activities, substrate accumulation and ganglioside decline in Parkinson's disease. *Mol Neurodegener*. 2019;14(1):40. doi 10.1186/s13024-019-0339-z

Kedariti M., Frattini E., Baden P., Cogo S., Civiero L., Ziviani E., Zilio G., … Di Fonzo A., Alcalay R.N., Rideout H., Greggio E., Plotegher N. LRRK2 kinase activity regulates GCase level and enzymatic activity differently depending on cell type in Parkinson's disease. *NPJ Parkinsons Dis*. 2022;8(1):92. doi 10.1038/s41531-022-00354-3

Khaligh A., Goudarzian M., Moslem A., Mehrtash A., Jamshidi J., Darvish H., Emamalizadeh B. RAB7L1 promoter polymorphism and risk of Parkinson's disease; a case-control study. *Neurol Res*. 2017; 39(5):468-471. doi 10.1080/01616412.2017.1297558

Kuwahara T., Iwatsubo T. The emerging functions of LRRK2 and Rab GTPases in the endolysosomal system. *Front Neurosci*. 2020;14: 227. doi 10.3389/fnins.2020.00227

Li H., Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi 10.1093/bioinformatics/btp324

Lill C.M. Genetics of Parkinson's disease. *Mol Cell Probes*. 2016; 30(6):386-396. doi 10.1016/j.mcp.2016.11.001

Liu Z., Bryant N., Kumaran R., Beilina A., Abeliovich A., Cookson M.R., West A.B. LRRK2 phosphorylates membrane-bound Rabs and is activated by GTP-bound Rab7L1 to promote recruitment to the trans-Golgi network. *Hum Mol Genet*. 2018;27(2):385-395. doi 10.1093/hmg/ddx410

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

MacLeod D.A., Rhinn H., Kuwahara T., Zolin A., Di Paolo G., MacCabe B.D., Marder K.S., Honig L.S., Clark L.N., Small S.A., Abeliovich A. Rab7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron*. 2013; 77(3):425-439. doi 10.1016/j.neuron.2012.11.033

Madero-Pérez J., Fdez E., Fernández B., Lara Ordóñez A.J., Blanca Ramírez M., Gómez-Suaga P., Waschbüsch D., … Beilina A., Gonnelli A., Cookson M.R., Greggio E., Hilfiker S. Parkinson disease-associated mutations in LRRK2 cause centrosomal defects via Rab8a phosphorylation. *Mol Neurodegener*. 2018;13(1):3. doi 10.1186/S13024-018-0235-y

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

Marie G., Dunning C.J., Gaspar R., Grey C., Brundin P., Sparr E., Linse S. Acceleration of α-synuclein aggregation by exosomes. *J Biol Chem*. 2015;290(5):2969. doi 10.1074/jbc.M114.585703

Mazzulli J.R., Xu Y.H., Sun Y., Knight A.L., McLean P.J., Caldwell G.A., Sidransky E., Grabowski G.A., Krainc D. Gaucher disease glucocerebrosidase and α-synuclein form a bidirectional pathogenic loop in synucleinopathies. *Cell*. 2011;146(1):37-52. doi 10.1016/j.cell.2011.06.001

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M.A. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303. doi 10.1101/gr.107524.110

Nalls M.A., Blauwendraat C., Vallerga C.L., Heilbron K., Bandres-Ciga S., Chang D., Tan M., … Silburn P.A., Vallerga C.L., Wallace L., Wray N.R., Zhang F. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-genome wide association study. *Lancet Neurol*. 2019;18(12):1091-1102. doi 10.1016/S1474-4422(19)30320-5

Nechushtai L., Frenkel D., Pinkas-Kramarski R. Autophagy in Parkinson's disease. *Biomolecules*. 2023;13(10):1435. doi 10.3390/biom13101435

Nelson M.P., Boutin M., Tse T.E., Lu H., Haley E.D., Ouyang X., Zhang J., Auray-Blais C., Shacka J.J. The lysosomal enzyme alpha-Galactosidase A is deficient in Parkinson's disease brain in association with the pathologic accumulation of alpha-synuclein. *Neurobiol Dis*. 2018;110:68-81. doi 10.1016/j.nbd.2017.11.006

Pchelina S., Baydakova G., Nikolaev M., Senkevich K., Emelyanov A., Kopytova A., Miliukhina I., Yakimovskii A., Timofeeva A., Berkovich O., Fedotova E., Illarioshkin S., Zakharova E. Blood lysosphingolipids accumulation in patients with Parkinson's disease with glucocerebrosidase 1 mutations. *Mov Disord*. 2018;33(8):1325-1330. doi 10.1002/mds.27393

Pihlstrøm L., Rengmark A., Bjørnaråa K.A., Dizdar N., Fardell C., Forsgren L., Holmberg B., Larsen J.P., Linder J., Nissbrandt H., Tysnes O.B., Dietrichs E., Toft M. Fine mapping and resequencing of the PARK16 locus in Parkinson's disease. *J Hum Genet*. 2015; 60(7):357-362. doi 10.1038/jhg.2015.34

Purlyte E., Dhekne H.S., Sarhan A.R., Gomez R., Lis P., Wightman M., Martinez T.N., Tonelli F., Pfeffer S.R., Alessi D.R. Rab29 activation of the Parkinson's disease-associated LRRK2 kinase. *EMBO J*. 2018;37(1):1-18. doi 10.15252/EMBJ.201798099

Rausch T., Fritz M.H.Y., Untergasser A., Benes V. Tracy: basecalling, alignment, assembly and deconvolution of sanger chromatogram trace files. *BMC Genomics*. 2020;21(1):230. doi 10.1186/S12864-020-6635-8

Reczek D., Schwake M., Schröder J., Hughes H., Blanz J., Jin X., Brondyk W., Van Patten S., Edmunds T., Saftig P. LIMP-2 is a receptor for lysosomal mannose-6-phosphate-independent targeting of beta-glucocerebrosidase. *Cell*. 2007;131(4):770-783. doi 10.1016/j.cell.2007.10.018

Rivero-Ríos P., Romo-Lozano M., Fernández B., Fdez E., Hilfiker S. Distinct roles for Rab10 and Rab29 in pathogenic LRRK2-mediated endolysosomal trafficking alterations. *Cells*. 2020;9(7):1719. doi 10.3390/cellS9071719

Rombach S.M., Dekker N., Bouwman M.G., Linthorst G.E., Zwinderman A.H., Wijburg F.A., Kuiper S., vd Bergh Weerman M.A., Groener J.E., Poorthuis B.J., Hollak C.E., Aerts J.M. Plasma globotriaosylsphingosine: diagnostic value and relation to clinical manifestations of Fabry disease. *Biochim Biophys Acta*. 2010;1802(9): 741-748. doi 10.1016/j.bbadis.2010.05.003

Rudakou U., Yu E., Krohn L., Ruskey J.A., Asayesh F., Dauvilliers Y., Spiegelman D., … Rouleau G.A., Hassin-Baer S., Fon E.A., Alcalay R.N., Gan-Or Z. Targeted sequencing of Parkinson's disease

loci genes highlights *SYT11*, *FGF20* and other associations. *Brain*. 2021;144(2):462-472. doi 10.1093/brain/awaa401

Satake W., Nakabayashi Y., Mizuta I., Hirota Y., Ito C., Kubo M., Kawaguchi T., … Yamamoto M., Hattori N., Murata M., Nakamura Y., Toda T. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet*. 2009;41(12):1303-1307. doi 10.1038/ng.485

Shi J.J., Mao C.Y., Guo Y.Z., Fan Y., Hao X.Y., Li S.J., Tian J., … Zuo C.Y., Liang Y.Y., Xu Y.M., Yang J., Shi C.H. Joint analysis of proteome, transcriptome, and multi-trait analysis to identify novel Parkinson's disease risk genes. *Aging*. 2024;16(2):1555-1580. doi 10.18632/aging.205444

Simón-Sánchez J., Schulte C., Bras J.M., Sharma M., Gibbs J.R., Berg D., Paisan-Ruiz C., … Chen H., Riess O., Hardy J.A., Singleton A.B., Gasser T. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet*. 2009;41(12):1308-1312. doi 10.1038/ng.487

Steger M., Tonelli F., Ito G., Davies P., Trost M., Vetter M., Wachter S., … Fell M.J., Morrow J.A., Reith A.D., Alessi D.R., Mann M. Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *eLife*. 2016;5:e12813. doi 10.7554/eLife.12813.001

Sun J., Deng L., Zhu H., Liu M., Lyu R., Lai Q., Zhang Y. Meta-analysis on the association between rs11868035, rs823144, rs3851179 and Parkinson's disease. *Meta Gene*. 2021;30:100949. doi 10.1016/j.mgene.2021.100949

Tucci A., Nalls M.A., Houlden H., Revesz T., Singleton A.B., Wood N.W., Hardy J., Paisán-Ruiz C. Genetic variability at the PARK16 locus. *Eur J Hum Genet*. 2010;18(12):1356-1359. doi 10.1038/ejhg.2010.125

Usenko T.S., Senkevich K.A., Bezrukova A.I., Baydakova G.V., Basharova K.S., Zhuravlev A.S., Gracheva E.V., … Palchikova E.I., Zalutskaya N.M., Emelyanov A.K., Zakharova E.Y., Pchelina S.N. Impaired sphingolipid hydrolase activities in dementia with Lewy bodies and multiple system atrophy. *Mol Neurobiol*. 2022;59(4):2277-2287. doi 10.1007/S12035-021-02688-0

Usenko T.S., Senkevich K.A., Basharova K.S., Bezrukova A.I., Baydakova G.V., Tyurin A.A., Beletskaya M.V., Kulabukhova D.G., Grunina M.N., Emelyanov A.K., Miliukhina I.V., Timofeeva A.A., Zakharova E.Y., Pchelina S.N. LRRK2 exonic variants are associated with lysosomal hydrolase activities and lysosphingolipid alterations in Parkinson's disease. *Gene*. 2023;882:147639. doi 10.1016/j.gene.2023.147639

Usenko T.S., Timofeeva A., Beletskaia M., Basharova K., Baydakova G., Bezrukova A., Grunina M., Emelyanov A., Miliukhina I., Zakharova E., Pchelina S. The effect of p.G2019S mutation in the *LRRK2* gene on the activity of lysosomal hydrolases and the clinical features of Parkinson's disease associated with p.N370S mutation in the *GBA1* gene. *J Integr Neurosci*. 2024;23(1):16. doi 10.31083/j.jin2301016

Vincze T., Posfai J., Roberts R.J. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res*. 2003;31(13):3688-3691. doi 10.1093/nar/gkg526

Wang S., Ma Z., Xu X., Wang Z., Sun L., Zhou Y., Lin X., Hong W., Wang T. A role of Rab29 in the integrity of the trans-golgi network and retrograde trafficking of mannose-6-phosphate receptor. *PLoS One*. 2014;9(5):e96242. doi 10.1371/journal.pone.0096242

Wei Y., Awan M.U.N., Bai L., Bai J. The function of Golgi apparatus in LRRK2-associated Parkinson's disease. *Front Mol Neurosci*. 2023;16:1097633. doi 10.3389/fnmol.2023.1097633

Xia H., Luo Q., Li X.X., Yang X.L. Association between PARK16 gene polymorphisms and susceptibility of Parkinson's disease in a Chinese population. *Genet Mol Res*. 2015;14(2):2978-2985. doi 10.4238/2015.april.10.7

Ysselstein D., Nguyen M., Young T.J., Severino A., Schwake M., Merchant K., Krainc D. LRRK2 kinase activity regulates lysosomal glucocerebrosidase in neurons derived from Parkinson's disease patients. *Nat Commun*. 2019;10(1):5570. doi 10.1038/s41467-019-13413-w

Zhu H., Tonelli F., Turk M., Prescott A., Alessi D.R., Sun J. Rab29-dependent asymmetrical activation of leucine-rich repeat kinase 2. *Science*. 2023;382(6677):1404-1411. doi 10.1126/science.adi9926

Zimprich A., Biskup S., Leitner P., Lichtner P., Farrer M., Lincoln S., Kachergus J., Hulihan M., Uitti R.J., Calne D.B., Stoessl A.J. Mutations in *LRRK2* cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron*. 2004;44(4):601-607. doi 10.1016/j.neuron.2004.11.005

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

Received November 7, 2024. Revised January 31, 2025. Accepted June 30, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):828-837

doi 10.18699/vjgb-25-90

The contribution of FOXO family transcription factor genes (*FOXO1*, *FOXO3*) to chronic obstructive pulmonary disease

V.A. Markelov [1, 2], L.Z. Akhmadishina [1, 3], T.R. Nasibullin [1], Y.G. Aznabaeva [2], O.V. Kochetova [1, 2], N.N. Khusnutdinova [1], S.M. Izmailova [2], N.Sh. Zagidullin [2], G.F. Korytina [1, 2]

[1] Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
[2] Bashkortostan State Medical University, Ufa, Russia
[3] Ufa State Petroleum Technological University, Ufa, Russia
guly_kory@mail.ru

Abstract. Chronic obstructive pulmonary disease (COPD) is a multifactorial disease of the respiratory system and is the third leading cause of death worldwide. In the framework of the most relevant concepts of COPD pathogenesis, the key focus is on accelerated cellular senescence. FOXO family transcription factors are important hub components of cellular senescence signaling pathways. The objective of the study is to identify the association of *FOXO1* (rs12585277, rs9549240), and *FOXO3A* (rs2253310, rs3800231) genes polymorphic variants with COPD and disease phenotypes. DNA samples from COPD patients (N = 710) and healthy individuals (N = 655) were used, polymorphic loci were analyzed by real-time PCR. For the first time, significant associations of *FOXO1* (rs12585277) and *FOXO3A* (rs2253310) gene polymorphic loci with COPD and disease phenotypes were shown. Association with COPD was established with *FOXO1* (rs12585277) (P_{adj} = 0.0018, OR = 1.44 for the AG genotype) and *FOXO3A* (rs2253310) ($P_{adj} = 5.926 \times 10^{-7}$, OR = 1.99 for the GG genotype). A significant genotype-dependent variation of smoking index (in pack/years), vital capacity and forced vital capacity was revealed for *FOXO1* (rs9549240, rs12585277) and *FOXO3A* (rs2253310) loci. Multiple regression and ROC analysis identified highly informative COPD risk model, which included polymorphic variants of the *FOXO1* (rs12585277) and *FOXO3A* (rs2253310) genes, smoking index and age ($P = 5.25 \times 10^{-93}$, AUC = 0.864). The multivariate regression model of the COPD "frequent exacerbator" phenotype included the AG genotype of *FOXO1* (rs12585277), smoking index and age (AUC = 0.897, $P = 4.1 \times 10^{-86}$). FOXO family transcription factors involved in autophagy, oxidative stress and cellular homeostasis may provide a platform for a new diagnostic and treatment strategy for COPD as potential biomarkers and targets for therapy.

Key words: chronic obstructive pulmonary disease; cellular senescence; oxidative stress; FOX transcription factor family; *FOXO1*; *FOXO3*

For citation: Markelov V.A., Akhmadishina L.Z., Nasibullin T.R., Aznabaeva Y.G., Kochetova O.V., Khusnutdinova N.N., Izmailova S.M., Zagidullin N.Sh., Korytina G.F. The contribution of FOXO family transcription factor genes (*FOXO1*, *FOXO3*) to chronic obstructive pulmonary disease. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025; 29(6):828-837. doi 10.18699/vjgb-25-90

Funding. The study was supported by the Russian Scientific Foundation Grant No. 23-25-00019.

Вклад генов транскрипционных факторов семейства FOXO (*FOXO1*, *FOXO3*) в развитие хронической обструктивной болезни легких

В.А. Маркелов [1, 2], Л.З. Ахмадишина [1, 3], Т.Р. Насибуллин [1], Ю.Г. Азнабаева [2], О.В. Кочетова [1, 2], Н.Н. Хуснутдинова [1], С.М. Измайлова [2], Н.Ш. Загидуллин [2], Г.Ф. Корытина [1, 2]

[1] Институт биохимии и генетики – обособленное структурное подразделение Уфимского федерального исследовательского центра Российской академии наук, Уфа, Россия
[2] Башкирский государственный медицинский университет, Уфа, Россия
[3] Уфимский государственный нефтяной технический университет, Уфа, Россия
guly_kory@mail.ru

Аннотация. Хроническая обструктивная болезнь легких (ХОБЛ) – многофакторное заболевание дыхательной системы, является третьей ведущей (после ишемической болезни сердца и инсульта) причиной смерти в мире. В рамках наиболее актуальных концепций патогенеза ХОБЛ ключевое значение уделяется ускоренному клеточному старению. Транскрипционные факторы семейства FOXO – важные ключевые компоненты сигнальных путей клеточного старения. Цель исследования – выявление ассоциации полиморфных вариантов генов *FOXO1* (rs12585277, rs9549240) и *FOXO3A* (rs 2253310, rs3800231) с развитием ХОБЛ и фенотипами заболевания. В рабо-

те использованы образцы ДНК больных ХОБЛ ($N = 710$) и здоровых индивидов ($N = 655$). Полиморфные локусы анализировали методом ПЦР в реальном времени. Впервые показаны значимые ассоциации полиморфных локусов генов *FOXO1* (rs12585277) и *FOXO3A* (rs2253310) с ХОБЛ и фенотипами заболевания. Ассоциация с ХОБЛ установлена с локусами генов *FOXO1* (rs12585277) ($P_{adj} = 0.0018$, OR = 1.44 генотип AG) и *FOXO3A* (rs2253310) ($P_{adj} = 5.926 \times 10^{-7}$, OR = 1.99 генотип GG). Обнаружена вариабельность показателей индекса курения (в пачках/лет), жизненной емкости легких и форсированной жизненной емкости легких в зависимости от полиморфных вариантов локусов *FOXO1* (rs9549240 и rs12585277) и *FOXO3A* (rs2253310). Идентифицирована значимая многофакторная регрессионная модель формирования ХОБЛ, в которую вошли полиморфные варианты генов *FOXO1* (rs12585277) и *FOXO3A* (rs2253310), индекс курения и возраст обследуемых ($P = 5.25 \times 10^{-93}$, AUC = 0.864). Многофакторная регрессионная модель развития фенотипа ХОБЛ с частыми обострениями включала генотип AG локуса *FOXO1* (rs12585277), индекс курения и возраст (AUC = 0.897, $P = 4.1 \times 10^{-86}$). Связанные с аутофагией, окислительным стрессом и клеточным гомеостазом транскрипционные факторы семейства FOXO как потенциальные биомаркеры и мишени для терапии могут стать основой для разработки новой стратегии диагностики и лечения хронической обструктивной болезни легких.

Ключевые слова: хроническая обструктивная болезнь легких (ХОБЛ); клеточное старение; аутофагия; окислительный стресс; транскрипционные факторы семейства FOXO; *FOXO1*; *FOXO3*

Introduction

Chronic obstructive pulmonary disease (COPD) is a complex respiratory disease affecting the distal respiratory tract and pulmonary parenchyma with the development of pulmonary emphysema (Agustí et al., 2023). COPD is the third leading cause of death in the world, which explains the ongoing search for new approaches to diagnosis, treatment and prevention of the disease development (Agustí et al., 2023). Although intensive research has been conducted on both the molecular background and various clinical aspects of COPD, the mechanisms underlying the pathogenesis of COPD and different disease phenotypes remain incompletely understood (Brandsma et al., 2020).

COPD development results from exposure to a complex of risk factors over a long period of time, with tobacco smoking being the main one. Cigarette smoke exposure to airways leads to oxidative stress and activation of inflammatory cells and lung tissue damage (Domej et al., 2014). The most actual framework for COPD pathogenesis focuses on accelerated cellular senescence (Luo et al., 2024). As a fundamental mechanism for maintaining tissues and organs homeostasis, cellular senescence is being mediated by multiple processes. The most important of them include DNA damage, telomere loss, mitochondrial dysfunction, and autophagy and proteostasis alterations (Li et al., 2024). Previously, we have demonstrated the contribution of sirtuin family genes and the PI3K/AKT signaling cascade to COPD (Korytina et al., 2023). The FOXO transcription factors are major targets of the PI3K/AKT signaling cascade, exerting insulin-dependent regulation of cellular metabolism (Farhan et al., 2020). The mammalian FOXO class of transcription factors currently includes four proteins: FOXO1 (FKHR), FOXO3a (FKHRL1), FOXO4 (AFX), and FOXO6 (Santos et al., 2023). The FOXO transcription factors regulate the expression of a number of genes for antioxidant defense, cell cycle and apoptosis, proliferation, metabolism, and are involved in repression of mitochondrial respiratory chain proteins (Hagenbuchner et al., 2013).

The aim of this study was to identify the association of FOXO family transcription factor gene polymorphic variants (*FOXO1*, *FOXO3*) with COPD and disease phenotypes.

Materials and methods

This study was designed according to the "case-control" principle. DNA samples from unrelated individuals, ethnic Tatars, who were residents of the Republic of Bashkortostan, were used in the study. The work was approved by the Ethics Committee of the Institute of Biochemistry and Genetics of the UFRC RAS (Protocol No. 19, dated November 1, 2022). Informed voluntary consent for the use of biological material in the study was obtained from all the participants. The inclusion and exclusion criteria had been described previously (Korytina et al., 2019). The diagnosis of COPD was established according to recommendations of the working group on the "Global Strategy for the Diagnosis, Treatment, and Prevention of Chronic Obstructive Pulmonary Disease" (http://goldcopd.org) on the basis of clinical and laboratory instrumental studies, including high-resolution computed tomography and spirometry.

The control group included unrelated individuals with no history of chronic diseases, including respiratory diseases and acute respiratory diseases at the time of biomaterial collection, matched for sex, age, smoking status, exposure to risk factors, region of residence, and ethnicity. The inclusion criteria in the control group were the normal parameters of respiratory function ($FEV_1/FVC > 70$ %, $FEV_1 > 80$ %) and age over 45 years. To identify genetic markers associated with COPD phenotypes, we performed a comparison

between the control group and patients differentiated by disease phenotype as we described previously (Korytina et al., 2020) (Supplementary Table S1)[1]. Two phenotypes were distinguished: group 1 – COPD with frequent exacerbations; group 2 – patients with rare exacerbations. In the Table S1, the characteristics of the groups are summarized.

Genotyping. DNA was isolated from peripheral blood leukocytes by phenol-chloroform extraction. The polymorphic loci of the *FOXO1* (rs12585277, rs9549240) and *FOXO3A* (rs2253310, rs3800231) genes were selected; the functional significance of SNPs was analyzed by means of RegulomeDB Version 1.1 (https://regulomedb.org), SNPinfo Web Server (https://snpinfo.niehs.nih.gov), and HaploReg v3 (Ward, Kellis, 2016) (Table S2). Bioinformatics analysis indicated that the selected SNPs affected gene function in different tissue types or were in linkage disequilibrium with functional loci of the gene. SNPs were analyzed using real-time polymerase chain reaction (PCR) with commercially available assays with fluorescent detection (https://www.oligos.ru, DNA Synthesis LLC, Russia) using BioRad CFX96™ (Bio-Rad Laboratories, Inc, USA). The methods of analysis had been previously described by our group in detail (Korytina et al., 2019)

Statistical analyses. The description of standard statistical analysis methods was provided previously (Korytina et al., 2019). Analysis of deviation of obtained genotype frequencies from the Hardy–Weinberg equilibrium and the association of SNPs with the disease in the basic allele test and in regression models (dominant, recessive, additive (per rare allele dose – increase in the rare allele dose in the series: homozygote for frequent allele (0) – heterozygote (1) – homozygote for rare allele (2)) and overdominance (for heterozygotes)) were performed using the SNPassoc package v. 2.0–2 for R (González et al., 2007). An SNP was considered to be associated at $P < 0.05$; correction for multiple comparisons was conducted using the assessment method for the proportion of received false-positive results, false discovery rate (FDR), via online tool (https://tools.carbocation.com/FDR). The haplotype frequencies and linkage disequilibrium structure LD (D′, r^2)) were calculated with Haploview 4.2. The contribution of polymorphic variants of the studied genes to the variability of quantitative traits indicating the severity of airway obstruction (lung function parameters – vital capacity (VC), forced vital capacity (FVC), forced expiratory volume in the first second (FEV_1)) and smoking index was determined using linear regression. The logistic regression approach with stepwise imputation of independent variables was used to create multivariate regression models. The SNPs of the studied genes, as well as clinical and demographic parameters (age, smoking index, smoking status, sex) were selected as independent variables.

A regression model is an equation in which the dependent variable is represented as a function of the independent variables (predictors). The log-regression equation has the form:

$$Y = \text{logit}\,(P) = \ln(P/(1-P)) = \\ = \beta 0 + \beta 1.X1 + \beta 2.X2 + \ldots\ \beta i.Xi,$$

where Y is the dependent variable (status: 0 – control, 1 – case), X is the independent variable, β0 is a constant, β is the regression coefficient for the independent variable, $X1...i$ is the value of the independent variable. The exponent of the individual regression coefficient β was interpreted as the odds ratio (OR). The hypothesis of significance of the whole model, taking into account all independent variables, was verified using the likelihood ratio (LR) test. To evaluate the effectiveness of the obtained prognostic model, the Nagelkerke R^2 value and receiver operating characteristic (ROC) curve analysis were performed. The capacity of the regression model to discriminate between positive and negative cases (patient or healthy status) was assessed by the area under the ROC curve (AUC); the AUC value ranged from 0.5 (no discriminatory capacity of the model) to 1.0 (perfect discriminatory capacity). Calculations were carried out using SPSS v. 22.

Results

The analysis of polymorphic loci of the *FOXO1* (rs12585277, rs9549240) and *FOXO3A* (rs2253310, rs3800231) genes was carried out in the formed groups of COPD patients and the control group. The observed genotype frequencies of all examined SNPs in the control group were in accordance with the Hardy–Weinberg test: *FOXO1* (rs12585277) (P_{H-W} = 0.597), *FOXO1* (rs9549240) (P_{H-W} = 0.341), and *FOXO3A* (rs2253310) (P_{H-W} = 0.3191), (rs3800231) (P_{H-W} = 0.3831). Subsequently, we estimated the statistical significance of differences in distribution of allele and genotype frequencies between the groups, and the odds ratio values for the minor allele of each locus were calculated (basic allelic test) (Table 1).

At the next stage, logistic regression was used to analyze the association of SNPs taking into account quantitative and binary traits (sex, age, smoking status and index), which were introduced into the regression equation as independent variables (Table 1).

Statistically significant differences in genotype frequency distributions between groups for *FOXO1*

[1] Supplementary Tables S1–S5 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Markelov_Engl_29_6.pdf

Table 1. Allele and genotype frequencies of the examined polymorphic loci of the *FOXO1* and *FOXO3A* genes in the groups of COPD patients and healthy individuals, association analysis with COPD

Gene RefSNP	Minor allele	Genotypes, alleles model	COPD *n* (%) (*N* = 710)	Control *n* (%) (*N* = 655)	P_{adj}	$P_{cor\text{-}FDR}$	OR_{adj} (95 % CI)
FOXO1 rs12585277 G>A	A	GG/GA/AA	202/398/110 (28.45/56.06/15.49)	236/307/112 (36.03/46.87/17.10)	0.002	0.006	–
		G/A	802/618 (56.48/43.52)	779/531 (59.47/40.53)	0.123	0.208	0.88 (0.76–1.03)
		GG GA+AA Dominant	202 (28.45) 508 (71.55)	236 (36.03) 419 (63.97)	0.003	0.008	1.00 1.42 (1.11–1.82)
		GG+AA GA	312 (43.94) 398 (56.06)	348 (53.13) 307 (46.87)	0.0018	0.006	1.00 1.44 (1.14–1.81)
		GG+GA AA Recessive	600 (84.51) 110 (15.49)	543 (82.9) 112 (17.10)	0.5	0.54	1.00 0.90 (0.66–1.23)
		Log-additive	–	–	0.12	0.208	1.15 (0.97–1.36)
FOXO1 rs9549240 T>G	T	GG/GT/TT	414/244/52 (58.31/34.37/7.32)	356/240/59 (54.35/36.64/9.01)	0.32	0.416	–
		G/T	1072/348 (75.49/24.51)	952/358 (72.67/27.33)	0.101	0.202	1.15 (0.97–1.37)
		GG GT+TT Dominant	414 (58.31) 296 (41.69)	356 (54.35) 299 (45.65)	0.17	0.209	1.00 0.85 (0.67–1.07)
		GG+TT GT	466 (65.66) 244 (34.37)	415 (63.36) 240 (36.64)	0.43	0.516	1.00 0.79 (0.52–1.21)
		GG+GT TT Recessive	658 (92.68) 52 (7.32)	596 (90.99) 59 (9.01)	0.52	0.54	1.00 0.85 (0.67–1.07)
		Log-additive	–	–	0.13	0.208	0.87 (0.73–1.04)
FOXO3A rs3800231 A>G	A	GG/GA/AA	184/380/146 (25.92/53.52/20.56)	199/336/120 (30.38/51.30/18.32)	0.22	0.31	–
		G/A	748/672 (52.68/47.32)	734/576 (56.03/43.97)	0.86	0.86	0.87 (0.75–1.02)
		GG GA+AA Dominant	184 (25.92) 526 (74.08)	199 (30.38) 456 (69.62)	0.096	0.202	1.00 1.25 (0.96–1.63)
		GG+AA GA	330 (46.48) 380 (53.52)	319 (48.70) 336 (51.30)	0.47	0.537	1.00 1.09 (0.86–1.38)
		GG+GA AA Recessive	564 (79.44) 146 (20.56)	535 (81.68) 120 (18.32)	0.33	0.416	1.00 1.16 (0.86–1.57)
		Log-additive	–	–	0.097	0.173	1.16 (0.97–1.38)
FOXO3A rs2253310 C>G	C	GG/GC/CC	572/129/9 (80.56/18.17/1.27)	447/183/25 (68.24/27.94/3.82)	9.733×10^{-7}	5.8398×10^{-6}	–
		G/C	1273/147 (89.65/10.35)	1077/233 (82.21/17.79)	1.54×10^{-7}	1.848×10^{-6}	0.53 (0.43–0.67)
		GG GC+CC Dominant	572 (80.56) 138 (19.44)	447 (68.24) 208 (31.76)	5.926×10^{-7}	4.7408×10^{-6}	1.00 0.52 (0.40–0.67)
		GG+CC GC	581 (81.83) 129 (18.17)	472 (72.06) 183 (27.94)	4.111×10^{-5}	0.0001	1.00 0.57 (0.44–0.75)
		GG+GC CC Recessive	701 (98.73) 9 (1.27)	630 (96.18) 25 (3.82)	0.0003	0.0012	1.00 0.32 (0.14–0.72)
		Log-additive	–	–	1.451×10^{-7}	1.848×10^{-6}	0.54 (0.43–0.68)

Note. *P* – significance level of differences in allele and genotype frequencies between the groups (χ^2 test for homogeneity of samples); regression analysis: number of individuals included in regression analysis (*N* = 1,365); P_{adj} – significance of the likelihood ratio test of the log-regression model controlling for sex, age, and smoking status and index; OR_{adj} – odds ratio controlling for all these factors, 95 % CI – 95 % confidence interval for OR; $P_{cor\text{-}FDR}$ – test significance after correction.

(rs12585277) (P = 0.002) were found. The association with COPD was established in a dominant model (P_{adj} = 0.003, OR = 1.42), the frequency of the heterozygous genotype in the COPD patient group reached 56.06 % versus 46.87 % in controls (P_{adj} = 0.0018, OR = 1.44).

Significant differences in the distribution of genotype and allele frequencies between the patients group and healthy individuals were found for the *FOXO3A* (rs2253310) ($P = 9.733 \times 10^{-7}$ and $P = 1.54 \times 10^{-7}$, respectively). Association with COPD was found in dominant ($P_{adj} = 5.926 \times 10^{-7}$, OR = 0.52), recessive ($P_{adj} = 3.360 \times 10^{-3}$, OR = 0.32), log-additive models ($P_{adj} = 1.451 \times 10^{-7}$, OR = 0.54), and with the heterozygous GC genotype ($P_{adj} = 4.111 \times 10^{-5}$, OR = 0.57). Notably, the frequent G allele ($P = 1.54 \times 10^{-7}$, OR = 1.87 95 % CI 1.50–2.33) and the GG genotype ($P_{adj} = 5.926 \times 10^{-7}$, OR = 1.99 95 % CI 1.57–2.47) were more often observed in the COPD group.

There was no statistically significant difference between the groups of patients and healthy individuals for the *FOXO1* (rs9549240) and *FOXO3A* (rs3800231) polymorphic loci.

Haplotype analysis of *FOXO1* and *FOXO3A* genes polymorphic loci

The linkage disequilibrium between loci rs12585277 and rs9549240 of the *FOXO1* gene ($D' = 0.6183$, $r^2 = 0.429$) and statistically significant differences in the pattern of haplotype frequency distribution of the *FOXO1* gene between the group of COPD patients and controls (P = 0.045) were revealed (Table S3). The frequency of the A-G haplotype at rs12585277 and rs9549240 loci was significantly higher in the group of COPD patients (24.26 *vs.* 18.81 % in controls, P_{adj} = 0.011, OR = 1.33).

The linkage disequilibrium between polymorphic loci rs3800231 and rs2253310 of the *FOXO3A* gene was observed ($D' = 0.3315$, $r^2 = 0.1452$) (Table S3). Statistically significant differences in the haplotype frequency distribution pattern were established between the group of COPD patients and controls (P = 0.00001). The frequency of haplotype A-G at rs3800231 and rs2253310 loci was significantly higher in the COPD group (40.27 *vs.* 32.56 % in controls, P_{adj} = 0.03, OR = 1.25), whereas haplotypes A-C and C-C were more frequent in the group of healthy individuals (P_{adj} = 0.0076, OR = 0.65 and P_{adj} = 0.0072, OR = 0.51, respectively).

Association analysis of polymorphic variants of the *FOXO1* and *FOXO3A* genes with different COPD phenotypes

In order to identify genetic markers associated with COPD phenotypes, we performed a comparison of the control group and patients differentiated by disease phenotypes according to the GOLD classification (http://goldcopd.org); this classification included an integral assessment of the COPD phenotype taking into account the number of exacerbations per year, results of COPD assessment test (CAT) and medical research council dyspnea scale (MRC), and lung function parameters (Table 2).

Statistically significant associations with the COPD phenotype with frequent exacerbations were obtained for *FOXO1* (rs12585277) in the dominant model (P_{adj} = 0.001, OR = 1.65) and for the heterozygous genotype (P_{adj} = 0.0005, OR = 1.653); and for *FOXO3A* (rs2253310) in the dominant (P_{adj} = 0.0002, OR = 0.55) and log-additive models (P_{adj} = 0.0001, OR = 0.58). It should be noted that the homozygotes for the frequent allele G of *FOXO3A* (rs2253310) were more frequent in patients (OR = 1.81 95 % CI 1.32–2.49). The frequency of the A-G haplotype at rs12585277 and rs9549240 loci of the *FOXO1* gene was higher in the COPD group (24.29 *vs.* 18.81 % in controls, P_{adj} = 0.022, OR = 1.36 95 % CI 1.05–1.76) (Table S4).

In COPD patients with rare exacerbations, a significant association was confirmed only for *FOXO3A* (rs2253310) in the dominant (P_{adj} = 0.00001, OR = 0.48) and log-additive (P_{adj} = 0.00001, OR = 0.5) models. The GG genotype of *FOXO3A* (rs2253310) (OR = 2.09 95 % CI 1.53–2.88) and the A-G haplotype at rs3800231 and rs2253310 loci of *FOXO3A* (P_{adj} = 0.032, OR = 1.31 95 % CI 1.02–1.67) were more frequent in the patient group.

Association of *FOXO1* and *FOXO3A* polymorphic loci with lung function parameters and smoking index

Smoking is a major risk factor for COPD and a trigger for the oxidative stress that results in DNA damage and cell apoptosis (Domej et al., 2014). The analysis of the quantitative parameter indicating the smoking intensity and history (smoking index) in the smokers' group, including both patients and healthy individuals, depending on the polymorphic variants of the *FOXO1* and *FOXO3A* genes was performed (Table S5). It has been established that the GT genotype of *FOXO1* (rs9549240) and the GG genotype of *FOXO3A* (rs2253310) are associated with higher smoking index values (P = 0.0042 and P = 0.012).

Lung function parameters including vital capacity (VC), forced vital capacity (FVC), forced expiratory volume in the first second (FEV_1), ratio of forced expiratory volume in 1 s to vital capacity (FEV_1/FVC) are key clinical variables indicating the degree of airway obstruction in COPD and disease progression. The indi-

Table 2. Results of association analysis of *FOXO1* and *FOXO3A* gene polymorphic loci in groups differentiated by COPD phenotypes

Gene, SNP	Minor allele	Genotype, model	OR_{adj} (95 % CI)	P_{adj}	$P_{cor\text{-}FDR}$
COPD with rare exacerbations (N = 991)					
FOXO1 rs12585277 G>A	A	GG GA+AA Dominant	1.00 1.22 (0.93–1.60)	0.16	0.18
		Log-additive	1.06 (0.89–1.28)	0.51	0.51
FOXO1 rs9549240 T>G	T	GG GT+TT Dominant	1.00 0.80 (0.61–1.04)	0.099	0.148
		Log-additive	0.84 (0.68–1.03)	0.094	0.148
FOXO3A rs3800231 A>G	A	GG GA+AA Dominant	1.00 1.27 (0.94–1.72)	0.12	0.154
		Log-additive	1.20 (0.99–1.46)	0.063	0.148
FOXO3A rs2253310 C>G	C	GG GC+CC Dominantя	1.00 0.48 (0.35–0.65)	4.961×10^{-7}	2.232×10^{-6}
		Log-additive	0.50 (0.38-0.67)	1.373×10^{-7}	1.235×10^{-6}
COPD with frequent exacerbations (N = 1,029)					
FOXO1 rs12585277 G>A	A	GG GA+AA Dominant	1.00 1.65 (1.22–2.24)	0.001	0.0022
		GG+AA GA	1.00 1.63 (1.24–2.15)	0.0005	0.0015
		Log-additive	1.32 (1.12–1.49)	0.045	0.081
FOXO1 rs9549240 T>G	T	GG GT+TT Dominant	1.00 0.90 (0.68–1.18)	0.44	0.44
		Log-additive	0 0.90 (0.73–1.11)	0.33	0.371
FOXO3A rs3800231 A>G	A	GG GA+AA Dominant	1.00 1.24 (0.90–1.71)	0.19	0.28
		Log-additive	0 1.12 (0.90–1.38)	0.31	0.371
FOXO3A rs2253310 C>G	C	GG GC+CC Dominant	1.00 0.55 (0.40–0.76)	0.0002	0.0009
		Log-additive	0.58 (0.44–0.77)	0.0001	0.0009

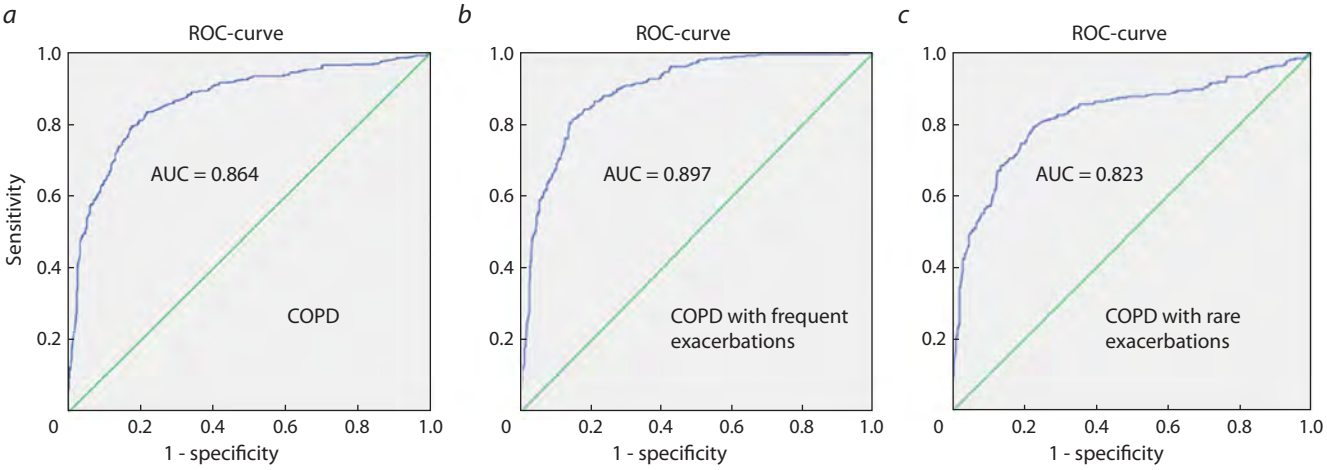
Note. N – number of individuals included in regression analysis; P_{adj} – significance of the likelihood ratio test of the log-regression model controlling for sex, age, and smoking status and index; OR_{adj} – odds ratio controlling for all these factors, 95 % CI – 95 % confidence interval for OR; $P_{cor\text{-}FD}$ – test significance after correction.

viduals with the homozygous GG genotype of *FOXO1* (rs9549240) had lower VC values (P = 0.0071); the carriers of the A allele in the homozygous and heterozygous state (dominant model) of *FOXO1* (rs12585277) had lower FVC values (P = 0.04).

Multiple regression and ROC analysis

At the final stage, using multiple regression analysis with a stepwise forward inclusion of predictors followed by ROC analysis, a search for complex clinical-genetic models of COPD development was carried out. The genotypes or alleles of the studied genes were selected as independent variables, subsequently clinical and demographic variables (sex, age, smoking status and smoking index) were added and the most significant multivariate regression models were selected. The lung function parameters were excluded, as they are classical and well-identified predictors of COPD and disease severity.

The informative regression model of COPD development included polymorphic variants of *FOXO1* (rs12585277) (AG genotype), and *FOXO3A* (rs2253310) (C allele) genes, smoking index and age of the subjects ($P = 5.25 \times 10^{-93}$) (Table 3). The ROC analysis of the model demonstrated its high ability to discriminate between COPD patients and healthy individuals (AUC = 0.864, sensitivity – 78.3 %, specificity – 82.3 %) (see the Figure *a*).



Area under the curve (ROC curve) to evaluate the effectiveness of prognostic regression models of COPD development (*a*); COPD phenotype with frequent exacerbations (*b*); COPD phenotype with rare exacerbations (*c*) including significant clinical, demographic and genetic predictors.

AUC is the area under the curve. Full model characteristics are presented in Table 3.

Table 3. The predictive regression models for COPD development

Independent variable	β	P_{wald}	OR	95 % CI_{OR}
COPD				
Smoking index	0.04	4×10^{-20}	1.04	1.03–1.05
Age	0.08	1.55×10^{-31}	1.08	1.07–1.1
FOXO1 (rs12585277) AG	0.33	0.0415	1.39	1.01–1.91
FOXO3A (rs2253310) C	–0.46	0.0037	0.62	0.46–0.86
Intercept	–5.66	9.4×10^{-41}	0.003	
$\chi^2 = 435.74$ $P = 5.25\times10^{-93}$ AUC = 0.864 (sensitivity – 78.3 %, specificity – 82.3 %, $R^2 = 0.483$)				
COPD with frequent exacerbations				
Smoking index	0.050	1.96×10^{-18}	1.05	1.04–1.06
Age	0.115	2.46×10^{-30}	1.12	1.1–1.14
FOXO1 (rs12585277) AG	0.530	0.0123	1.7	1.12–2.57
Intercept	–8.756889	6.5×10^{-42}	0.0002	
$\chi^2 = 398.76$ $P = 4.1\times10^{-86}$ AUC = 0.897 (sensitivity – 69.6 %, specificity – 90.9 %, $R^2 = 0.57$)				
COPD with rare exacerbations				
Smoking index	0.035	8.4×10^{-14}	1.04	1.03–1.05
Age	0.064	5.7×10^{-15}	1.07	1.05–1.08
FOXO3A (rs2253310) C	–0.469	0.0168	0.63	0.43–0.92
Intercept	–4.992	4.7×10^{-26}	0.01	
$\chi^2 = 225.29$ $P = 1.4\times10^{-48}$ AUC = 0.823 (sensitivity – 54.5 %, specificity – 91.7 %, $R^2 = 0.384$)				

Note. β – the beta coefficient for the independent variable; P_{wald} – the significance for Wald statistics (Wald statistic is the regression coefficient divided by the square of the standard error: β/SE^2), indicates the significance of the independent variable; OR – odds ratio, represents the exponent of the beta coefficient (exp β) for the independent variable; χ^2 – the likelihood ratio test (LR), is necessary to test the hypothesis of the significance of the regression model, taking into account all independent variables; *P* – the value for the likelihood ratio test; R^2 – Nagelkerke R^2 model quality indicator – reflects the proportion of variability in the trait; intercept for the regression equation, the value of the dependent variable at which the independent variable is equal to zero; AUC – area under the curve; sensitivity – the proportion of correctly classified patients with a given diagnosis; specificity – the proportion of correctly classified healthy individuals. The ROC curves are presented in the Figure.

A significant multifactorial regression model of COPD phenotype with frequent exacerbations included the AG genotype of *FOXO1* (rs12585277), smoking index and age; the regression model was characterized by a high ability to discriminate between patients and healthy individuals (AUC = 0.897, $P = 4.1 \times 10^{-86}$). However, sensitivity, a measure of the model's ability to correctly classify patients with a given COPD phenotype from healthy individuals, was only 69.6 %; at the same time, the model accurately classified healthy individuals, as it possessed high specificity (90.9 %) (see the Figure *b*).

The multifactorial regression model for COPD phenotype with rare exacerbations included the C allele of *FOXO3A* (rs2253310), smoking index and age ($P = 1.4 \times 10^{-48}$). The ROC analysis of the obtained model demonstrated its moderate discriminatory ability (AUC = 0.823), low sensitivity (54.5 %), but high specificity (91.7 %) (see the Figure *c*).

Discussion

The association of *FOXO1* (rs12585277) with COPD and the COPD phenotype with frequent exacerbations was established; haplotype A-G at rs12585277 and rs9549240 of *FOXO1* was more frequent in the general COPD group and among patients with frequent exacerbations. The *FOXO1* (rs12585277) AG genotype is part of multivariate regression model of COPD development and the COPD phenotype with frequent exacerbations, as well as predictors such as age and smoking index. A decline in lung function values indicating the severity of airway obstruction was revealed: FVC in carriers of homozygous and heterozygous genotype for the rare A allele of *FOXO1* (rs12585277), and VC in carriers of the GG genotype of the *FOXO1* locus (rs9549240). The variability of smoking index parameter was determined depending on *FOXO1* (rs9549240) genotypes.

FOXO1 is located on chromosome 13q14.11 (https://www.ncbi.nlm.nih.gov/gene/2308). Previously, a number of polymorphic loci of *FOXO1* have been shown to be associated with type 2 diabetes mellitus and obesity (Hussain et al., 2022; Santana et al., 2024). The studies on the association of polymorphic variants of the *FOXO1* gene with COPD have not been conducted yet. T. Xue et al. (2024) have identified increased levels of FOXO1 mRNA and protein in the lung tissue of mice with lung emphysema model. Meanwhile, FOXO1 expression has been shown to be downregulated in the blood of COPD patients (Zhu et al., 2020).

The most significant associations with COPD were observed with *FOXO3A* (rs2253310); the G allele and the GG genotype were more frequent in the patient group. This association maintained significance regardless of the disease phenotype. The GG genotype was associated with an increased smoking index among all smokers; this quantitative measure characterizes smoking duration and intensity. The increase in smoking index is a major risk factor for COPD; the obtained results may be related to a higher prevalence of individuals with a high smoking index (more than 40 pack/years) among COPD patients with the GG genotype of *FOXO3A (*rs2253310). According to the multiple regression analysis results, the rs2253310 locus is a significant predictor for COPD development in general, in addition to *FOXO1* (rs12585277), age and smoking index.

The *FOXO3A* gene is located on chromosome 6q21 (https://www.ncbi.nlm.nih.gov/gene/2309). The association of polymorphic variants in the FOXO3A gene with longevity (Soerensen et al., 2015) and a number of age-associated diseases (Klinpudtan et al., 2022; Cao et al., 2023) has been shown. Association studies of *FOXO3A* gene polymorphic loci with COPD have not been previously conducted.

Systemic effects are typical for COPD, resulting in the development of severe complications that additionally exacerbate the disease progression in some patients (Agustí et al., 2023). Due to changes in the COPD diagnosis and prevention strategy (http://goldcopd.org), a great attention of researchers is currently focused on the identification of different disease phenotypes markers and effective determination of the patients with frequent exacerbations, as this category of COPD patients is characterized by a rapid progression of airway obstruction and increased mortality (Geerdink et al., 2016). It was shown that heterozygous genotype AG of *FOXO1* (rs12585277) and genotype GG of *FOXO3A (*rs2253310) were significantly more frequent in the group of COPD patients with frequent exacerbations. The results of multivariate regression analysis demonstrated that the regression model for discriminating COPD patients with frequent exacerbations from healthy individuals had the highest efficiency parameters (such as AUC and R^2 levels). It can be explained by the greater homogeneity of the patients' group with this phenotype. The obtained data suggest that the most informative genetic marker for COPD phenotype with frequent exacerbations, among those that showed an association, is the AG genotype of *FOXO1* (rs12585277).

The association of *FOXO1* and *FOXO3A* gene polymorphic loci with COPD or disease phenotypes has not been studied before. For the first time, we investigated the role of genes encoding FOXO transcription factors in the disease development. The current interest is based on the fact that FOXOs regulate the expression of pro-

teins related to autophagy, oxidative stress and cellular metabolism (Hagenbuchner, Ausserlechner, 2013; Gui, Burgering, 2022).

The activation of FOXO1 has been demonstrated to suppress oxidative stress-induced apoptosis of epithelial cells in a model of bronchopulmonary dysplasia (Zang et al., 2023). FOXO1 stimulates the expression ofa 150 kDa oxygen-regulated protein (ORP150) and in this way protects airway epithelial cells from endoplasmic reticulum stress mediated by cigarette smoke exposure (Liu et al., 2018). The inhibition of FOXO1 stimulates the related processes of autophagy and endoplasmic reticulum stress (Guo et al., 2022) and induces phenotypic conversion of pulmonary macrophages, which contributes to inflammation and airway remodeling (Chung et al., 2019).

It has been shown that FOXO3 activity suppresses cellular senescence and pathological airway remodeling induced by cigarette smoke exposure (Yao et al., 2012); on the other hand, inhibition of FOXO3 expression promotes the accumulation of NF-kB in the nucleus and stimulates its pro-inflammatory transcriptional activity (Di Vincenzo et al., 2018). All these processes are important pathogenetic mechanisms contributing to COPD development. The oxidative stress induced by cigarette smoke stimulates the transcriptional activity of FOXO3, leading to the activation of FOXO1 expression and stimulating its binding to the promoters of autophagy protein genes (*ATG5*, *ATG12*, *ATG16*), beclin protein 1 (*BECLIN1*) and microtubule-associated protein 3 alpha light chain 1 (*LC3*) genes (Bagam et al., 2021). The antioxidant function of FOXO3 is determined by increased expression of the *SOD2*, *CAT,* and *GPX1* genes, which play a key role in the regulation of reactive oxygen species (ROS) homeostasis in lung cells as a response to oxidative stress (Mahlooji et al., 2022).

H. Jiang et al. (2023) established that activation of FOXO3A under stress factors leads to cellular adaptation and reduced cellular senescence, whereas suppression of FOXO3A activity is associated with greater mitochondrial injury in pulmonary epithelial cells. The insufficient level of FOXO3, which may be associated with functional polymorphic variants as well, leads to suppression of antioxidant gene expression, resulting in the development of oxidative stress as a response to cigarette smoke exposure (Hwang et al., 2011).

Therefore, FOXO transcription factors play a key role in the normal functioning of mitochondria, preventing the oxidative stress and in this way inhibiting the progression of lung epithelial cellular senescence (Chen et al., 2021), a critical pathogenetic mechanism of COPD development.

Conclusion

For the first time, significant associations of *FOXO1* (rs12585277) and *FOXO3A* (rs2253310) gene polymorphic loci with COPD and disease phenotypes have been shown. The FOXO family transcription factors related to autophagy, oxidative stress and cellular homeostasis, as potential biomarkers and targets for therapy, may provide a platform for a new diagnostic and treatment strategy for COPD.

References

Agustí A., Celli B.R., Criner G.J., Halpin D., Anzueto A., Barnes P., Bourbeau J. Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary. *Eur Respir J.* 2023;61(4): 230-239. doi 10.1183/13993003.00239-2023

Bagam P., Kaur G., Singh D.P., Batra S. In vitro study of the role of FOXO transcription factors in regulating cigarette smoke extract-induced autophagy. *Cell Biol Toxicol*. 2021;37(4):531-553. doi 10.1007/s10565-020-09556-y

Brandsma C.A., Van den Berge M., Hackett T.L., Brusselle G., Timens W. Recent advances in chronic obstructive pulmonary disease pathogenesis: from disease mechanisms to precision medicine. *J Pathol.* 2020;250(5):624-635. doi 10.1002/path.5364

Cao G., Lin M., Gu W., Su Z., Duan Y., Song W., Liu H., Zhang F. The rules and regulatory mechanisms of FOXO3 on inflammation, metabolism, cell death and aging in hosts. *Life Sci*. 2023;328:121877. doi 10.1016/j.lfs.2023.121877

Chen J.X., Yang L., Sun L., Chen W., Wu J., Zhang C.F., Liu K.Y., Bai L., Lu H.G., Gao T., Tian H., Jiang S.L. Sirtuin 3 ameliorates lung senescence and improves type II alveolar epithelial cell function by enhancing the FoxO3a-dependent antioxidant defense mechanism. *Stem Cells Dev*. 2021;30(17):843-855. doi 10.1089/scd.2021.0099

Chung S., Kim J.Y., Song M.A., Park G.Y., Lee Y.G., Karpurapu M., Englert J.A., Ballinger M.N., Pabla N., Chung H.Y., Christman J.W. FoxO1 is a critical regulator of M2-like macrophage activation in allergic asthma. *Allergy.* 2019;74(3):535-548. doi 10.1111/all.13626

Di Vincenzo S., Heijink I.H., Noordhoek J.A., Cipollina C., Siena L., Bruno A., Ferraro M., Postma D.S., Gjomarkaj M., Pace E. SIRT1/FoxO3 axis alteration leads to aberrant immune responses in bronchial epithelial cells. *J Cell Mol Med*. 2018;22(4):2272-2282. doi 10.1111/jcmm.13509

Domej W., Oettl K., Renner W. Oxidative stress and free radicals in COPD – implications and relevance for treatment. *Int J Chron Obstruct Pulmon Dis.* 2014;9:1207-1224. doi 10.2147/COPD.S51226

Farhan M., Silva M., Xingan X., Huang Y., Zheng W. Role of FOXO transcription factors in cancer metabolism and angiogenesis. *Cells.* 2020;9(7):1586. doi 10.3390/cells9071586

Geerdink J.X., Simons S.O., Pike R., Stauss H.J., Heijdra Y.F., Hurst J.R. Differences in systemic adaptive immunity contribute to the 'frequent exacerbator' COPD phenotype. *Respir Res.* 2016;17(1):140. doi 10.1186/s12931-016-0456-y

González J.R., Armengol L., Solé X., Guinó E., Mercader J.M., Estivill X., Moreno V. SNPassoc: an R package to perform whole genome association studies. *Bioinformatics.* 2007;23(5):644-645. doi 10.1093/bioinformatics/btm025

Gui T., Burgering B.M.T. FOXOs: masters of the equilibrium. *FEBS J.* 2022;289(24):7918-7939. doi 10.1111/febs.16221

Guo J., Nie J., Chen Z., Wang X., Hu H., Xu J., Lu J., Ma L., Ji H., Yuan J., Xu B. Cold exposure-induced endoplasmic reticulum stress regulates autophagy through the SIRT2/FoxO1 signaling pathway. *J Cell Physiol.* 2022;237(10):3960-3970. doi 10.1002/jcp.30856

Hagenbuchner J., Ausserlechner M.J. Mitochondria and FOXO3: breath or die. *Front Physiol*. 2013;4:147. doi 10.3389/fphys.2013.00147

Hussain S., Yadav S.S., Dwived.i P., Banerjee M., Usman K., Nath R., Khattri S. SNPs of *FOXO1* and their interactions contributes to the enhanced risk of diabetes among elderly individuals. *DNA Cell Biol*. 2022;41(4):381-389. doi 10.1089/dna.2021.1139

Hwang J.W., Rajendrasozhan S., Yao H., Chung S., Sundar I.K., Huyck H.L., Pryhuber G.S., Kinnula V.L., Rahman I. FOXO3 deficiency leads to increased susceptibility to cigarette smoke-induced inflammation, airspace enlargement, and chronic obstructive pulmonary disease. *J Immunol.* 2011;187(2): 987-998. doi 10.4049/jimmunol.1001861

Jiang H., Xu Y., Jiang Y., Li Y. FOXO3 activation prevents cellular senescence in emphysema induced by cigarette smoke. *COPD.* 2023; 20(1):80-91. doi 10.1080/15412555.2022.2164262

Klinpudtan N., Allsopp R.C., Kabayama M., Godai K., Gondo Y., Masui Y., Akagi Y. The association between longevity-associated FOXO3 allele and heart disease in septuagenarians and octogenarians: the SONIC study. *J Gerontol A Biol Sci Med Sci*. 2022;77(8): 1542-1548. doi 10.1093/gerona/glab204

Korytina G.F., Akhmadishina L.Z., Aznabaeva Y.G., Kochetova O.V., Zagidullin N.S., Kzhyshkowska J.G., Zagidullin S.Z., Viktorova T.V. Associations of the NRF2/KEAP1 pathway and antioxidant defense gene polymorphisms with chronic obstructive pulmonary disease. *Gene*. 2019;692:102-112. doi 10.1016/j.gene.2018.12.061

Korytina G.F., Akhmadishina L.Z., Kochetova O.V., Aznabaeva Y.G., Izmailova S.M., Zagidullin S.Z., Victorova T.V. Association of *CRP*, *CD14*, pro-inflammatory cytokines and their receptors (*TNFA*, *LTA*, *TNFRSF1A*, *TNFRSF1B*, *IL1B*, and *IL6*) genes with chronic obstructive pulmonary disease development. *Russ J Genet*. 2020;56,972-981. doi 10.1134/S1022795420080086 (in Russian)

Korytina G.F., Akhmadishina L.Z., Markelov V.A., Aznabaeva Y.G., Kochetova O.V., Nasibullin T.R., Larkina A.P., Khusnutdinova N.N., Zagidullin N.S., Victorova T.V. Role of PI3K/AKT/mTOR signaling pathway and sirtuin genes in chronic obstructive pulmonary disease development. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2023;27(5):512-521. doi 10.18699/VJGB-23-62

Li Y., Tian X., Luo J., Bao T., Wang S., Wu X. Molecular mechanisms of aging and anti-aging strategies. *Cell Commun Signal*. 2024;22(1): 285. doi 10.1186/s12964-024-01663-1

Liu J.Q., Zhang .L., Yao J., Yao S., Yuan T. AMPK alleviates endoplasmic reticulum stress by inducing the ER-chaperone ORP150 via FOXO1 to protect human bronchial cells from apoptosis. *Biochem Biophys Res Commun*. 2018;497(2):564-570. doi 10.1016/j.bbrc.2018.02.095

Luo X., Zeng W., Tang J., Liu W., Yang J., Chen H., Jiang L., Zhou X., Huang J., Zhang S., Du L., Shen X., Chi H., Wang H. Multi-modal transcriptomic analysis reveals metabolic dysregulation and immune responses in chronic obstructive pulmonary disease. *Sci Rep.* 2024; 14(1):22699. doi 10.1038/s41598-024-71773-w

Mahlooji M.A., Heshmati A., Kheiripour N., Ghasemi H., Asl S.S., Solgi G., Ranjbar A., Hosseini A. Evaluation of protective effects of curcumin and nanocurcumin on aluminium phosphide-induced subacute lung injury in rats: modulation of oxidative stress through SIRT1/FOXO3 signalling pathway. *Drug Res* (*Stuttg*). 2022;72(2): 100-108. doi 10.1055/a-1647-2418

Santana C.V.N., Magno L.A.V., Ramos A.V., Rios M.A., Sandrim V.C., De Marco L.A., de Miranda D.M., Romano-Silva M.A. Genetic variations in *AMPK*, *FOXO3A*, and *POMC* increase the risk of extreme obesity. *J Obes.* 2024;2024:3813621. doi 10.1155/2024/3813621

Soerensen M., Nygaard M., Dato S., Stevnsner T., Bohr V.A., Christensen K., Christiansen L. Association study of FOXO3A SNPs and aging phenotypes in Danish oldest-old individuals. *Aging Cell*. 2015;14(1):60-66. doi 10.1111/acel.12295

Ward L.D., Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res*. 2016;44(D1):D877-D881. doi 10.1093/nar/gkv1340

Xue T., Dong F., Gao J., Zhong X. Identification of related-genes of T cells in lung tissue of chronic obstructive pulmonary disease based on bioinformatics and experimental validation. *Sci Rep*. 2024;14(1): 12042. doi 10.1038/s41598-024-62758-w

Yao H., Chung S., Hwang J.W., Rajendrasozhan S., Sundar I.K., Dean D.A., McBurney M.W., Guarente L., Gu W., Rönty M., Kinnula V.L., Rahman I. SIRT1 protects against emphysema via FOXO3-mediated reduction of premature senescence in mice. *J Clin Invest.* 2012;122(6):2032-2045. doi 10.1172/JCI60132

Zang L., Chi J., Bi S., Tao Y., Wang R., Li L. SIRT3 improves alveolar epithelial cell damage caused by bronchopulmonary dysplasia through deacetylation of FOXO1. *Allergol Immunopathol* (*Madr*). 2023;51(2):191-204. doi 10.15586/aei.v51i2.710

Zhu M., Ye M., Wang J., Ye L., Jin M. Construction of potential miRNA-mRNA regulatory network in COPD plasma by bioinformatics analysis. *Int J Chron Obstruct Pulmon Dis.* 2020;15:2135-2145. doi 10.2147/COPD.S255262

Conflict of interest. The authors declare no conflict of interest.
Received February 9, 2025. Revised March 24, 2025. Accepted April 18, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):838-846

doi 10.18699/vjgb-25-91

Using polygenic scores to assess liability to antisocial behavior

A.V. Kazantseva [1], D.V. Yakovleva [1, 2], Yu.D. Davydova [1], E.K. Khusnutdinova [1]

[1] Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
[2] Ufa University of Science and Technology, Ufa, Russia
Kazantsa@mail.ru

Abstract. To date, several genome-wide association studies (GWAS) of antisocial behavior (ASB) have been conducted in Europeans, which promoted research aimed at evaluating liability to ASB-related phenotypes in independent samples. Such studies implemented a polygenic score (PGS) approach, which represents a composite score considering a number of "risky" alleles. Since no GWAS of ASB has been conducted in Russians, the present study aimed to perform a replication study of liability to severe criminal behavior (homicide) in individuals from Russia using PGS. Moreover, we sought to obtain the best model considering PGS and potential social factors as predictors. Genotyping of the "top" ten SNPs previously identified in GWAS meta-analysis of ASB (*CADM2*, *REV3L*, *FOXP1*, *FOXP2*, *BDNF*, *FURIN*, *XKR6*, *TMEM18*, *SORCS3*, and *ZIC4* genes) was conducted via real-time PCR in 227 homicide offenders and 254 healthy donors from the Volga-Ural region of Russia. Multiple regression models included "weighted" and "unweighted" PGS and potential social factors as predictors. The best regression model of liability to severe ASB was based on genetic effects of examined SNPs and social predictors, including traumatic brain injury, severe chronic disease, and tobacco smoking, which was more pronounced among subjects with a family history of mental illness ($p = 2 \times 10^{-13}$). PGS alone explained a small proportion of variance in liability to ASB (1.1–1.5 %), while the inclusion of social parameters increased variance explained (16.2–21.2 %). Revealed findings evidence a higher impact of social factors than a composite effect of selected "top" SNPs in predicting liability to ASB in the examined cohort. A higher probability of ASB was linked to comorbid substance abuse, traumatic brain injury, and family history of mental illness, which may also represent a result of a "risky" genetic profile.

Key words: aggression; homicide; G × E interaction; polygenic score; regression model; ROC-analysis; social factors

For citation: Kazantseva A.V., Yakovleva D.V., Davydova Yu.D., Khusnutdinova E.K. Using polygenic scores to assess liability to antisocial behavior. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):838-846. doi 10.18699/vjgb-25-91

Funding. The study was supported by the State Contract of the Ministry of Science and Higher Education of the Russian Federation (No. 1022040500074-9).

Использование полигенных показателей для оценки предрасположенности к манифестации антисоциального поведения

А.В. Казанцева [1], Д.В. Яковлева [1, 2], Ю.Д. Давыдова [1], Э.К. Хуснутдинова [1]

[1] Институт биохимии и генетики – обособленное структурное подразделение Уфимского федерального исследовательского центра Российской академии наук, Уфа, Россия
[2] Уфимский университет науки и технологий, Уфа, Россия
Kazantsa@mail.ru

Аннотация. Проведенные к настоящему времени полногеномные анализы ассоциаций (GWAS) антисоциального поведения (АП) в европейских популяциях стали предпосылкой для дальнейших исследований по оценке предрасположенности к развитию схожих фенотипов в независимых выборках. В таких работах используются полигенные показатели, которые представляют собой обобщенный балл, учитывающий число «рисковых» аллелей по каждому включенному генетическому локусу. Поскольку в РФ не было проведено GWAS АП, цель настоящего исследования – проведение репликативного исследования предрасположенности к манифестации крайних форм АП (убийств) в российской когорте с использованием подсчета полигенных показателей. Кроме того, задачей было также выявление наилучшей модели, основанной на включении полигенного показателя и социальных факторов в качестве предикторов. Генотипирование 10 «топовых» SNP, идентифицированных ранее в метаанализе GWAS АП (в генах *CADM2*, *REV3L*, *FOXP1*, *FOXP2*, *BDNF*, *FURIN*, *XKR6*, *TMEM18*, *SORCS3*, *ZIC4*), проведено с помощью ПЦР в реальном времени у лиц, совершивших убийства ($N = 227$), и в контрольной группе ($N = 254$) из Волго-Уральского региона РФ. Множественный регрессионный анализ основывался на включении «взвешенных» и «невзвешенных» полигенных показателей и потенциальных социальных факторов в качестве предикторов. Наилучшая регрессионная модель предрасположенно-

сти к манифестации крайних форм АП содержала данные генетического профиля по 10 локусам и социальным факторам (черепно-мозговая травма, тяжелые хронические заболевания в анамнезе, табакокурение) и была наиболее значимой для лиц с семейной отягощенностью психическими расстройствами ($p = 2 \times 10^{-13}$). Введение только полигенного показателя в модель объясняло небольшой процент вариации в предрасположенности к АП (1.1–1.5 %), тогда как добавление социальных предикторов увеличивало процент объясненной вариации (16.2–21.2 %). Полученные результаты указывают на большую значимость социальных факторов по сравнению с кумулятивным эффектом 10 локусов в предикции развития АП в исследуемой выборке. Повышенная вероятность его манифестации связана с наличием коморбидного аддиктивного поведения, черепно-мозговой травмы и семейной отягощенности психопатологиями, что может быть результатом наличия «рискового» генетического профиля.

Ключевые слова: агрессия; убийство; ген-средовые взаимодействия; полигенная оценка; регрессионная модель; ROC-анализ; социальные факторы

Introduction

Aggressive behavior (AB) and antisocial behavior (ASB) represent a destructive form of social interaction aimed at causing damage to another object and resulting in its frustration. From the evolutionary point of view, enhanced aggression was required for the survival of human groups (Baron, Richardson, 2004), thus promoting certain biological benefits. Although it is suggested that the aggression level in modern society is decreased compared with early humans, it still remains significant. To be more precise, the level of severe crimes, including homicides and intentional inflictions of severe harm, accounted for 117.3 and 567.1 thousand cases in 2022 in Russia (according to the data from the Ministry of Internal Affairs of the Russian Federation, http://www.crimestat.ru). According to the data from the World Health Organization (https://www.who.int/data), the homicide rate remains significant worldwide and was estimated at 5.8 cases per 100,000 of population in 2021 in the United States (in comparison, 6.7 cases in Russia; 0.5–4.5 cases in Europe; 5–100 cases in South American countries, and 5–20 cases in Africa).

In turn, during past years, several specifically cruel cases of murder, domestic violence, and antisocial behavior at schools have shocked Russia and the neighboring countries. However, it remains impossible to predict the occurrence of severe cruelty before the crimes have been conducted. In this regard, it seems important to determine significant factors underlying ASB, which can help to predict a higher probability of manifesting cruelty and antisocial behavior. It should be noted that ASB usually manifests in the form of certain psychiatric diseases, including oppositional defiant disorder, conduct disorder, and antisocial personality disorder (Pezzoli et al., 2025). Therefore, these phenotypes can share etiology and underlying factors.

According to previous research, the main factors predisposing to ASB or related phenotypes include biological, psychological, and environmental ones (Fritz et al., 2023). Examination of biological factors, which contribute 50 to 80 % of variance in aggression (Manchia, Fanos, 2017; Odintsova et al., 2023), is mainly focused on the analysis of genetic and epigenetic effects. Logically, genetic variants (SNPs) in the genes attributed to neurotransmitter release, reuptake, and binding (Davydova et al., 2020a; Antón-Galindo et al., 2023), oxytocin and arginine vasopressin signaling (Davydova et al., 2020b; Kazantseva et al., 2021), and others (Pezzoli et al., 2025) have been tested for their relation to individual variance in aggressive behavior. However, the results of multiple studies demonstrate inconsistent findings. Another methodological approach, i. e., genome-wide association studies (GWAS), enables to identify associated SNPs under a hypothesis-free paradigm. Although to date several GWASs of antisocial behavior have been carried out, these studies differ in the examined phenotypes frequently linked with ASB (combined phenotype of externalizing behavior (Karlsson Linnér et al., 2021), impulsivity (Deng et al., 2023), problems with self-regulation (Heilbronner et al., 2021), irritability (Mbatchou et al., 2021), risky behavior (Karlsson Linnér et al., 2019)) or age groups (children (Pappa et al., 2016), adults (Tielbeek et al., 2017)). Moreover, summarized findings from ~ 1.5 million subjects identified more than 500 SNPs related to liability to externalizing behavior, including antisocial behavior, attention-deficit/hyperactivity disorder (ADHD), and addiction in a European cohort (Karlsson Linnér et al., 2021).

One of the possible applications of GWAS findings is to use them for the calculation of polygenic scores (PGS) on the basis of effect estimates obtained for each SNP in the training sample. In turn, inclusion of PGS in mathematical models can gain prediction of enhanced risk of certain complex phenotypes. To date, several attempts seeking to replicate GWAS findings in an independent sample using PGS from ASB phenotype have been made (Karlsson Linnér et al., 2021; Li et al., 2023; Tesli et al., 2024; Acland et al., 2025), which succeeded in determining some proportion of variance in liability to conduct disorder, substance use disorders, smoking, ADHD, criminal behavior, depression, posttraumatic stress disorder, unemployment, and suicidal attempts. One of the possible limitations of using PGS for predicting ASB is the ethnic origin of the examined population, since differences in allele and genotype frequencies between ethnic groups can change SNPs' effect (Kazantseva et al., 2016). To date, no GWAS of liability to homicidal conduct has been carried out in subjects from Russia. Therefore, it is relevant to check if it is applicable to use the effect estimates obtained from combined ASB phenotype and different ethnic groups to predict the probability of conducting severe ASB in the Russian cohort.

Undoubtedly, specific environmental/social factors acting at various stages of ontogenesis affect genes' activity via epigenetic changes in regulation of genes responsible for manifesting aggression (Borinskaya et al., 2021). In this context, the analysis of potential social factors together with genetic effects (PGS) can help to increase the prognostic significance of the final model. In addition, it is established that ASB is highly accumulated in certain groups, including subjects with comorbid mental disorders (Ip et al., 2021; Wang et al., 2024; Pezzoli et al., 2025), family history of mental illness

(Han et al., 2024), addiction (Karlsson Linnér et al., 2021; Antón-Galindo et al., 2023), and unfavorable rearing conditions (Burt, 2022).

Considering the existing findings of ASB meta-analysis of European populations (Karlsson Linnér et al., 2021) and absent GWAS data for individuals from Russia, the present study aimed to evaluate the applicability of calculated polygenic scores based on existing GWAS data to predict severe ASB (homicide) in the Russian cohort. Moreover, to enhance the prognostic ability of regression models, we sought to obtain the best model with the optimal sensitivity and specificity, which assumes PGS and potential social factors as predictors.

Materials and methods

The study sample comprised 227 criminal offenders who conducted homicide and were directed to a forensic examination of present mental disorders in the Republican Clinical Psychiatric Hospital (Ufa, Russia). Only individuals without mental illness who were proven to be sane by the Court were included in the study. The examined sample consisted mainly of men (93 %) with a mean age of 41.5 ± 14.5 years. Ethnic content of the sample was the following: 48 % Russians, 34.8 % Tatars, and 17.2 % Bashkirs. The data on the social/clinical background of enrolled subjects were obtained via a survey and included the information on present and past tobacco smoking, alcohol/opiate abuse, family history of mental illness or criminal behavior, suicidal attempts, level of education, maltreatment in childhood, severe chronic disease in anamnesis, and type of ASB (proactive or reactive aggression).

The control group was selected on the basis of correspondence to the group of criminal offenders by age, ethnicity, and gender. In total, we examined DNA samples obtained from 254 individuals who reported no family history of mental illness and were non-registered in the psychiatric database of the Republic of Bashkortostan. The study was approved by the local bioethical committee at the Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences (Ufa, Russia) (protocol code 15, date of approval, October 12, 2017) in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

SNP selection for PGS calculation from GWAS meta-analysis of ASB (Karlsson Linnér et al., 2021) was based on the following criteria: the lowest level of significance ($p < 10^{-18}$); selection of a single SNP from a set of proxy SNPs; minor allele frequency (MAF) above 0.05 in Europeans (based on 1000 Genomes); and known regulatory effect of the SNP based on the RDB (Regulome Database, https://regulomedb.org/regulome-search) and CADD (Combined Annotation Dependent Depletion, https://cadd.gs.washington.edu) databases. The final list of selected SNPs included *CADM2* rs993137, *REV3L* rs458806, *FOXP1* rs11720703, *FOXP2* rs1476535, *BDNF* rs6265, *FURIN* rs4702, *XKR6* rs4240671, *TMEM18* rs6711254, *SORCS3* rs11596214, and *ZIC4* rs2279829, which were used for PGS calculation, and is reported in Table 1. Genotyping of previously extracted DNA of the control group and criminal offenders was carried out using real-time PCR with KASP chemistry (LGC Genomics, UK).

All examined SNPs corresponded to the Hardy–Weinberg equilibrium ($p > 0.05$). At the second stage, we calculated PGS based on effect estimates obtained from R. Karlsson Linnér et al. (2021). Namely, PGS for each individual from our sample was calculated on the basis of inclusion of 1) SNPs under $p < 0.1$ ("weighted" effect), 2) all SNPs (using "weighted" effect), 3) all SNPs (using "unweighted" effect). Calculation of "weighted" and "unweighted" PGS was previously explained in detail (Kazantseva et al., 2023a). Briefly, individual PGS was calculated as the weighted/unweighted sum of the number of effect alleles at a certain SNP multiplied by the effect estimate (PLINK v.1.09).

Subsequently, a series of multiple logistic regressions was performed to obtain models that can predict liability to ASB in the total groups of homicide offenders, as well as in subgroups of subjects with proactive forms of aggression, comorbid substance use, or known family history of mental illness or criminal behavior. Initially, only PGS as a predictor was included, which was followed by a backward selection procedure to obtain a list of statistically significant social parameters to be

Table 1. Examined top SNPs linked to antisocial behavior: data from previous ASB GWAS and the VUR cohort

Gene	SNP	EA/OA	EE^{K}	EAF^{K}	EAF^{ASB}	EAF^{CT}	p^{K}	p^{VUR}
CADM2	rs993137	C/T	0.020	0.383	0.251	0.302	4.61×10^{-53}	**0.081**
REV3L	rs458806	C/T	0.016	0.178	0.313	0.252	1.30×10^{-29}	**0.043**
FOXP1	rs11720703	T/C	0.013	0.471	0.391	0.382	2.87×10^{-27}	0.795
FOXP2	rs1476535	T/C	0.013	0.451	0.459	0.445	3.41×10^{-26}	0.688
BDNF	rs6265	C/T	0.015	0.814	0.841	0.852	1.78×10^{-24}	0.647
FURIN	rs4702	G/A	0.012	0.442	0.405	0.461	1.08×10^{-23}	**0.075**
XKR6	rs4240671	G/A	0.012	0.509	0.563	0.543	4.80×10^{-23}	0.528
TMEM18	rs6711254	A/G	0.015	0.173	0.190	0.161	1.89×10^{-22}	0.257
SORCS3	rs11596214	G/A	0.011	0.606	0.593	0.533	6.25×10^{-21}	**0.065**
ZIC4	rs2279829	C/T	0.013	0.788	0.801	0.771	2.88×10^{-18}	0.247

Note. EA/OA – effect allele/other allele; EE – effect estimate; EAF – effect allele frequency; ASB – criminal offenders from VUR; CT – control group from VUR; VUR – Volga-Ural sample; K – data from R. Karlsson Linnér et al. (2021); *p* – *p*-values. *P*-values obtained in the present sample at a trend level ($p < 0.1$) are marked in bold.

included as predictors together with PGS (R v.4.4.2). To select the best predicting model, we have compared data on the lowest *p*-value, the highest proportion of variance (Nagelkerke pseudo-R^2) explaining liability to ASB, and the highest area under the ROC curve (AUC) for each model.

Results

At the initial stage of the study, we examined the presence of significant differences between the criminal offenders and the control group in the examined social factors (Table 2). We have observed the differences in the proportion of individuals characterized by severe somatic diseases and traumatic brain injuries in anamnesis ($p = 1.2 \times 10^{-12}$), depending on education level ($p = 4.5 \times 10^{-16}$) and present smoking ($p = 4.0 \times 10^{-7}$) between the groups.

For the genetic part of the present study we selected the "top" ten SNPs ($p < 2.9 \times 10^{-18}$) identified in the previous meta-analysis GWAS of ASB (Karlsson Linnér et al., 2021). Effect estimates for alleles used for calculation of "weighted PGS" as well as effect allele frequencies in the VUR sample are given in Table 1. In addition, we have tested for statistically significant differences in allele frequencies of examined SNPs between criminal offenders and the control group in the examined cohort from the VUR, which enabled us to confirm a coincidence of four SNPs, although at a trend level ($p < 0.1$): *CADM2* rs993137, *REV3L* rs458806, *FURIN* rs4702, and *SORCS3* rs11596214.

Primary logistic regression models that included PGS (based on four SNPs) revealed a small proportion of variance in liability to antisocial behavior in the total group ($r^2 = 0.9$ %, $p = 0.014$), among subjects with a proactive form of aggression ($r^2 = 0.9$ %, $p = 0.017$), with comorbid substance abuse ($r^2 = 0.9$ %, $p = 0.027$), and with a family history of mental illness ($r^2 = 1.5$ %, $p = 0.014$) (Table 3, Fig. 1). At the initial stage of regression analysis, we have included all social factors, including sex and ethnicity as covariates, together with PGS.

As expected, inclusion of potential social parameters as predictors enabled an increase in the statistical significance of the models, which resulted in 14.5 % (ASB), 15.8 % (proactive ASB), 21.0 % (ASB with comorbid addiction), and 21.2 % of variance (ASB with family history of mental illness) being explained. It should be mentioned that valuable social factors comprised of traumatic brain injury (TBI) or

Table 2. Characteristics of the examined groups of criminal offenders and healthy donors and analysis of differences in social factors between the groups

Social factor	Group	ASB (*N* = 227)		Control group (*N* = 254)		β	*p*-value
		N	%	*N*	%		
Sex	Men	211	92.9	235	92.5	–0.11	0.73
	Women	16	7.1	19	7.5		
Ethnicity	Russians	109	48	120	47.2	–0.003	0.99
	Tatars	79	34.8	92	36.2	–0.06	0.83
	Bashkirs	39	17.2	42	16.6	0.03	0.91
Aggression type	Proactive	211	92.9	–	–	–	–
	Reactive	16	7.1	–	–		
Family history of mental illness	Yes	88	38.8	–	–	–	–
	No	139	61.2	254	100		
TBI/disease	Yes	136	59.9	62	24.4	1.53	**1.2×10^{-12}**
	No	91	40.1	192	75.6		
Education level	High	92	40.5	213	83.9	–2.0	**4.5×10^{-16}**
	Low	135	59.5	41	16.1		
Present smoking	Yes	160	70.4	112	44.1	1.1	**4.0×10^{-7}**
	No	67	29.6	142	55.9		
Past smoking	Yes	179	78.9	202	79.5	–0.05	0.84
	No	48	21.1	52	20.5		
Alcohol/opiate abuse	Yes	150	66.1	–	–	–	–
	No	77	33.9	254	100		
Maltreatment in childhood	Yes	31	13.7	24	9.4	0.36	0.73
	No	196	86.3	230	90.6		
Suicidal attempts	Yes	45	19.8	–	–	–	–
	No	182	80.2	254	100		

Note. Statistically significant differences between the groups based on *p*-value < 0.05 are shown in bold. TBI/disease – traumatic brain injury or severe chronic disease in anamnesis. Dashes indicate non-applicable data.

Table 3. Regression models of liability to ASB based on a polygenic score and social factors as predictors

Model	Predictor	ASB (*N* = 227)			Proactive ASB (*N* = 211)			Addictive ASB (*N* = 150)			FamPsy ASB (*N* = 88)		
		β	SE	*p*-value	β	SE	*p*-value	β	SE	*p*-value	β	SE	*p*-value
1	PGS_0.1	105.0	42.8	0.014	106.6	44.8	0.017	106.1	48.1	0.027	141.4	57.9	0.014
	Model *p*-value	0.014			0.017			0.027			0.014		
	Adjusted r^2	0.009			0.009			0.009			0.015		
	AUC	0.569			0.569			0.571			0.589		
2	PGS_0.1	84.8	55.4	0.126	86.6	57.4	0.131	98.2	63.5	0.122	148.7	75.2	0.048
	TBI/disease	1.4	0.2	9.8×10^{-9}	1.3	0.2	1.1×10^{-7}	1.3	0.3	5.7×10^{-7}	1.6	0.3	1.9×10^{-7}
	Past smoking	–1.3	0.3	8.8×10^{-4}	–1.4	0.4	4×10^{-4}	–1.1	0.5	0.031	–1.2	0.6	0.048
	Present smoking	1.7	0.3	1.4×10^{-7}	1.7	0.3	4.1×10^{-7}	2.2	0.4	4.1×10^{-7}	1.8	0.5	4.4×10^{-4}
	Model *p*-value	4.24×10^{-14}			7.7×10^{-16}			$< 10^{-16}$			1.3×10^{-13}		
	Adjusted r^2	0.145			0.158			0.210			0.212		
	AUC	0.752			0.744			0.788			0.780		
3	PGS_W	220.4	69.4	1.4×10^{-3}	223.1	72.3	0.002	195.8	77.7	0.011	206.6	94.0	0.028
	Model *p*-value	1.3×10^{-3}			0.002			0.011			0.028		
	Adjusted r^2	0.015			0.015			0.011			0.012		
	AUC	0.579			0.580			0.577			0.576		
4	PGS_W	193.3	92.4	0.036	191.9	95.3	0.044	178.8	104.4	0.086	245.2	125.8	0.051
	TBI/disease	1.4	0.2	1.3×10^{-8}	1.3	0.2	1.3×10^{-7}	1.3	0.3	5.7×10^{-7}	1.6	0.3	2.1×10^{-7}
	Past smoking	–1.2	0.3	1.4×10^{-3}	–1.3	0.4	7.1×10^{-4}	–1.1	0.5	0.034	–1.2	0.6	0.045
	Present smoking	1.7	0.3	2.8×10^{-7}	1.6	0.3	9×10^{-7}	2.2	0.4	6.9×10^{-7}	1.8	0.5	5.3×10^{-4}
	Model *p*-value	$< 10^{-16}$			3.3×10^{-16}			$< 10^{-16}$			1.4×10^{-13}		
	Adjusted r^2	0.170			0.162			0.211			0.212		
	AUC	0.758			0.751			0.792			0.800		
5	PGS_UW	2.9	0.9	0.002	2.9	1.0	0.003	2.5	1.1	0.019	2.6	1.3	0.045
	Model *p*-value	0.002			0.003			0.019			0.045		
	Adjusted r^2	0.013			0.013			0.010			0.010		
	AUC	0.588			0.589			0.584			0.584		
6	PGS_UW	2.5	1.3	0.047	2.5	1.3	0.053	2.2	1.4	0.131	3.1		0.078
	TBI/disease	1.4	0.2	1.2×10^{-8}	1.3	0.2	1.2×10^{-7}	1.4	0.3	5.6×10^{-7}	1.6	0.3	2.1×10^{-7}
	Past smoking	–1.2	0.3	0.0013	–1.3	0.4	6.5×10^{-4}	–1.1	0.5	0.032	–1.2	0.6	0.044
	Present smoking	1.6	0.3	3.1×10^{-7}	1.6	0.3	9.6×10^{-7}	2.2	0.4	6.9×10^{-7}	1.8	0.5	5.7×10^{-4}
	Model *p*-value	$< 10^{-16}$			3.3×10^{-16}			$< 10^{-16}$			2.0×10^{-13}		
	Adjusted r^2	0.169			0.161			0.210			0.210		
	AUC	0.760			0.753			0.794			0.802		

Note. ASB – antisocial behavior; FamPsy ASB – ASB in individuals with a family history of mental illness; β – regression coefficient for each predictor in the model; SE – standard error of β; AUC – area under curve; TBI/disease – traumatic brain injury or severe chronic disease in anamnesis. PGS_0.1 was based on effect estimates for *REV3L* rs458806, *FOXP1* rs11720703, *XKR6* rs4240671, and *SORCS3* rs11596214; PGS_W and PGS_UW were PGS based on "weighted" and "unweighted" effect estimates for ten SNPs, correspondingly.

severe chronic disease in anamnesis (β = 1.4, $p = 9.8\times10^{-9}$) and present smoking (β = 1.7, $p = 1.4\times10^{-7}$) were associated with enhanced liability to aggression, while past smoking demonstrated a positive effect on ASB decrease (β = –1.3, $p = 8.8\times10^{-4}$). The impact of other social factors together with sex and ethnicity remained insignificant after the backward selection procedure. Therefore, inclusion of the mentioned social parameters allowed us to explain up to 16.1 % of variance in developing ASB. According to determined models, we can conclude that they possess the highest prediction ability for developing ASB in individuals who have relatives with mental disorders or criminal behavior (AUC = 0.780) or have alcohol/opiate addiction (AUC = 0.788) (Table 3).

At the second stage of our analysis, we calculated PGS based on effect estimates for all examined SNPs, even if they were non-significant in the VUR sample (Table 1). Therefore,

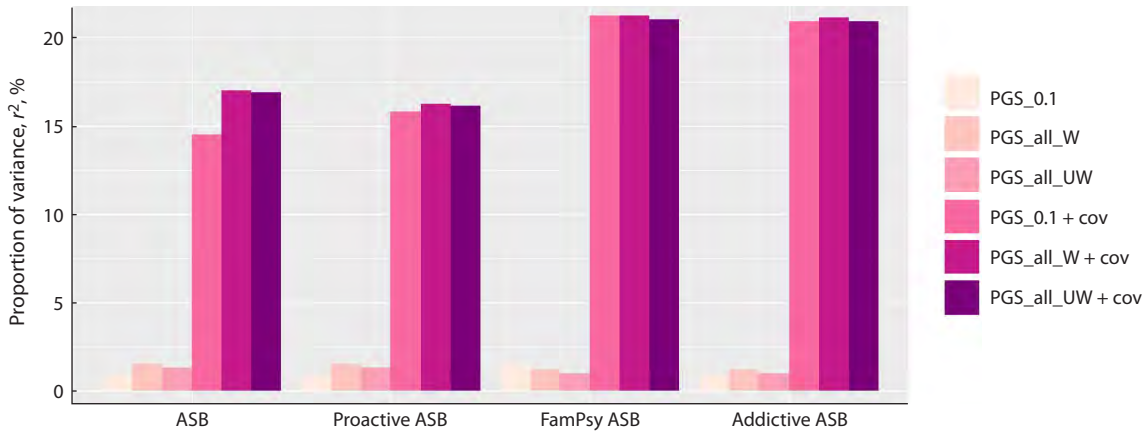


Fig. 1. Proportion of variance (adjusted r^2) in liability to antisocial behavior in the examined cohort explained by predictors included in multiple regression models based on PGS calculation of SNPs with $p < 0.1$ (PGS_0.1), "weighted effects" of all SNPs (PGS_all_W), "unweighted effects" of all SNPs (PGS_all_UW) with inclusion of social predictors (PGS_0.1 + cov, PGS_all_W + cov, PGS_all_UW + cov).

Examined groups of ASB: total group of homicide offenders (ASB); homicide offenders with a proactive type of ASB (Proactive ASB), family history of mental illness or criminal behavior (FamPsy ASB), or substance abuse (Addictive ASB).

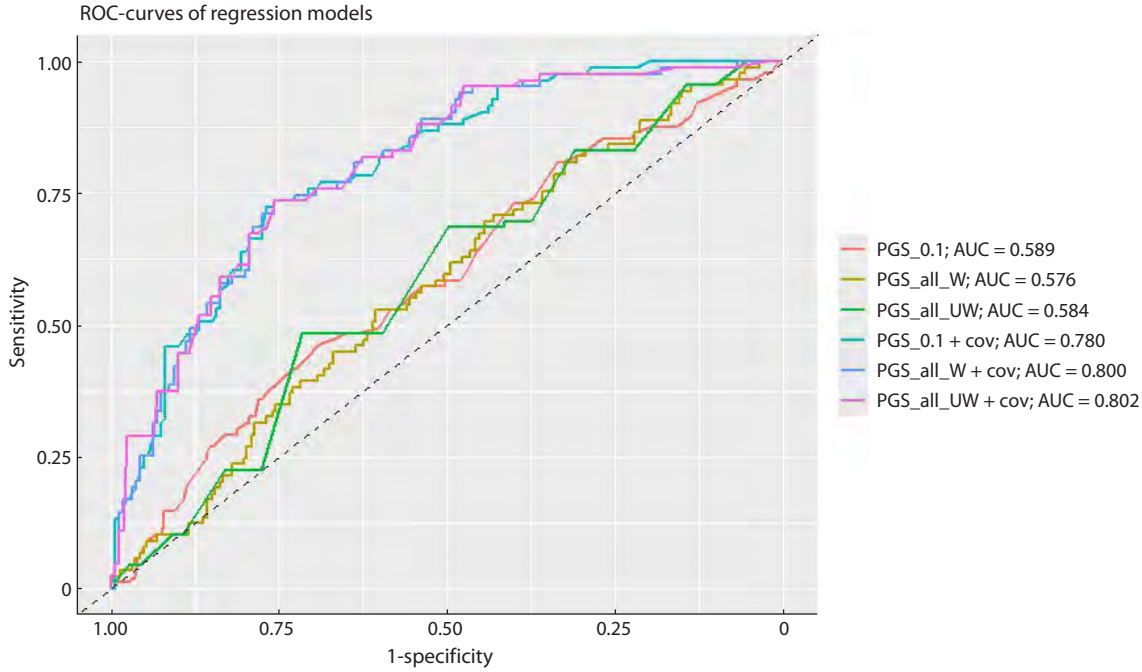


Fig. 2. ROC curves and areas under the curves (AUCs) for various regression models predicting liability to manifest ASB in individuals with a family history of mental illness or criminal behavior (abbreviations are given as in Fig. 1).

regression models, which implemented "weighted" (PGS_W) and "unweighted" (PGS_UW) PGS, slightly enhanced the proportion of variance in liability to ASB compared to previous models 1 and 2 (Table 2). Namely, a combined effect of ten genetic variants explained 1.1–1.5 % ("weighted effect") and 1.0–1.3 % ("unweighted effect") of variance in predisposition to homicide violence. Previously mentioned social predictors remained significant and, together with PGS, enabled to enhance the proportion of variance explained (16.2–21.2 % in "weighted" PGS, 16.1–21.0 % in "unweighted" PGS).

However, it seems that the inclusion of a larger number of non-significant SNPs had a very small effect on improving the predicting abilities of the models. Nevertheless, models with ten vs. four SNPs in PGS demonstrated slightly higher prognostic ability for ASB in the total sample and in individuals with a proactive form of aggressive behavior or comorbid substance abuse (Table 2, Models 4, 6). We have also constructed ROC curves and calculated comparative areas under the curves (AUCs) for all analyzed models (Fig. 2). Finally, our findings indicate that the best regression model has higher prognostic ability (r^2 = 21 %) and a moderate measure of classifier performance (AUC = 0.802) to designate subjects at high risk for developing ASB if they have family history of mental disorders.

Discussion

In the present study, we tested different logistic regression models, which were based on calculated polygenic scores, aimed at predicting liability to homicide in individuals from the Volga-Ural region of Russia. Based on our findings, the highest prediction ability for developing ASB was observed for individuals with a family history of mental disorders and those with substance abuse comorbidity. The data revealed are not surprising, since externalizing pathology is frequently accumulated in families (Acheson et al., 2018; Han et al., 2024) due to shared genetic profiles between biological parents and offspring. On the other hand, it was reported that the same genes/genetic variants were linked to different psychiatric conditions, addiction, and antisocial behavior (Ip et al., 2021; Antón-Galindo et al., 2023; Li et al., 2023; Wang et al., 2024; Pezzoli et al., 2025), which can be explained by genes' pleiotropy in various complex traits (Watanabe et al., 2019).

Since no significant difference in predicting ASB risk in the VUR sample was observed among models based on "weighted" and "unweighted" effects of SNPs, it can be concluded that effect estimates from GWAS of Europeans seem to be inappropriate for individuals from Russia. Therefore, future research should be focused on conducting GWAS of ASB in a Russian cohort followed by verification in the same-ethnicity independent sample. Published studies, which sought to replicate findings obtained for different populations, succeeded in using PGS from ASB to predict liability to externalizing behavior in both Europeans and African Americans (Brislin et al., 2024), although representing a small cumulative effect of genetic variants.

Our findings indicate a very small impact of selected SNPs on predicting ASB, which was based on the effect estimates from the study of summarized phenotype of externalizing pathology. The data obtained support previous findings on the small effect (0.1–4.0 %) of analyzed genetic variants (even at a genome-wide level) as polygenic scores on predicting ASB (Tielbeek et al., 2017, 2022; Tesli et al., 2024). Our previous research also revealed a small proportion of variance explained in aggression level in a general population of Russia, which was attributed to the combined effect of 30 genetic variants (Kazantseva et al., 2023b).

It is known that environmental factors play a modulating role in establishing specific patterns of behavior (Kazantseva et al., 2014), including ASB-related ones. In particular, harsh parenting (Burt, 2022), school violence (Acland et al., 2025), and affiliation with delinquent peers (Schwartz et al., 2019) were assumed to increase a risk for manifesting ASB. Regression models designed in the present study also point to a more pronounced effect of environmental factors in establishing ASB than that of the genetic component. These findings are at odds with existing studies, which also depicted valuable impact of such social factors as community violence (Musci et al., 2019), harsh parenting (Acland et al., 2025), and low parental education level (Barnes et al., 2019) under gene-by-environment interactions.

In the present study, we have observed a significant effect of present smoking and history of traumatic brain injury/severe chronic disorders on manifesting criminal behavior. One of the probable links between smoking and ASB is attributed to the influence of nicotine on the CNS via exaggerated stress sensitivity (Weltens et al., 2021) and changed in epigenetic regulation (Gould et al., 2023). It should be noted that the usually accepted environmental effects can also be due to the impact of certain genetic and epigenetic profiles, which are inherited (McAdams et al., 2013). In this regard, present smoking may represent the result of activity of the genes responsible for developing addiction and externalizing behavior. Moreover, the negative effect of smoking promoting the development of ASB later in life was only evident for individuals with predisposing genetic patterns. Namely, individuals who were subjected to prenatal smoking exposure (their mothers smoked during pregnancy) demonstrated an enhanced risk of ASB only if they were genetically related to their mothers. At the same time, no link between maternal smoking and offspring' ASB was observed if children were developed from a donated egg (van Goozen et al., 2022).

Another significant factor affecting liability to ASB in our cohort is traumatic brain injury in anamnesis, which confirms previous data (Ryan et al., 2021; Theadom et al., 2024). It was suggested that TBI can cause abnormal morphometry of the central executive network in the brain, which can result in worsening of executive functions (Ryan et al., 2021) or exacerbate other valuable triggers, including social deprivation (Guskiewicz et al., 2003), thus promoting ASB. In summary, reported findings on the higher effect of social factors on developing ASB in the Russian cohort can probably capture the effect of other genes on the occurrence of such an "environment".

Future research should integrate various methodological approaches, including those measuring brain activity and connectivity underlying specificity of individual behavioral responses, and consider the impact of genetic and environmental factors. For instance, there is some evidence of a link between amygdala hyper-reactivity and increased impulsivity and reduced self-regulation as a response to threatening stimuli (Dotterer et al., 2017). Another study identified a link between diminished P3(P300) amplitude of electrical potential, which was obtained as a response to a visual oddball task, and manifestation of externalizing phenotypes (Iacono, 2018; Brislin et al., 2024).

The present study has several limitations. First, the set of SNPs used for PGS calculation is rather small, which can mirror the low proportion of explained variance in liability to ASB attributed to genetic impact. To be more precise, calculated PGS in the previous meta-analysis (Karlsson Linnér et al., 2021) enabled to explain 3–4 % of variance in manifesting a combined phenotype of antisocial behavior when PGS was estimated on genetic data from 579 SNPs at the genome-wide significance level.

In turn, the present study has been focused on biallelic polymorphisms only, while other structural variations in the genome, such as tandem repeats and microdeletions/duplications, which can also contribute to genetically caused manifestations of aggression, remained unstudied within the present research. Although the examined sample represents a specific cohort of individuals with a severe form of aggressive behavior (homicide), the sample size is small, which can result in type I and II errors and requires future enlargement of the examined sample. Moreover, the obtained PGS models are limited to a number of analyzed social factors,

while other probably relevant factors including child-parent relationship, belonging to a criminal organization, physical or sexual violence, social isolation, personality type, etc. were non-examined. Finally, since the majority of enrolled offenders were characterized by excessive alcohol/opiate use, we cannot rule out whether the reported findings are attributed to present heavy alcohol drinking.

Conclusion

In summary, the present study represents an attempt to create a prognostic model for developing antisocial behavior in a Russian cohort based on genetic data reported for European populations. Revealed findings present evidence for a higher impact of social factors rather than a composite effect of the selected "top" SNPs in predicting liability to ASB. Nevertheless, the best model was able to explain up to 21.2 % of variance in liability to ASB, especially in subjects with a family history of mental illness or criminal behavior, which was based on the genetic profile of ten SNPs and such social parameters as traumatic brain injury, severe chronic disease in anamnesis, and tobacco smoking. Future research in this field has to be focused on performing GWAS in a Russian cohort of criminal offenders (or persons with other types of antisocial behavior) to identify genetic loci and their effect estimates specific to the main ethnic groups from Russia. Obviously, such analyses will enable the design of models of liability to manifest ASB with higher prediction probabilities.

References

Acheson A., Vincent A.S., Cohoon A.J., Lovallo W.R. Defining the phenotype of young adults with family histories of alcohol and other substance use disorders: studies from the family health patterns project. *Addict Behav*. 2018;77:247-254. doi 10.1016/j.addbeh.2017.10.014

Acland E.L., Pocuca N., Paquin S., Boivin M., Ouellet-Morin I., Andlauer T.F.M., Gouin J.P., Côté S.M., Tremblay R.E., Geoffroy M., Castellanos-Ryan N. Polygenic risk and hostile environments: links to stable and dynamic antisocial behaviors across adolescence. *Dev Psychopathol*. 2025;37(1):464-476. doi 10.1017/S095457942400004X

Antón-Galindo E., Cabana-Domínguez J., Torrico B., Corominas R., Cormand B., Fernàndez-Castillo N. The pleiotropic contribution of genes in dopaminergic and serotonergic pathways to addiction and related behavioral traits. *Front Psychiatry*. 2023;14:1293663. doi 10.3389/fpsyt.2023.1293663

Barnes J.C., Liu H., Motz R.T., Tanksley P.T., Kail R., Beckley A.L., Belsky D.W., Domingue B.W., Moffitt T.E., Pratt T.C., Wertz J. The propensity for aggressive behavior and lifetime incarceration risk: a test for gene-environment interaction (G × E) using whole-genome data. *Aggres Violent Behav*. 2019;49:101307. doi 10.1016/j.avb.2019.07.002

Baron R.A., Richardson D.R. Human Aggression. New York: Plenum Press, 2004

Borinskaya S.A., Rubanovich A.V., Larin A.K., Kazantseva A.V., Davydova Y.D., Genetozov E.V., Khusnitdinova E.K., Yankovsky N.K. Epigenome-wide association study of CpG methylation in aggressive behavior. *Russ J Genet*. 2021;57(12):1454-1460. doi 10.1134/S1022795421120048

Brislin S.J., Salvatore J.E., Meyers J.M., Kamarajan C., Plawecki M.H., Edenberg H.J., Kuperman S., … Kramer J.R., Chan G., Porjesz B.; COGA Collaborators; Dick D.M. Examining associations between genetic and neural risk for externalizing behaviors in adolescence and early adulthood. *Psychol Med*. 2024;54(2):267-277. doi 10.1017/S0033291723001174

Burt S.A. The genetic, environmental, and cultural forces influencing youth antisocial behavior are tightly intertwined. *Annu Rev Clin Psychol*. 2022;18:155-178. doi 10.1146/annurev-clinpsy-072220-015507

Davydova Y., Kazantseva A., Enikeeva R., Mustafin R., Malykh S., Lobaskova M., Valinurov R., Akhmerova I., Khusnutdinova E. The involvement of hypothalamic-pituitary-adrenal and monoaminergic systems genes in developing aggressive behavior. *Eur Neuropsychopharmacol*. 2020a;31(S1):S16-S17. doi 10.1016/j.euroneuro.2019.12.023

Davydova Y.D., Kazantseva A.V., Enikeeva R.F., Mustafin R.N., Lobaskova M.M., Malykh S.B., Gilyazova I.R., Khusnutdinova E.K. The role of oxytocin receptor (*OXTR*) gene polymorphisms in the development of aggressive behavior in healthy individuals. *Russ J Genet*. 2020b;56(9):1129-1138. doi 10.1134/S1022795420090057

Deng W.Q., Belisario K., Gray J.C., Levitt E.E., Mohammadi-Shemirani P., Singh D., Pare G., MacKillop J. Leveraging related health phenotypes for polygenic prediction of impulsive choice, impulsive action, and impulsive personality traits in 1534 European ancestry community adults. *Genes Brain Behav*. 2023;22(3):e12848. doi 10.1111/gbb.12848

Dotterer H.L., Hyde L.W., Swartz J.R., Hariri A.R., Williamson D.E. Amygdala reactivity predicts adolescent antisocial behavior but not callous-unemotional traits. *Dev Cogn Neurosci*. 2017;24:84-92. doi 10.1016/j.dcn.2017.02.008

Fritz M., Soravia S.M., Dudeck M., Malli L., Fakhoury M. Neurobiology of aggression-review of recent findings and relationship with alcohol and trauma. *Biology* (*Basel*). 2023;12(3):469. doi 10.3390/biology12030469

Gould T.J. Epigenetic and long-term effects of nicotine on biology, behavior, and health. *Pharmacol Res*. 2023;192:106741. doi 10.1016/j.phrs.2023.106741

Guskiewicz K.M., McCrea M., Marshall S.W., Cantu R.C., Randolph C., Barr W., Onate J.A., Kelly J.P. Cumulative effects associated with recurrent concussion in collegiate football players: the NCAA Concussion Study. *JAMA*. 2003;290(19):2549-2555. doi 10.1001/jama.290.19.2549

Han M.J., Kim S.T., Park C.I., Hwang S.S., Kim H.W., Kang J.I., Kim S.J. Serial mediating effects of childhood trauma and conduct behaviors on the impact of family history among patients with alcohol use disorder. *Sci Rep*. 2024;14(1):7196. doi 10.1038/s41598-024-57861-x

Heilbronner U., Papiol S., Budde M., Andlauer T.F.M., Strohmaier J., Streit F., Frank J., … Stürmer T., Müller-Myhsok B., Nöthen M.M., Rietschel M., Schulze T.G. "The Heidelberg Five" personality dimensions: genome-wide associations, polygenic risk for neuroticism, and psychopathology 20 years after assessment. *Am J Med Genet B Neuropsychiatr Genet*. 2021;186(2):77-89. doi 10.1002/ajmg.b.32837

Iacono W.G. Endophenotypes in psychiatric disease: prospects and challenges. *Genome Med*. 2018;10(1):11. doi 10.1186/s13073-018-0526-5

Ip H.F., Van der Laan C.M., Krapohl E.M.L., Brikell I., Sánchez-Mora C., Nolte I.M., St Pourcain B., … Lundström S., Plomin R., Bartels M., Nivard M.G., Boomsma D.I. Genetic association study of childhood aggression across raters, instruments, and age. *Transl Psychiatry*. 2021;11(1):413. doi 10.1038/s41398-021-01480-x

Karlsson Linnér R., Biroli P., Kong E., Meddens S.F.W., Wedow R., Fontana M.A., Lebreton M., … Lee J.J., Cesarini D., Benjamin D.J., Koellinger P.D., Beauchamp J.P. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet*. 2019;51(2):245-257. doi 10.1038/s41588-018-0309-3

Karlsson Linnér R., Mallard T.T., Barr P.B., Sanchez-Roige S., Madole J.W., Driver M.N., Poore H.E., … Waldman I.D., Palmer A.A., Harden K.P., Koellinger P.D., Dick D.M. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nat Neurosci*. 2021;24(10):1367-1376. doi 10.1038/s41593-021-00908-3

Kazantseva A.V., Kutlumbetova Yu.Yu., Malykh S.B., Lobaskova M.M., Khusnutdinova E.K. Arginine-vasopressin receptor gene (*AVPR1A*, *AVPR1B*) polymorphisms and their relation to personality traits. *Russ J Genet*. 2014;50(3):298-307. doi 10.1134/S1022795414030041

Kazantseva A.V., Malykh S.B., Khusnutdinova E.K. Molecular-genetic studies of personality: from gene candidate analyses to genome-wide association studies. In: Malykh S.B., Kovas Y.V., Gaysina D.A. (Eds) Genomics of Behavior: Children Development and Education. Tomsk: Tomsk State University, 2016;178-209 (in Russian)

Kazantseva A.V., Davydova Y.D., Enikeeva R.F., Valinurov R.G., Gareeva A.E., Khusnutdinova N.N., Khusnutdinova E.K. The association study of polymorphic variants of hypothalamic-pituitary-adrenal system genes (*AVPR1B*, *OXTR*) and aggressive behavior manifestation: a focus on social environment. *Res Results Biomed*. 2021;7(3):232-244. doi 10.18413/2658-6533-2021-7-3-0-3

Kazantseva A., Davydova Y., Enikeeva R., Mustafin R.N., Malykh S., Lobaskova M., Kanapin A., Prokopenko I., Khusnutdinova E. A combined effect of polygenic scores and environmental factors on individual differences in depression level. *Genes*. 2023a;14(7):1355. doi 10.3390/genes14071355

Kazantseva A.V., Davydova Y.D., Enikeeva R.F., Yakovleva D.V., Mustafin R.N., Lobaskova M.M., Malykh S.B., Khusnutdinova E.K. Individual variance in human aggression: a combined effect of polygenic score and social/lifestyle factors. *Russ J Genet*. 2023b;59(S2):S227-S236. doi 10.1134/s1022795423140065

Li W., Zhou H., Thygesen J.H., Heydtmann M., Smith I., Degenhardt F., Nöthen M., Morgan M.Y., Kranzler H.R., Gelernter J., Bass N., McQuillin A. Genome-wide association study of antisocial personality disorder diagnostic criteria provides evidence for shared risk factors across disorders. *Psychiatr Genet*. 2023;33(6):233-242. doi 10.1097/YPG.0000000000000352

Manchia M., Fanos V. Targeting aggression in severe mental illness: the predictive role of genetic, epigenetic, and metabolomic markers. *Prog Neuropsychopharmacol Biol Psychiatry*. 2017;77:32-41. doi 10.1016/j.pnpbp.2017.03.024

Mbatchou J., Barnard L., Backman J., Marcketta A., Kosmicki J.A., Ziyatdinov A., Benner C., … Baras A., Reid J., Abecasis G., Maxwell E., Marchini J. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*. 2021;53(7):1097-1103. doi 10.1038/s41588-021-00870-7

McAdams T.A., Gregory A.M., Eley T.C. Genes of experience: explaining the heritability of putative environmental variables through their association with behavioural and emotional traits. *Behav Genet*. 2013;43(4):314-328. doi 10.1007/s10519-013-9591-0

Musci R.J., Bettencourt A.F., Sisto D., Maher B., Masyn K., Ialongo N.S. Violence exposure in an urban city: a G×E interaction with aggressive and impulsive behaviors. *J Child Psychol Psychiatry*. 2019;60(1):72-81. doi 10.1111/jcpp.12966

Odintsova V.V., Hagenbeek F.A., van der Laan C.M., van de Weijer S., Boomsma D.I. Genetics and epigenetics of human aggression. *Handb Clin Neurol*. 2023;197:13-44. doi 10.1016/B978-0-12-821375-9.00005-0

Pappa I., St Pourcain B., Benke K., Cavadino A., Hakulinen C., Nivard M.G., Nolte I.M., … Middeldorp C.M., Oldehinkel A.J., Pennell C.E., Boomsma D.I., Tiemeier H. A genome-wide approach to children's aggressive behavior: the EAGLE consortium. *Am J Med Genet B Neuropsychiatr Genet*. 2016;171(5):562-572. doi 10.1002/ajmg.b.32333

Pezzoli P., McCrory E.J., Viding E. Shedding light on antisocial behavior through genetically informed research. *Annu Rev Psychol*. 2025;76(1):797-819. doi 10.1146/annurev-psych-021524-043650

Ryan N.P., Catroppa C., Hughes N., Painter F.L., Hearps S., Beauchamp M.H., Anderson V.A. Executive function mediates the prospective association between neurostructural differences within the central executive network and anti-social behavior after childhood traumatic brain injury. *J Child Psychol Psychiatry*. 2021;62(9):1150-1161. doi 10.1111/jcpp.13385

Schwartz J.A., Solomon S.J., Valgardson B.A. Socialization, selection, or both? The role of gene-environment interplay in the association between exposure to antisocial peers and delinquency. *J Quant Criminol*. 2019;35:1-26. doi 10.1007/s10940-017-9368-3

Tesli N., Jaholkowski P., Haukvik U.K., Jangmo A., Haram M., Rokicki J., Friestad C., Tielbeek J.J., Næss Ø., Skardhamar T., Gustavson K., Ask H., Fazel S., Tesli M., Andreassen O.A. Conduct disorder – a comprehensive exploration of comorbidity patterns, genetic and environmental risk factors. *Psychiatry Res*. 2024;331:115628. doi 10.1016/j.psychres.2023.115628

Theadom A., Jones K., Starkey N., Barker-Collo S., Ameratunga S., Faulkner J., Ao B.T., Feigin V. Symptoms and engagement in antisocial behavior 10 years after mild traumatic brain injury within a community civilian sample: a prospective cohort study with age-sex matched control group. *Arch Phys Med Rehabil*. 2024;105(2):295-302. doi 10.1016/j.apmr.2023.07.016

Tielbeek J.J., Johansson A., Polderman T.J.C., Rautiainen M.R., Jansen P., Taylor M., Tong X., … Faraone S.V., Popma A., Medland S.E., Posthuma D.; Broad Antisocial Behavior Consortium collaborators. Genome-wide association studies of a broad spectrum of antisocial behavior. *JAMA Psychiatry*. 2017;74(12):1242-1250. doi 10.1001/jamapsychiatry.2017.3069

Tielbeek J.J., Uffelmann E., Williams B.S., Colodro-Conde L., Gagnon É., Mallard T.T., Levitt B.E., … Fisher S.E., Moffitt T.E., Caspi A., Polderman T.J.C., Posthuma D. Uncovering the genetic architecture of broad antisocial behavior through a genome-wide association study meta-analysis. *Mol Psychiatry*. 2022;27(11):4453-4463. doi 10.1038/s41380-022-01793-3

van Goozen S.H.M., Langley K., Hobson C.W. Childhood antisocial behavior: a neurodevelopmental problem. *Annu Rev Psychol*. 2022;73:353-377. doi 10.1146/annurev-psych-052621-045243

Wang S., Dan Y.L., Yang Y., Tian Y. The shared genetic etiology of antisocial behavior and psychiatric disorders: insights from pleiotropy and causality analysis. *J Affect Disord*. 2024;365:534-541. doi 10.1016/j.jad.2024.08.149

Watanabe K., Stringer S., Frei O., Umićević Mirkov M., de Leeuw C., Polderman T.J.C., van der Sluis S., Andreassen O.A., Neale B.M., Posthuma D. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*. 2019;51(9):1339-1348. doi 10.1038/s41588-019-0481-0

Weltens I., Bak M., Verhagen S., Vandenberk E., Domen P., van Amelsvoort T., Drukker M. Aggression on the psychiatric ward: prevalence and risk factors. A systematic review of the literature. *PLoS One*. 2021;16(10):e0258346. doi 10.1371/journal.pone.0258346

Conflict of interest. The authors declare no conflict of interest.
Received March 24, 2025. Revised May 22, 2025. Accepted June 4, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):847-855

doi 10.18699/vjgb-25-92

Endogenous oxytocin and intermale interactions after oxytocin administrations in Norway rats selected for behavior

S.G. Shikhevich [1], R.V. Kozhemyakina [1], R.G. Gulevich [1], Yu.E. Herbeck [1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] University of Haifa, Haifa, Israel
shikhsvt@bionet.nsc.ru

Abstract. The neuropeptide oxytocin (OT) secreted by specialized neurons in the hypothalamus affects social behavior and aggression in various animal species in a dose-dependent manner. Our earlier studies showed that OT administration by nasal application to adult and adolescent Norway rat males selected for enhanced aggressive response to humans reduced aggression upon the opponent in the resident-intruder test. By contrast, OT administration to rats selected for tame behavior exerted no effect on behavior or even enhanced aggression. It was still unknown how selection for behavior affected the endogenous oxytocinergic system in rats. Here we study the populations of OT-containing cells in the paraventricular and supraoptic nuclei of the hypothalamus in intact tame and aggressive rats with regard to lateralization, as the hypothalamus is known to be functionally asymmetrical. We have also assessed blood OT changes after nasal OT application to rats selected for behavior. As it is known that the effect of OT on rat aggressiveness may depend on the basal level of the latter, we have analyzed the effect of OT administration on behavior in tame and aggressive rats interacting on neutral ground, where the aggressiveness of males manifests itself less than in the defense of territory in the resident-intruder test. The asymmetry in the numbers of OT-containing cells in the left and right halves of the paraventricular and supraoptic nuclei has been observed only in tame rats. The number of such cells in the right half of tame rats is greater than in aggressive. In contrast, the blood OT level in tame rats is significantly lower than in aggressive ones both in the intact animals and after OT administration. Oxytocin administration to aggressive rats shortens aggressive interactions and lateral threats and reduces the number of the latter as compared to animals of the same behavior pattern having received saline. This observation may point to an anti-aggressive effect of OT. In tame rats, though, OT administration increases the number of hind leg kicks and kicking duration. It appears that the differences in the endogenous OTergic system of hypothalamus found in this study are associated with both the behavior formed during selection and different responses to exogenous OT in tame and aggressive animals.

Key words: oxytocin; selection; behavior; rat; aggressiveness; immunohistochemistry; hypothalamus

For citation: Shikhevich S.G., Kozhemyakina R.V., Gulevich R.G., Herbeck Yu.E. Endogenous oxytocin and intermale interactions after oxytocin administrations in Norway rats selected for behavior. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):847-855. doi 10.18699/vjgb-25-92

Funding. The maintenance of rat strains in the vivarium for conventional animals, ICG SB RAS, Novosibirsk, was supported by State Budgeted Project FWNR-0259-2022-0019.
Studies of the endogenous OTergic system and behavior of Norway rats after OT administration were supported by the Russian Science Foundation, project 21-44-04405.

Acknowledgements. Examination of hypothalamus nuclei in Norway rats was supported by the Shared Access Center for Microscopy of Biologic Objects, Siberian Branch of the Russian Academy of Sciences. We thank A.A. Sorokoumova for help in figure design and V.V. Gulevich for manuscript translation.

Authors' contribution. Yu.E.H and R.G.G. conceived of the presented idea and planned the experiment. S.G.Sh. and R.V.K. collected data. S.G.Sh. processed the data. All authors discussed the results and contributed to the final manuscript.

Эндогенный окситоцин и межсамцовые взаимодействия после введения окситоцина у серых крыс, селекционируемых по поведению

С.Г. Шихевич [1], Р.В. Кожемякина [1], Р.Г. Гулевич [1], Ю.Э. Гербек [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Хайфский университет, Хайфа, Израиль
shikhsvt@bionet.nsc.ru

Аннотация. Известно, что нейропептид окситоцин (ОТ), секретируемый специализированными нейронами в гипоталамусе, оказывает дозозависимое влияние на социальное поведение и агрессию у различных видов животных. Ранее нами было показано, что у взрослых и неполовозрелых самцов серых крыс, селекционируемых на усиление агрессивного поведения по отношению к человеку, назальные аппликации окситоцина вызывали антиагрессивный эффект по отношению к оппоненту в тесте резидент–интрудер, в то время как у крыс, селекционируемых на ручное поведение, аппликации окситоцина или не влияли на поведение, или вызывали усиление агрессивности. Однако оставалось неизвестным, как влияет отбор по поведению на эндогенную окситоцинергическую систему у крыс. В данной работе исследовали количество содержащих окситоцин клеток в паравентрикулярном (ПВЯ) и супраоптическом (СОЯ) ядрах гипоталамуса у интактных ручных и агрессивных крыс, учитывая фактор латерализации, поскольку было известно о функциональной асимметрии гипоталамуса. Наряду с этим оценивали, как изменяется уровень окситоцина в крови после назальных аппликаций этого нейропептида у крыс, селекционируемых по поведению. Так как эффекты окситоцина на агрессивность крыс могут зависеть от степени ее проявления, в данной работе исследовали влияние аппликаций окситоцина на поведение у ручных и агрессивных крыс при взаимодействии на нейтральной территории, где агрессивность самцов проявляется слабее, чем при защите собственной территории в тесте резидент–интрудер. Показано, что только для ручных крыс характерна асимметрия по количеству содержащих окситоцин клеток, локализованных в правых и левых частях СОЯ и ПВЯ гипоталамуса. Причем количество таких клеток в правой половине СОЯ у ручных крыс оказалось больше, чем у агрессивных, в то время как уровень окситоцина в крови у ручных крыс как в контрольной группе, так и после аппликаций окситоцина, напротив, был достоверно ниже, чем у агрессивных. Аппликации окситоцина у агрессивных крыс вызывали уменьшение продолжительности агрессивных взаимодействий и боковых стоек угроз, а также количества последних по сравнению с животными того же поведения, получающими физраствор, что может свидетельствовать об антиагрессивном эффекте окситоцина, тогда как у ручных крыс аппликации окситоцина, наоборот, приводили к увеличению количества ударов задними лапами и их продолжительности. По-видимому, найденные нами различия в эндогенной окситоцинергической системе гипоталамуса могут быть связаны и с поведением, сформированным в процессе отбора, и с различной реакцией на введение окситоцина у ручных и агрессивных крыс.

Ключевые слова: окситоцин; отбор; поведение; крыса; агрессивность; иммуногистохимия; гипоталамус

Introduction

Oxytocin (OT) had long been connected only with the reproductive function, maternal behavior, and nursing. Large-scale studies of the entire range of OT's physiological effects commenced in the mid-20th century. The results obtained in animals and, later, in humans, indicate that OT is essential for mitigating anxiety (Neumann, Slattery, 2016; Yoon, Kim, 2020; Takayanagi, Onaka, 2021) and aggression (Calcagnoli et al., 2013; de Jong, Neumann, 2018; Herbeck, Gulevich, 2019; Marsh et al., 2021) and for improving memory and learning (Aydogan et al., 2018).

The hypothalamic neuropeptide oxytocin is synthesized primarily in neurons of the paraventricular (PVN) and supraoptic (SON) hypothalamus nuclei. Then it is delivered to the posterior pituitary via axons and stored in vesicles until release to the systemic blood flow (Castel et al., 1984). Also, OT is transported via collateral branches of axons of the hypothalamo–neurohypophyseal tract to various parts of the forebrain, where OT receptors are expressed. There, it acts on various behavioral aspects (Jurek, Neumann, 2018; Grinevich, Neumann, 2021). Physiological stimuli, such as childbirth, lactation, stress, or emotions, induce a fast release of preaccumulated OT to blood and various parts of the brain (Eliava et al., 2016; Tang et al., 2020; Grinevich, Neumann, 2021).

The hypothalamus plays the key role in motivation formation and behavior triggering (Simonov 1987, 1993; Sudakov, 1993). The data on the functional hypothalamus asymmetry are scarce (Pavlova, 2001; Kiss, 2020). In particular, comparison of the efficiency of the stimulation of the right and left hypothalami in rabbits by the conjugated neuronal impulsation method in order to induce motivational and emotional responses showed that the left hypothalamus contributes more to defensive motivation, whereas the right hypothalamus, to emotional positive affects (Pavlova, 2001). The Norway rats selected for friendly and aggressive attitude to humans that were used in our studies differ significantly in the defense response to the experimenter's glove. This difference may be associated with the location of OT-secreting neurons in the animals.

I. Neumann et al. (2013) found that OT levels rose in blood and in dialysates of the dorsal hippocampus and amygdule of Wistar rat and C57Bl/6 mouse males 70 min after intranasal OT administration. They assume that the doses of exogenous OT in these brain regions were significant for behavior regulation.

F. Calcagnoli et al. (2014) examined Norway rats taken from the wild in Groningen (Netherlands), which differed from laboratory strains in greater and broadly variable male-to-male aggression. The scientists found that the OT mRNA level in PVN but not in SON inversely correlated with aggressiveness. The most aggressive males had lower OT mRNA levels in PVN than less aggressive ones. Oxytocin administration to the cerebral ventricular system of Norway male rats reduced their aggressiveness in a dose-dependent manner, which was most pronounced in males with higher aggression (Calcagnoli et al., 2013). Human studies also revealed differences in the efficiency of nasal OT adminis-

tration on prosocial behavior (signs of positive interactions with other individuals) depending on the preexperimental social motivation in the subject. The improvement of prosocial behavior in response to intranasal OT administration was most pronounced in people with low social motivation, but exogenous OT could aggravate interpersonal anxiety in people with a low level of social safety (Bartz et al., 2015; Soriano et al., 2020).

Studies of Norway rats selected for aggressive and tolerant attitude to humans can help understand features of the functioning of the endogenous OTergic system and the role of OT in behavior regulation. The Norway rat population selected for tame behavior demonstrates a complete absence of defense responses and high tolerance of handling by humans. In contrast, the aggressive rats are characterized not only by high aggression toward humans but also by invariable and extreme intraspecies male-to-male aggression (Plyusnina, Solov'eva, 2010; Plyusnina et al., 2011).

It had been found that the activity of both the peripheral and central segments of the hypothalamic–pituitary–adrenal (HPA) axis in tame rats was higher than in aggressive or unselected animals (Plyusnina, Oskina, 1997; Herbeck et al., 2017). The aggression of male rats selected for aggressive behavior toward the opponent in the test on neutral ground was weaker than in the resident–intruder test (RIT), where animals under study defended their own space (Plyusnina, Solov'eva, 2010). Nasal OT administration to adult and adolescent aggressive males suppressed aggression in the RIT test. In tame rats, OT administration at the 2 μg/μL concentration to adult and adolescent males either exerted no effect on behavior or even enhanced signs of aggression (Gulevich et al., 2019; Kozhemyakina et al., 2020).

It is unknown how rat selection for behavior affects the parameters of the hypothalamus OTergic system and OT level in peripheral blood or whether the effect of OT applications on rat behavior will remain in the test on neutral ground, where male aggression is milder than in the RIT test.

In this view, the goal of the present study was to investigate features of the function of the endogenous OTergic system in tame and aggressive male rats and the effect of nasal OT administration on blood OT level and manifestation of aggression in the test on neutral ground.

Materials and methods

Experimental animals. Experiments were conducted at the conventional vivarium of the Institute of Cytology and Genetics (Novosibirsk) with adult male Norway rats (*Rattus norvegicus*). Two populations of the animals, hereafter referred to as tame and aggressive, had been selected for absence or enhancement of aggressive response to humans, respectively, for 84–93 generations. Animals were kept in metal cages, 50 × 33 × 20 cm, four males per cage, and exposed to artificial light schedule 12/12. Food and water were given *ad libitum*. All tests were done from 14:00 to 18:00 local time.

All international, national, and/or institutional principles of animal care and use were followed. All procedures in experiments with animals met the ethical standards mandated by Russian Federation legislation, principles of the Basel Declaration, and recommendations of the Bioethics Committee at the Institute of Cytology and Genetics of the Russian Academy of Sciences (protocol No. 8, dated March 19, 2012).

Immunohistochemical examination of SON and PVN in the hypothalamus of intact tame and aggressive rats. The experiment was conducted with rats of selection generation 89. Brains were fixed by perfusion through the aorta left ventricle with a peristaltic pump with phosphate-buffered saline (1×PBS) for 5–10 min and with 4 % paraformaldehyde for 5–10 min. The extracted brains were dehydrated by immersion in 30 % sucrose at +4 °C for 3–4 weeks and stored in Tissue-Tek® O.C.T. Compound at –70 °C until cryosection. Frontal sections (30 μm) were prepared in a Microm HM-505 N cryostat (Microm, Germany) at –20 °C and mounted on Polysine® slides (Menzel-Gläser, Germany). The slides were stored at –20 °C until further examinations.

Slides were stained by the conventional method with the Rabbit specific HPR/DAB (ABC) Detection IHC kit ab64261 (Abcam, United Kingdom). Antibodies (Anti-Oxytocin-neurophysin 1 antibody ab2078 (Abcam)) were diluted 1/10,000. Slides were incubated with the antibodies for 3 days.

Oxytocin-containing cells in SONs and PVNs in slides were counted under an AXIO optical microscope (Zeiss, Germany) (Fig. 1). Oxytocin-containing cells in PVNs were arranged into compact groups; we chose the region with the greatest density of stained cells. Its area was 4.7 μm^2, and cells were counted there in all slides examined. We found no such specific region in SONs and determined the number of cells per 1 μm^2 of the section.

Nasal oxytocin applications and blood oxytocin assay thereafter. The experiment was conducted with rats of selection generation 93. Experimental groups included 10–12 animals. To minimize the stressing effect of the application of OT or saline, the animals had been habituated to handling for 7 days. On day 8, one group (OT group) received one application of 20 μL OT at the concentration of 2 μg/μL. Another group (saline group) received 20 μL of saline. Animals were decapitated 40 min after the procedure, and peripheral blood samples were taken to assay OT in the plasma. Intact rats were used as the control. Samples were taken into VACUETTE tubes (4 mL, 13 × 75 mm) with K3 EDTA and protease inhibitor aprotinin PREMIUM. Blood was centrifuged no later than 20 min after sampling. Plasma was immediately frozen at –20 °C and stored at –70 °C.

Oxytocin was assayed with DetectX® Oxytocin ELISA Kit (Arbor Assays, USA). To extract OT for assay, 200 mL of 0.05 M Tris HCl pH 8.0 and 5 μL of DTT (BioChemica, Pakistan) were added to 100 μL of plasma, and the mixture

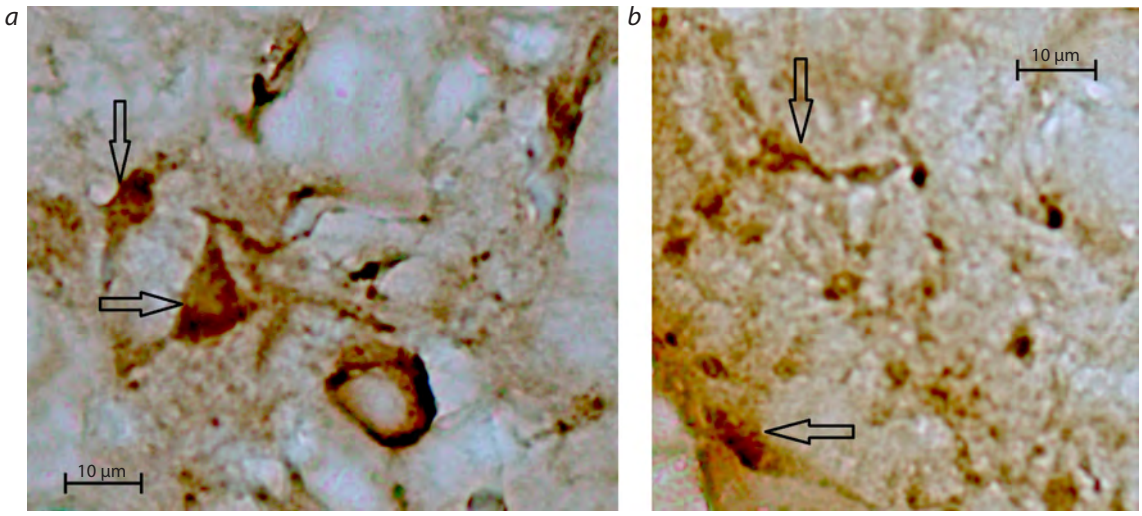


Fig. 1. Immunohistochemical staining of oxytocin-containing cells in (*a*) paraventricular (PVN) and (*b*) supraoptical (SON) nuclei of the rat hypothalamus.

Camera lens 40×.

was incubated at 37 °C for 45 min. Iodoacetamide (0.5 M, 15 μL) was added, and the mixture was incubated at room temperature for 20 min. Then 640 μL of 80 % acetonitrile was added, and the mixture was centrifuged at 14,000g for 15 min. The supernatant was collected and dried with Concentrator plus (Eppendorf, Germany).

Study of male-to-male interactions on neutral ground. The experiment was conducted with rats of selection generation 84. Experimental groups included 9–10 animals. Two weeks before the experiment, animals were placed into individual cages. In one group (OT group), each rat received one application of 20 μL of 1 μg/μL OT onto epithelium around the nostrils. Another group received 20 μL saline (saline group). After 40 min, an experimental animal was placed into an unfamiliar cage (40 × 40 × 60 cm) divided into two equal compartments with a partition (Plyusnina, Solov'eva, 2010). Simultaneously, a Wistar male of close bodyweight was placed into the second compartment, and the partition was removed. The behavior was camcorded for further analysis. The test lasted for 10 min.

Agonistic behavior was assessed by the following parameters: latency of the first aggressive interaction, number and duration of attacks, chases, hind leg kicks, rearings, backfalls, aggressive grooming, and lateral threats (Plyusnina, Solov'eva, 2010; Soriano et al., 2020). The overall time of aggressive behavior included the duration of all these postures and movements associated with competition or conflict between animals. In addition, we assessed the overall time of social nonaggressive behavior, which included approaches and sniffings.

Statistical evaluation. The results were evaluated with Statistica 10.0 (Stat Soft™, USA). Distribution normality was checked by the Kolmogorov–Smirnov test. Data on the numbers of OT-containing cells in hypothalamus nuclei and blood OT level, fitting into the normal distribution, were analyzed by parametric methods: Student's test and ANOVA for independent measurements followed by *post-hoc* Fisher's LSD test.

Data on the numbers of OT-containing cells in whole SONs and PVNs were analyzed by Student's *t*-test, and the numbers in the left and right halves of SON and PVN were assessed by two-way ANOVA, where one of the factors was selection for behavior, and the other, lateralization, that is, whether the count referred to the right or left portion of SON or PVN.

Blood OT level data were assessed by two-way ANOVA, where one factor was selection for behavior and the other, application of OT or saline. The test of data on behavior features did not confirm distribution normality; therefore, we applied the nonparametric Mann–Whitney U-test for further processing. The data on behavior features are presented in the form of box-plots with maximum, minimum, and median values, where 50 % of the results for a given sample are within the box, 25 % from the minimum value to the box bottom, and 25 % from the maximum to the box top.

Results

Immunohistochemistry

In our study, the number of OT-containing cells in the entire PVN did not differ significantly between tame and aggressive rats (774.76 ± 38.98 and 826.16 ± 35.80, respectively), whereas in the entire SON it tended to be greater in tame rats than in aggressive ones: 434.10 ± 28.76 *vs.* 331.68 ± 37.16 ($p < 0.06$).

Data on the numbers of OT-containing cells in separate halves of SON and PVN were processed with two-way ANOVA, where one factor was selection for behavior and the other, lateralization (count in the left or right half of SON or PVN).

The results indicate that selection for behavior did not affect the numbers of OT-containing cells in separate halves of PVN, whereas the effect of lateralization was significant ($F_{1.65}$ = 8.08, $p < 0.01$). No interaction of these factors was observed. *Post-hoc* analysis showed that the number of OT-containing cells in the left half of PVN was greater than in the right in tame rats but not in aggressive ones ($p < 0.05$, Fig. 2).

In contrast to PVN, the numbers of OT-containing cells in separate SON halves were significantly affected by the selection for behavior factor ($F_{1.56} = 4.2$, $p < 0.05$) but not by lateralization. The factor interaction was insignificant. According to the subsequent *post-hoc* analysis, the number of OT-containing cells in the right half of SON was significantly greater in tame rats than in aggressive ones ($p < 0.01$, Fig. 3), which agrees with the aforementioned trend toward the greater number of these cells in the entire SON in tame rats as compared to aggressive ones. In addition, the number of OT-containing cells in the right half of SON in tame rats was significantly greater than in the left half ($p < 0.05$).

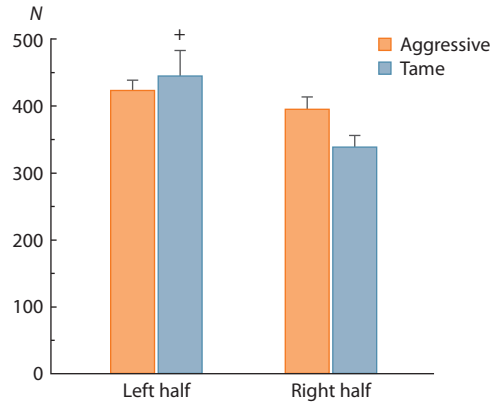


Fig. 2. The numbers of oxytocin-containing cells in the left and right halves of rat PVNs.

Two-way analysis of variance (ANOVA) followed by *post-hoc* Fisher's test: $^{+}$ $p < 0.05$ as compared to the right PVN. *N* – number of cells.

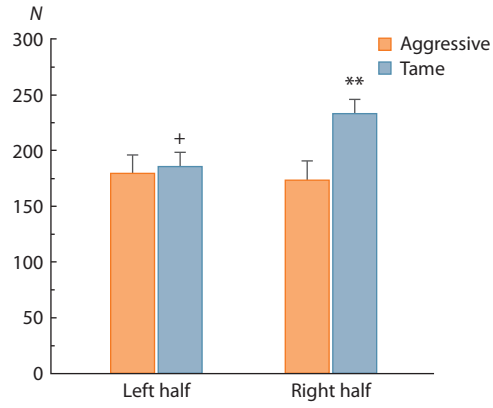


Fig. 3. The numbers of oxytocin-containing cells in the left and right halves of rat SON.

Two-way analysis of variance (ANOVA) followed by *post-hoc* Fisher's test: ** $p < 0.01$ as compared to aggressive rats, $^{+}$ $p < 0.05$ as compared to the right half of SON. *N* – number of cells.

Blood oxytocin after nasal applications

Two-way ANOVA demonstrated a significant influence of rat selection for behavior on the OT level in blood plasma ($F_{2.55} = 23.65$, $p < 0.001$) but no influence of OT administration. No factor interaction was found either. The subsequent *post-hoc* analysis showed that blood plasma OT level in tame rats was significantly lower than in aggressive ones in all groups examined: $p < 0.05$ in the control and saline groups; $p < 0.001$ in the OT group (Fig. 4). Saline applications did not affect blood OT in either tame or aggressive rats. The blood OT level in tame OT rats was higher than in the intact control ($p < 0.05$).

Effect of nasal oxytocin applications on rat behavior in the test for male-to-male interactions on neutral ground

Figure 5 presents data on behavior patterns in the test for male-to-male interactions on neutral ground, where significant differences were observed. In our experiments, the saline group of aggressive rats significantly exceeded the corresponding tame group in the overall time of aggressive interactions and the number and duration of lateral threats: $p < 0.03$, U = 17.5, Z = 2.2; $p < 0.035$, U = 18, Z = 2.16; $p < 0.04$, U = 19, Z = 2.08, respectively. The latency of aggression in aggressive animals after saline applications was significantly shorter than in tame ones: $p < 0.02$, U = 16, Z = –2.33.

Oxytocin applications to aggressive rats significantly reduced the overall time of aggressive interactions and the number and duration of lateral threats in comparison to rats of the same behavior group having received saline: $p < 0.04$, U = 19, Z = 2.08; $p < 0.035$, U = 18, Z = 2.16; $p < 0.025$, U = 17, Z = 2.24, respectively. Tame rats having received OT showed a significant increase in the number and duration of hind leg kicks as compared to the corresponding saline group: $p < 0.02$, U = 16, Z = –2.33 in both cases. Also, tame

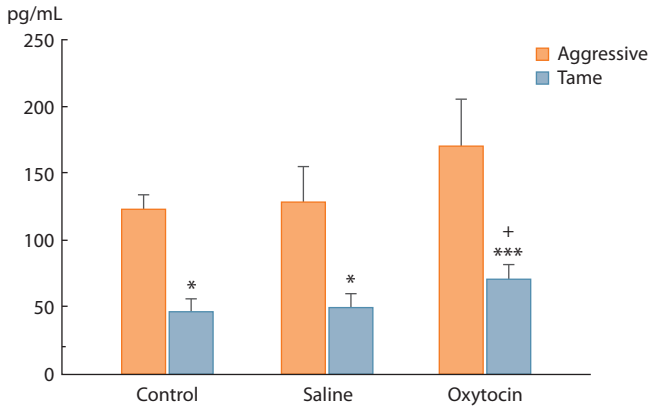


Fig. 4. Oxytocin levels, pg/mL, in rat blood plasma 40 min after the application of OT (20 μL, 2 μg/μL) or saline (20 μL).

Two-way analysis of variance (ANOVA) followed by *post-hoc* Fisher's test: * $p < 0.05$, *** $p < 0.001$ as compared to aggressive rats; $^{+}$ $p < 0.05$ as compared to the control.

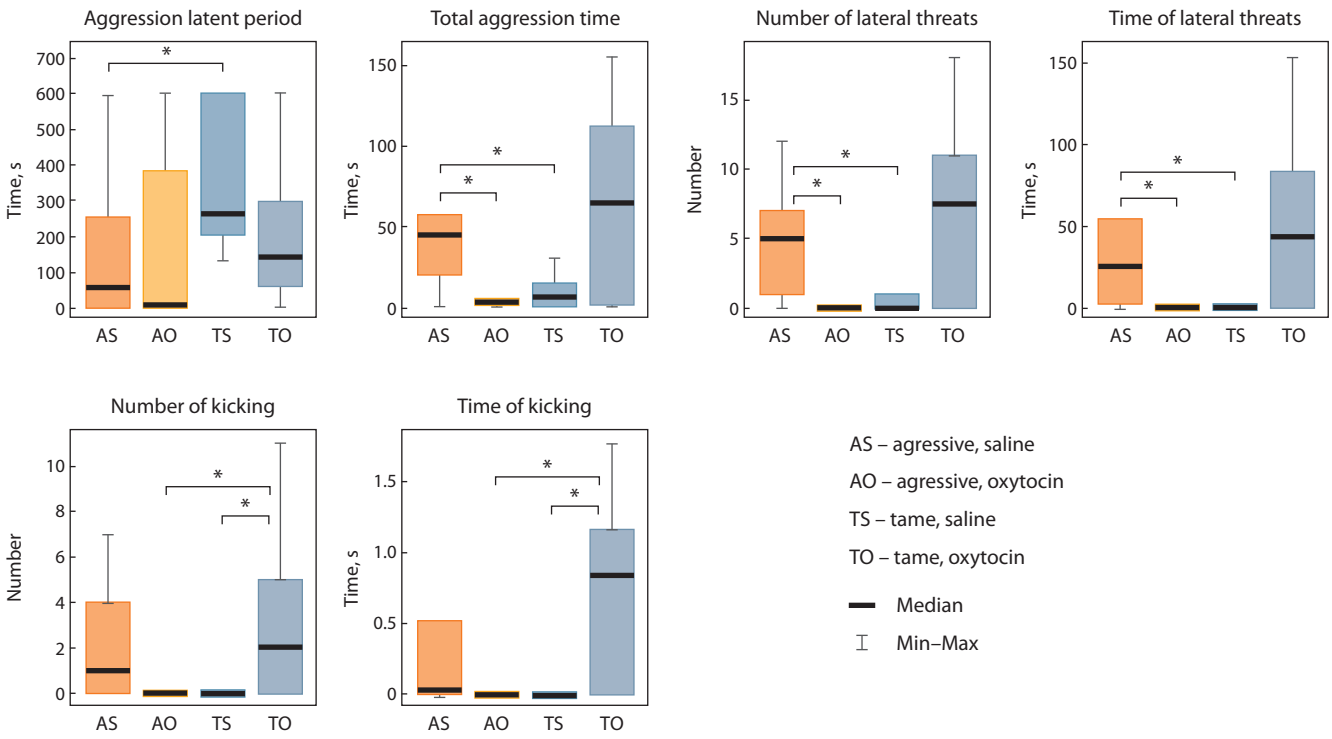


Fig. 5. Effect of nasal oxytocin applications (20 μL, 1 μg/μL) on behavior patterns in the test for male-to-male interactions on neutral ground.
*$p < 0.05$, Mann–Whitney U-test.

OT rats show a tendency toward longer aggressive interactions, and toward a greater number and duration of lateral threats than saline animals.

Comparison of behaviors in aggressive and tame rats after OT applications indicates that the number and duration of hind leg kicks in aggressive rats significantly decreased compared to tame ones: $p < 0.02$, U = 16, Z = –2.33 in both cases. Oxytocin administration leveled the differences between tame and aggressive rats in the overall time of aggressive interactions and the number and duration of lateral threats noted after saline administration.

Discussion

The immunohistochemical study of hypothalamus nuclei revealed an asymmetry in the numbers of OT-containing cells in SON and PVN in tame rats but not in aggressive ones. Data on the functional asymmetry of hypothalamus in rabbits obtained by conjugated neuronal impulsation in order to induce motivational and emotional responses indicate that the left hypothalamus contributed most to defensive motivation, and the right, to emotionally positive responses (Pavlova, 2001).

Our results indicate that the number of OT-containing cells in the right SON of rats selected for tame behavior is significantly greater than in aggressive ones (Fig. 2). As the rats selected for tame or aggressive behavior profoundly differ in the defense response to humans, we conjecture that the observed asymmetry in the numbers of OT-containing cells in the SON and PVN of tame rats and the greater number of such cells in the right half of SON in tame rats than in aggressive ones contribute to the difference in their attitude to humans.

The greater number of OT-containing cells in the right SON of tame rats as compared to aggressive ones is consistent with the observed tendency for the greater number of these cells in the entire SON ($p = 0.07$). In other words, it is reasonable to suggest that the number of OT-containing cells inversely correlates with rat aggressiveness, which is generally consistent with lower aggressiveness in aggressive rats after OT administration by injection to a brain ventricle or by nasal applications (Calcagnoli et al., 2013, 2015; Gulevich et al., 2019).

Experiments with Groningen male rats reveal an inverse correlation between OT mRNA levels in PVN and animal aggressiveness (Calcagnoli et al., 2014). These data, as well as those obtained in our studies, indicate that the functional parameters of the endogenous OTergic system in the hypothalamus of more aggressive rats are lower than in less aggressive (in Groningen rats) or tame (in our selection model) animals. Also, although aggressive male rats differed in the number of OT-containing cells in the right SON in our study, the Groningen rat groups differing in aggressiveness differed in the levels of OT mRNA in PVN (Calcagnoli et al., 2014). These features in the location of differences in rats studied by us and in the Groningen rats differing in aggressiveness may be related to different approaches to the formation of aggressive behavior and its evaluation. The aggressiveness

of Groningen rats was evaluated from the percentage of attack duration in ten RITs, and the aggressiveness in our study had been formed by long-term selection for attitude to humans.

It has been shown that blood OT in aggressive rats is significantly higher than in tame ones. As OT release to blood does not always correlate with local OT release from axon terminals in various regions of the brain (Knobloch et al., 2012; Grinevich et al., 2015), the assessment of brain OT system activity judging by blood OT level may be ambiguous, and OT variation in plasma does not necessarily correlate with animal behavior (Neumann, 2008). Forty minutes after nasal OT application, its blood level in tame rats was significantly higher than in intact animals but not higher than in the saline group, although this parameter did not differ significantly between the intact control and the saline group (Fig. 4). In Wistar rats, the blood OT level after OT application increased after 70, 100, and 130 min (Neumann et al., 2013). It seems that more time after nasal applications than in our study is required for OT elevation in blood and for the manifestation of its effects on behavior, which we observed as early as 40 min after applications both on neutral ground and in RIT (Gulevich et al., 2019; Kozhemyakina et al., 2020).

The behavior of aggressive rats in the test for male-to-male interactions on neutral ground points to an antiaggression effect of OT. The aggressive animals having received one OT application (20 μL, 1 μg/μL) differed from the *saline* aggressive group in showing significantly shorter aggressive interactions and lateral threats and in a smaller number of the latter. Similar effects of OT applications were noted in experiments with Groningen male rats in RIT (Calcagnoli et al., 2015) and in rats having been selected for aggressive behavior in interactions with an opponent male rat or with the experimenter's glove (Gulevich et al., 2019; Kozhemyakina et al., 2020). In aggressive rats, aggression lowered after single applications of OT at the concentration of 2 μg/μL and after five daily applications of OT at a lower concentration, 1 μg/μL, that is, at OT concentrations varying within a certain range. As it had been shown that the aggression of aggressive rats on neutral ground was weaker than in animals defending their own territory in RIT (Plyusnina, Solov'eva, 2010), it is conceivable that OT mitigates aggression regardless of its manifestation in different tests.

Tests of tame rats for behavior on neutral ground revealed an opposite effect of OT applications (20 μL, 1 μg/μL): a significant increase in the number of hind leg kicks and kicking duration. Tame OT rats also showed trends toward longer aggressive interactions, longer lateral threats, and a greater number of the latter as compared to tame saline animals (Fig. 5). The change of these behavior parameters is indicative of aggressive behavior enhancement in tame rats. Previous data of RIT after 5-day OT applications at the same dose as in this study (20 μL, 1 μg/μL) showed not only more hind leg kicks in tame rats but also longer aggressive interactions, longer lateral threats, and more attacks (Gulevich et al., 2019). Single OT applications at the higher concentration of 2 μg/μL exerted no significant effect on tame rat behavior in RIT (Gulevich et al., 2019).

Further studies are needed for the understanding of this inverse effect of higher OT doses. The enhancement of certain aggressive behavioral acts in tame rats having received exogenous OT in the stressing situation of interaction with an opponent on neutral ground as compared to the control is consistent with data from other scientists. For instance, studies on domesticated animals (dogs, cattle, and pigs) and humans demonstrate that OT in stress can induce a paradoxical response to neutral or even affiliative (or positive) interaction by enhancing aggression and stress response (Rault et al., 2013; Hernádi et al., 2015; Yayou et al., 2015; Crespi, 2016; Shamay-Tsoory, Abu-Akel, 2016).

An autoradiographic study of Groningen rats displaying maximum aggressiveness showed that OT receptor expression in the central amygdule and bed nucleus of stria terminalis was higher than in less aggressive animals (Calcagnoli et al., 2014). As thought by the scientists, this fact partially compensates for the lower OT expression in PVN in aggressive rats and sensitizes them to exogenous OT. The expression of OT receptors in both brain regions correlated with the duration of offensive aggression in a 10-min RIT. Thus, it can be expected that the expression of OT receptors in rats selected for aggressive behavior in the central amygdule and bed nucleus of stria terminalis would also be higher than in tame animals. Further inquiry is needed to evaluate OT receptor expression in these brain regions in rats selected for behavior.

As mentioned above, the greatest prosocial effects of nasal OT administration to humans were observed in persons with poor social motivation or those avoiding social communication (Bartz et al., 2015; Soriano et al., 2020). In our study, the antiaggression OT effect was noted only in aggressive rats, the behavior of which is characterized not only by high aggressiveness toward humans but also by stable and extreme manifestation of intraspecific male-to-male aggression; thus, these rats can serve as models for studying mechanisms that underlie neuropsychiatric diseases – autism, social anxiety, and depression – and for elaborating approaches to their treatment.

Studies on rats and voles point to a functional relationship between the OTergic system and the HPA axis (Neumann et al., 2000; Engelmann et al., 2004; Smith, Wang, 2014). The elevated activity of the endogenous OTergic system in lactation or after social interactions is followed by lower ACTH and corticosterone secretion at rest and after stressing events, such as the presence of a conspecific opponent (Neumann et al., 2000; Engelmann et al., 2004; Smith, Wang, 2014). The inhibiting effect of OT on the HPA axis in stress may be mediated by downregulation of the *Crh* gene for the corticotropin-releasing hormone in the hypothalamus (Jurek et al., 2015). It is conceivable that the greater number of OT-containing cells in the right SON of tame rats than in aggressive ones is associated with the lower *Crh*

expression in the hypothalamus, which had been found in our earlier studies of both intact animals and those having received OT applications (Herbeck et al., 2017; Gulevich et al., 2019).

Conclusion

The mitigation or, in contrast, enhancement of aggressive behavior induced by exogenous OT in rats selected for behavior is determined by differences in the activity of the endogenous OTergic system and, probably, its functional relationship with the HPA axis, which is essential in behavior regulation and is altered considerably by selection for behavior.

References

Aydogan G., Jobst A., Loy F., Dehning S., Zill P., Müller N., Kocher M. The effect of oxytocin on group formation and strategic thinking in men. *Horm Behav*. 2018;100:100-106. doi 10.1016/j.yhbeh.2018.02.003

Bartz J.A., Lydon J.E., Kolevzon A., Zaki J., Hollander E., Ludwig N., Bolger N. Differential effects of oxytocin on agency and communion for anxiously and avoidantly attached individuals. *Psychol Sci*. 2015;26(8):1177-1186. doi 10.1177/0956797615580279

Calcagnoli F., de Boer S.F., Althaus M., de Boer J.A., Koolhaas J.M. Antiaggressive activity of central oxytocin in male rats. *Psychopharmacology*. 2013;229(4):639-651. doi 10.1007/s00213-013-3124-7

Calcagnoli F., de Boer S.F., Beiderbec D.I., Althausc M., Koolhaas J.M., Neumann I.D. Local oxytocin expression and oxytocin receptor binding in the male rat brain is associated with aggressiveness. *Behav Brain Res*. 2014;261:315-322. doi 10.1016/j.bbr.2013.12.050

Calcagnoli F., Kreutzmann J.C., de Boer S.F., Althaus M., Koolhaas J.M. Acute and repeated intranasal oxytocinadministration exerts antiaggressive and pro-affiliative effects in male rats. *Psychoneuroendocrinology*. 2015;51:112-121. doi 10.1016/j.psyneuen.2014.09.019

Castel M., Gainer H., Dellmann H.D. Neuronal secretory systems. *Int Rev Cytol*. 1984;88:303-459. doi 10.1016/s0074-7696(08)62760-6

Crespi B.J. Oxytocin, testosterone, and human social cognition. *Biol Rev Camb Philos Soc*. 2016;91(2):390-408. doi 10.1111/brv.12175

de Jong T.R., Neumann I.D. Oxytocin and aggression. *Curr Top Behav Neurosci*. 2018;35:175-192. doi 10.1007/7854_2017_13

Eliava M., Melchior M., Knobloch-Bollmann H.S., Wahis J., da Silva Gouveia M., Tang Y., Ciobanu A.C., … Poisbeau P., Seeburg P.H., Stoop R., Charlet A., Grinevich V. A new population of parvocellular oxytocin neurons controlling magnocellular neuron activity and inflammatory pain processing. *Neuron*. 2016;89(6):1291-1304. doi 10.1016/j.neuron.2016.01.041

Engelmann M., Landgraf R., Wotjak C.T. The hypothalamic-neurohypophysial system regulates thehypothalamic-pituitary-adrenal axis under stress: an old concept revisited. *Front Neuroendocrinol*. 2004;25:132-149. doi 10.1016/j.yfrne.2004.09.001

Grinevich V., Neumann I. Brain oxytocin: how pazzle stones from animal studies translate into psychiatry. *Mol Psychiatry*. 2021;26(1):265-279. doi 10.1038/s41380-020-0802-9

Grinevich V., Desarménien M., Chini B., Tauber M., Muscatelli F. Ontogenesis of oxytocin pathways in the mammalian brain: late maturation and psychosocial disorders. *Front Neuroanat*. 2015;8:164. doi 10.3389/fnana.2014.00164

Gulevich R., Kozhemyakina R., Shikhevich S., Konoshenko M., Herbeck Yu. Aggressive behavior and stress response after oxytocin administration in male Norway rats selected for different attitudes to humans. *Physiol Behav*. 2019;199:201-218. doi 10.1016/j.physbeh.2018.11.030

Herbeck Y.E., Gulevich R.G. Neuropeptides as facilitators of domestication. *Cell Tissue Res*. 2019;375(1):295-307. doi 10.1007/s00441-018-2939-2

Herbeck Y.E., Amelkina O.A., Konoshenko M.Yu., Shikhevich S.G., Gulevich R.G., Kozhemyakina R.V., Plyusnina I.Z., Oskina I.N. Effects of neonatal handling on behavior and stress response in rats selected for their reaction towards humans. *Russ J Genet Appl Res*. 2017;7(1):71-81. doi 10.1134/S2079059717010051

Hernádi A., Kis A., Kanizsár O., Tóth K., Miklósi B., Topál J. Intranasally administered oxytocin affects how dogs (*Canis familiaris*) react to the threatening approach of their owner and an unfamiliar experimenter. *Behav Processes*. 2015;119:1-5. doi 10.1016/j.beproc.2015.07.001

Jurek B., Neumann I. The oxytocin receptor: from intracellular signaling to behavior. *Physiol Rev*. 2018;98(3):1805-1908. doi 10.1152/Physrev.00031.2017

Jurek B., Slattery D.A., Hiraoka Y., Liu Y., Nishimori K., Aguilera G., Neumann I.D., van den Burg E.H. Oxytocin regulates stress-induced *Crf* gene transcription through CREB-regulated transcription coactivator 3. *J Neurosci*. 2015;5(35):12248-12260. doi 10.1523/JNEUROSCI.1345-14.2015

Kiss D.S., Toth I., Jocsak G., Barany Z., Bartha T., Frenyo L.V., Horvath T.L., Zsarnovszky A. Functional aspects of hypothalamic asymmetry. *Brain Sci*. 2020;10(6):389. doi 10.3390/brainsci10060389

Knobloch S., Charlet A., Hoffmann L.C., Eliava M., Khrulev S., Cetin A.H., Osten P., Schwarz M.K., Seeburg P.H., Stoop R., Grinevich V. Evoked axonal oxytocin release in the central amygdala attenuates fear response. *Neuron*. 2012;73(3):553-566. doi 10.1016/j.neuron.2011.11.030

Kozhemyakina R.V., Shikhevich S.G., Konoshenko M.Yu., Gulevich R.G. Adolescent oxytocin treatment affects resident behavior in aggressive but not tame adult rats. *Physiol Behav*. 2020;224:113046. doi 10.1016/ j.physbeh.2020.113046

Marsh N., Marsh A.A., Lee M.R., Hurlemann R. Oxytocin and the neurobiology of prosocial behavior. *Neuroscientist*. 2021;27(6):604-619. doi 10.1177/1073858420960111

Neumann I. Brain oxytocin: a key regulator of emotional and social behaviours in both females and males. *J Neuroendocrinol*. 2008;20(6):858-865. doi 10.1111/j.1365-2826.2008.01726.x

Neumann I.D., Slattery D.A. Oxytocin in general anxiety and social fear: a translational approach. *Biol Psychiatry*. 2016;79(3):213-221. doi 10.1016/j.biopsych.2015.06.004

Neumann I., Wigger A., Torner L., Holsboer F., Landgraf R. Brain oxytocin inhibits basal and stress-induced activity of the hypothalamo-pituitary adrenal axis in male and female rats: partial action within the paraventricular nucleus. *J Neuroendocrinol*. 2000;12(3):235-243. doi 10.1046/j.1365-2826.2000.00442.x

Neumann I., Maloumby R., Beiderbeck D.I., Lukas M., Landgraf R. Increased brain and plasma oxytocin after nasal and peripheral administration in rats and mice. *Psychoneuroendocrinology*. 2013;38(10):1985-1993. doi 10.1016/j.psyneuen.2013.03.003

Pavlova I.V. Functional brain asymmetry in motivational and emotional conditions: Doctor Sci (Biol.) Dissertation. Moscow, 2001 (in Russian)

Plyusnina I., Oskina I. Behavioral and adrenocortical responses to open-field test in rats selected for reduced aggressiveness toward humans. *Physiol Behav*. 1997;61(3):381-385. doi 10.1016/s0031-9384(96)00445-3

Plyusnina I., Solov'eva M. Intraspecific intermale aggression in tame and aggressive Norway rats. *Zhurnal Vysshei Nervnoi Deyatelnosti imeni I.P. Pavlova*. 2010;60(2):175-183 (in Russian)

Plyusnina I.Z., Solov'eva M.Y., Oskina I.N. Effect of domestication on aggression in gray Norway rats. *Behav Genet*. 2011;41(4):583-592. doi 10.1007/s10519-010-9429-y

Rault J.-L., Carter C.S., Garner J.P., Marchant J.N., Richert B.T., Lay D.C. Jr. Repeated intranasal oxytocin administration in early life dysregulates the HPA axis and alters social behavior. *Physiol Behav*. 2013;112-113:40-48. doi 10.1016/j.physbeh.2013.02.007

Shamay-Tsoory S.G., Abu-Akel A. The social salience hypothesis of oxytocin. *Biol Psychiatry*. 2016;79(3):194-202. doi 10.1016/j.biopsych.2015.07.020

Simonov P.V. The Motivated Brain. Moscow: Nauka Publ., 1987 (in Russian)

Simonov P.V. The nerve centers of the emotions. *Zhurnal Vysshei Nervnoi Deyatelnosti imeni I.P. Pavlova*. 1993;43(3):514-529 (in Russian)

Smith A.S., Wang Z. Hypothalamic oxytocin mediates social buffering of the stress response. *Biol Psychiatry*. 2014;76(4):281-288. doi 10.1016/j.biopsych.2013.09.017

Soriano J.R., Daniels N., Prinsen J., Alaerts K. Intranasal oxytocin enhances approach-related EEG frontal alpha asymmetry during engagement of direct eye contact. *Brain Commun*. 2020;2(2):fcaa093. doi 10.1093/braincomms/fcaa093

Sudakov K.V. The neurophysical grounds of dominating motivation. *Vestnik Rossiiskoy Akademii Meditsinskikh Nauk = Annals of the Russian Academy of Medical Sciences*. 1993;7:42-48 (in Russian)

Takayanagi Y., Onaka T. Roles of oxytocin in stress responses, allostasis and resilience. *Int J Mol Sci*. 2021;23(1):150. doi 10.3390/ijms23010150

Tang Y., Benusiglio D., Lefevre A., Hilfiger L., Althammer F., Bludau A., Hagiwara D., … Stern J.E., Leng G., Neumann I.D., Charlet A., Grinevich V. Social touch promotes interfemale communication via activation of parvocellular oxytocin neurons. *Nat Neurosci*. 2020;23(9):1125-1137. doi 10.1038/s41593-020-0674-y

Yayou K., Ito S., Yamamoto N. Relationships between postnatal plasma oxytocin concentrations and social behaviors in cattle. *Anim Sci J*. 2015;86(8):806-813. doi 10.1111/asj.12363

Yoon S., Kim Yu. The role of the oxytocin system in anxiety disorders. *Adv Exp Med Biol*. 2020;1191:103-120. doi 10.1007/978-981-32-9705-0_7

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

Received December 4, 2024. Revised March 12, 2025. Accepted June 6, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):856-867

doi 10.18699/vjgb-25-93

Agrobacterium-derived DNA sequences in phylogenetic studies of plants

T.V. Matveeva [1] ✉, P.M. Zhurbenko [2], G.V. Khafizova [3], A.D. Shaposhnikov [1], R.R. Zhidkin [1], A.V. Rodionov [1, 2]

[1] Saint-Petersburg State University, St. Petersburg, Russia
[2] Komarov Botanical Institute of the Russian Academy of Sciences, St. Petersburg, Russia
[3] University of Houston, Houston, United States
✉ radishlet@gmail.com

Abstract. One of the main methods for obtaining transgenic plants is *Agrobacterium*-mediated transformation. This process relies on the ability of certain soil bacteria, specifically from the genera *Agrobacterium* and *Rhizobium*, to transfer and integrate a fragment of their plasmid into the chromosome of the recipient plant. This transferred DNA is referred to as T-DNA. Laboratory studies have demonstrated that whole plants can be regenerated from transgenic cells. It soon became evident that similar processes occur in nature, leading to the emergence of naturally transgenic plants, or natural GMOs. Thus, naturally transgenic plants possess homologues of the T-DNA genes from agrobacteria in their genomes (cellular T-DNA, or cT-DNA). These sequences are inherited through multiple sexual generations and retain their functionality. Furthermore, the potential for using newly acquired plant sequences in phylogenetic studies has been established, as cT-DNAs are clearly defined, highly specific, and recognizable DNA fragments that differ from typical plant DNA sequences. They are not found in untransformed ancestors, and their integration at specific chromosomal sites marks a monophyletic group of species. This review highlights the diversity of cellular T-DNAs and their potential use as phylogenetic markers. It includes a description of the main methodological approaches to such studies and discusses specific examples that clarify controversial points in the phylogeny of the genera *Nicotiana*, *Camellia*, *Vaccinium*, and *Arachis*. An important aspect of phylogenetic analysis based on cT-DNA is the assembly of individual alleles, which enables the tracking of interspecific hybridization events. This approach demonstrated the incomplete process of speciation within the *Thea* section of the genus *Camellia* and confirmed the role of interspecific hybridization in the breeding of North American blueberries. The review also addresses the dating of transformation events based on cT-DNA, which are organized in the form of imperfect repeats, as well as the application of phylogenetic studies to investigate the biodiversity of agrobacterial T-DNA genes.

Key words: agrobacterium-mediated transformation; cellular T-DNA; phylogenetics; *Nicotiana*; *Camellia*; *Vaccinium*; *Arachis*

For citation: Matveeva T.V., Zhurbenko P.M., Khafizova G.V., Shaposhnikov A.D., Zhidkin R.R., Rodionov A.V. Agrobacterium-derived DNA sequences in phylogenetic studies of plants. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):856-867. doi 10.18699/vjgb-25-93

Funding. The work was supported by the Russian Science Foundation, grant 25-26-00123.

Последовательности ДНК агробактериального происхождения в филогенетических исследованиях растений

Т.В. Матвеева [1] ✉, П.М. Журбенко [2], Г.В. Хафизова [3], А.Д. Шапошников [1], Р.Р. Жидкин [1], А.В. Родионов [1, 2]

[1] Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
[2] Ботанический институт им. В.Л. Комарова Российской академии наук, Санкт-Петербург, Россия
[3] Хьюстонский университет, Хьюстон, США
✉ radishlet@gmail.com

Аннотация. Агробактериальная трансформация – наиболее распространенный метод получения трансгенных растений. Метод основан на способности определенных почвенных бактерий родов *Agrobacterium* и *Rhizobium* переносить и интегрировать в хромосому растения-реципиента фрагмент своей плазмиды. Этот фрагмент получил название Т-ДНК (transferred DNA – переносимая ДНК). В лабораторных условиях было показано, что из трансгенных клеток можно регенерировать целые растения. Вскоре стало ясно, что подобные процессы происходят и в природе, приводя к появлению природно-трансгенных растений или природных генно-инженерно-модифицированных организмов (ГМО). Таким образом, природно-трансгенными называются растения, у которых в геномах присутствуют гомологи генов Т-ДНК агробактерий (клеточные Т-ДНК, клТ-ДНК). Эти последовательности наследуются в ряду половых поколений и сохраняют свою функциональность. Кроме того,

продемонстрирована возможность использования новоприобретенных растениями последовательностей в филогенетических исследованиях, поскольку клТ-ДНК являются четко определенными, высокоспецифичными и узнаваемыми фрагментами ДНК, не похожими на последовательности ДНК растений. Они не встречаются у нетрансформированных предков, и их интеграция в определенном хромосомном сайте маркирует монофилетическую группу видов. В представленном обзоре освещены вопросы разнообразия клТ-ДНК, возможности их применения как филогенетических маркеров, в том числе описаны основные методические подходы таких работ, на конкретных примерах рассмотрены возможности уточнения спорных моментов в филогении родов *Nicotiana*, *Camellia*, *Vaccinium* и *Arachis*. Важным моментом филогенетического анализа на основе клТ-ДНК является реконструкция отдельных аллелей. Это дает возможность отслеживать факты межвидовой гибридизации. Именно этот подход позволил продемонстрировать незавершенный процесс видообразования в пределах секции *Thea* рода *Camellia*, а также подтвердил использование межвидовой гибридизации в ходе селекции североамериканских голубик. В обзоре рассмотрены вопросы датировки событий трансформации на основе клТ-ДНК, организованных в виде несовершенных повторов, а также использования филогенетических исследований при изучении вопросов биоразнообразия генов Т-ДНК агробактериального происхождения.

Ключевые слова: агробактериальная трансформация; клеточная Т-ДНК; филогенетика; *Nicotiana*; *Camellia*; *Vaccinium*; *Arachis*

Introduction

Agrobacterium-mediated transformation is currently the most common method for producing transgenic plants to meet the needs of agriculture, medicine, veterinary science, and other sectors of the national economy. This method relies on the ability of soil bacteria, specifically *Agrobacterium tumefaciens* (Smith and Townsend 1907) Conn 1942 (Approved Lists 1980) and *Rhizobium rhizogenes* (Riker et al. 1930) Young et al., 2001, to transfer a fragment of their plasmid and integrate it into plant DNA (Schell, Van Montagu, 1977; Bahramnejad et al., 2019). The transferred sequences are referred to as T-DNA (transferred DNA).

Under natural conditions, the transfer and integration of T-DNA into the host plant chromosome typically stimulate the growth of transgenic tissue due to the expression of T-DNA genes that regulate morphogenesis (Nester, 2015). However, the regeneration of fully transgenic plants from such tissues has also been documented (Tepfer, 1990; Christey, 2001). Numerous observations suggest that this process can occur in nature without human intervention, resulting in the emergence of plants containing bacterial sequences in their genomes, which can be passed down through successive sexual generations (Matveeva, 2021). The first naturally transgenic plants were identified among species of the genus *Nicotiana* in 1983 (White et al., 1982, 1983), and currently, dozens of genera of dicots have been documented to contain naturally transgenic species (Matveeva, Otten, 2019; Matveeva, 2021). T-DNA found in the genomes of natural GMOs (nGMOs) is referred to as cellular T-DNA (cT-DNA) (Matveeva, 2021). Comparing the number of species with sequenced genomes to the number of nGMOs among them, we find that traces of *Agrobacterium*-mediated transformation have been preserved in seven percent of dicotyledonous plant species (Matveeva, Otten, 2019). This provides scientists with extensive material to study the functions of bacterial genes in plants and their evolutionary trajectories.

Additionally, interesting data have been obtained using T-DNA as a phylogenetic marker, as newly acquired sequences offer several advantages for studying the origin and evolution of nGMO species (Matveeva et al., 2011; Chen et al., 2022; Zhidkin et al., 2023; Bogomaz et al., 2024). cT-DNA are clearly defined, highly specific, and recognizable DNA fragments that differ from plant DNA sequences. They are absent in untransformed ancestors, and their integration at specific chromosomal sites marks a monophyletic group of species. Typically, cT-DNAs consist of single insertions, which is a significant advantage over classical nuclear markers. Furthermore, cT-DNAs can be quite long and ancient, accumulating single nucleotide substitutions (SNS), which allows for the construction of phylogenetic trees. The probability that the same T-DNA sequence will integrate into the genomes of two independent phylogenetic branches at the same target site with identical boundaries seems unlikely, but not impossible. This makes T-DNA insertions, along with transposon insertions into the genome, extremely important synapomorphies (Shedlock, Okada, 2000; Doronina et al., 2022) and, therefore, powerful tools for systematics. Their genome-wide analysis may also help identify the causes of phylogenetic signal conflicts should they arise (Kuritzin et al., 2016). Lastly, repetitive sequences in cT-DNA can be used to estimate the time since transformation or serve as relative time markers (Chen et al., 2022).

T-DNA structures

The physical structure of T-DNA varies among plasmids from different *Agrobacterium* strains and can be classified based on the number of plant-transferable fragments encoded in a single plasmid. T-DNA can be continuous, as seen in mannopine, mikimopine, and cucumopine strains (Jouanin, 1984; Hansen et al., 1991). This continuous fragment of DNA is flanked by border sequences at both ends, as exemplified by plasmid pRi8196. In contrast, other plasmids, such as pRiA4, have T-DNA divided into two segments: TL-DNA and TR-DNA (White et al., 1985), separated by a non-plant-transferable DNA region of about 15 kb that acts as a spacer. Regardless of their structural differences, all aforementioned T-DNAs contain genes for the synthesis of opines, which serve as a food source for *Agrobacterium*, as well as oncogenes, the products of which induce plant cell division.

The structures of nGMO cT-DNAs are more diverse, with those containing only opine synthesis genes being the most prevalent. This phenomenon can be attributed to at least three factors. First, in the known T-DNAs, the opine synthesis genes

are typically located closer to the right border. During transformation, when single-stranded T-DNA is excised from the plasmid, the VirD2 protein covalently binds to the 5′ end of the T-DNA at its right border and subsequently directs the T-strand through the type IV protein secretion system into the plant cell (Gelvin, 2021). Deletions that occur during the integration of T-DNA into the plant chromosome tend to be more extensive at the 3′ end than at the 5′ end, which is spatially protected by the VirD2 protein (Gelvin, 2021). Secondly, it is plausible that there exist *Agrobacterium* strains, the T-DNA of which contains only opine synthase genes. Thirdly, it is possible that significant portions of the T-DNA may have been lost after integration into the chromosome during the evolution of the descendants of a natural transformant, retaining only the opine synthesis genes. These genes do not influence morphogenesis as significantly as oncogenes, meaning they do not lead to growth and developmental abnormalities (Matveeva, Otten, 2021).

The other most common structures are extended T-DNA fragments containing both oncogenes and opine synthesis genes. These cT-DNAs are typically represented by inverted imperfect repeats, a feature that can be used to date transformation events. The least common are cT-DNAs containing only oncogenes (Matveeva, Otten, 2019; Matveeva, 2021).

Let us now explore in more detail the issue of dating the emergence of nGMOs throughout evolution.

Approaches to determining the timing of transformation events in nGMOs

The dating of transformation events in nGMOs can be performed using the molecular clock method. The peculiarities of T-DNA integration into the genome of an infected plant lead to the formation of repeating T-DNA sequences or regions (Tzfira et al., 2004; Singer, 2018). These repeats are long cT-DNAs, represented by imperfect inverted repeats or their fragments formed during region deletions or insertions (Matveeva, Otten, 2019; Matveeva, 2020). By analyzing the differences in the nucleotide sequences of these repeats, it is possible to estimate the approximate time of divergence and, consequently, the time of T-DNA integration into the genome of the future nGMO, using the nucleotide substitution coefficient (Gaut et al., 1996).

The transformation time can be calculated using the following formula:

$$T = \frac{d}{2r},$$

where T – the approximate integration time of cT-DNA; d – the ratio of nucleotide differences between two repeats; r – the average nucleotide substitution rate of 6.5×10^{-9} substitutions per site per year (Gaut et al., 1996; Lynch, Conery, 2000). This approach was used to date T-DNA integration in representatives of the *Camellia* genus, which possess various structural types of cT-DNA (Chen et al., 2022, 2023). The dating revealed that the oldest integration event occurred approximately 7.5 million years ago, while the youngest took place around 0.04 million years ago (Chen et al., 2023)[1].

[1] It was clarified through correspondence with the authors that the insertion times reported in the article were calculated using an erroneous formula, resulting in values that were overestimated by a factor of two. This review presents the corrected values.

However, this method has several limitations that hinder its application for determining the integration times of many cT-DNAs. The first limitation is the inability to detect inverted repeats in all cT-DNAs, as most nGMO species do not contain extended cT-DNAs where repeating direct and inverted sequences are typically found (Matveeva, 2021). Secondly, not all extended cT-DNAs contain repeats; for instance, only 6 out of 12 types of cT-DNAs in members of the genus *Camellia* possess repeats that facilitate the determination of integration time (Chen et al., 2023). An additional critical consideration when selecting sequences for dating is the necessity to exclude the possibility that these repeats formed as a result of other genomic rearrangements, such as the activity of mobile genomic elements.

The method of dating cT-DNA should be employed in conjunction with other approaches to accurately assess the nucleotide substitution rate in a particular species and to compare divergence times with closely related taxa for validation. Although this method of dating nGMO transformation requires further development and verification, it offers valuable insights into the evolutionary processes that nGMOs have undergone following *Agrobacterium*-mediated transformation.

Multiple transformation events in the evolution of the genus *Nicotiana*

The study of genetic transformation in plants occurring naturally, without human intervention, began with the species *Nicotiana glauca* Graham, which was found to contain sequences homologous to agrobacterial T-DNA in its nuclear genome (White et al., 1983). The first identified cT-DNA was designated gT. This sequence was organized as an imperfect inverted repeat and consisted of one copy of the *rolB* homolog, along with two copies each of the *rolC*, *ORF13*, *ORF14*, and *mis* homologs (Suzuki et al., 2002). The gT sequence served as a reference point for the search for cT-DNA in the genomes of other *Nicotiana* L. species (Furner et al., 1986; Intrieri, Buiatti, 2001). To date, 16 naturally transgenic species of this genus have been identified (Otten, 2020). The genus *Nicotiana* comprises twelve sections (Knapp et al., 2004), six of which have naturally occurring transgenic representatives described.

The search for and analysis of cT-DNA in plant genomes were initially conducted at the level of individual genes (Furner et al., 1986; Meyer et al., 1995; Intrieri, Buiatti, 2001). During this time, efforts were made to reconstruct the phylogenetic relationships between cT-DNA of various tobacco species and T-DNA of *Agrobacterium*. However, subsequent studies have called some earlier evolutionary models into question. The transition to whole-genome data has enabled researchers not only to search for cT-DNAs and analyze their composition but also to estimate their quantity and localization within the plant genome. The first such analysis was conducted for *N. tomentosiformis* Goodsp. (Chen et al., 2014), where four distinct cT-DNA types (TA, TB, TC, and TD) were identified (see the Table), differing from the previously studied gT in *N. glauca* (Suzuki et al., 2002). This finding highlighted the importance of assessing the number of cT-DNAs in the genome and their localization sites as a means to identify the descendants of specific transformation events.

Multiple extended cT-DNAs in tobacco species have been used to date transformation events in evolution. In *N. to-*

cT-DNA structures in *N. tomentosiformis*

Name	List of genes (arms of imperfect inverted repeats are shown in parentheses)
TA	(*orf8*, *orf3*, *rolA*, *rolB*, *rolC*, *orf13*, *orf14*, *mis*) (*mis*, *orf14*, *orf13*, *rolC*, *rolB*, *rolA*, *orf8*)
TB	(*mis*, *orf14*) (*orf14*, *mis*, *ags*-like, *mas1′*, *mas2′*)
TC	(*ocl*, *orf2*, *orf3*, *orf8*, *rolA*, *rolB*) (*rolB*, *rolA*, *orf8*, *orf3*, *orf2*, *c*)
TD	(*orf14*), *orf55*, *orf 15*, *orf511*, (*orf1*)

mentosiformis, all four cT-DNAs were shown to result from independent *Agrobacterium*-mediated transformation events that occurred at different times. The cT-DNA TC was identified as the oldest of the four, with an estimated age of 1 million years (Chen et al., 2014). Later, in the genome of *N. otophora* Griseb., a species phylogenetically close to *N. tomentosiformis*, two TC copies were found that differed by 4 % and were located at the same site as the TC of *N. tomentosiformis*. The common localization site suggests that the cT-DNA TC was most likely acquired by a common ancestral species of *N. tomentosiformis* and *N. otophora* (Fig. 1).

During the evolution of this species, a TC duplication likely occurred, leading to the formation of *N. otophora*, which carries two TC copies in its genome, and *N. tomentosiformis*, which lost one TC copy during speciation (Chen et al., 2018). Attempts to reconstruct the evolutionary events, including speciation, in the genus *Nicotiana* are complicated by its characteristic reticulate or mesh-like evolution, marked by partial fusion of ancestral branches and the formation of hybrid forms (Knapp et al., 2004). For example, a well-known representative of the genus *Nicotiana*, *N. tabacum* L. (cultivated tobacco), is an interspecific hybrid of *N. tomentosiformis* and *N. sylvestris* Speg. (Yukawa et al., 2006). The genome of *N. tabacum* contains three of the four cT-DNAs found in *N. tomentosiformis* (TA, TB, TD), suggesting that *Agrobacterium*-mediated transformation events in *N. tomentosiformis* preceded the speciation of *N. tabacum* (Chen et al., 2014). This assumption is supported by flow cytometry and genomic hybridization estimates, indicating that the age of *N. tabacum* is less than 600 years (Leitch et al., 2008). The TC cT-DNA, a copy of which is present in the genome of the ancestral species *N. tomentosiformis*, was likely lost during the speciation of *N. tabacum* (Chen et al., 2014) (Fig. 1). Using TC cT-DNA dating data and speciation time estimates, it is possible to trace the TC path from its initial transfer into the plant genome to its loss or consolidation in various species that originated from the ancestral form at different stages of the genus's evolution.

All the representatives described above belong to the *Tomentosae* and *Nicotiana* sections. For representatives of the *Noctiflorae* section (*N. glauca*, *N. noctiflora* Hook.), three different cT-DNAs acquired through independent transformation events have also been documented (one insertion in the genome of *N. glauca* and two in *N. noctiflora*) (Khafizova et al., 2023). However, additional naturally transgenic representatives with sequenced genomes are needed to elucidate the phylogenetic relationships in this branch.

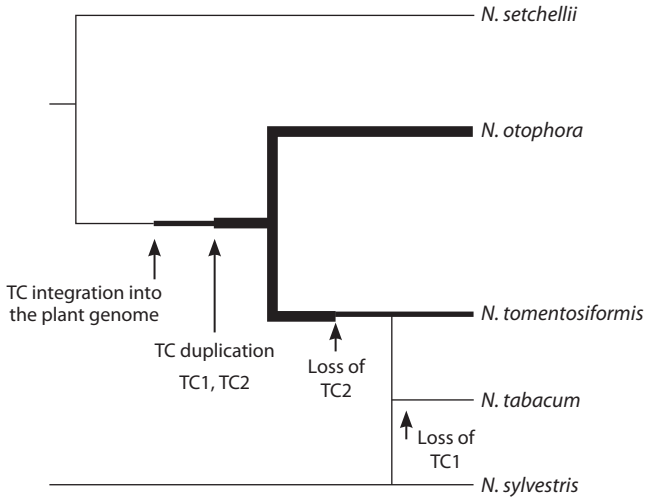


Fig. 1. Model of TC cT-DNA evolution in *Nicotiana* genomes.
The width of the lines reflects the presence of one or two TC copies in the species genome, or their absence (according to Chen et al., 2018).

So, the degree of study of the taxon is directly related to the availability of material and its economic significance. The intraspecific diversity of cultivated tobacco is also of interest; therefore, in the next section, we will focus on the application of cT-DNA in the study of this topic.

Intraspecific variability of cT-DNA in tobacco

Nicotiana tabacum, or cultivated tobacco, is an allotetraploid formed through the interspecific hybridization of *N. tomentosiformis* and *N. sylvestris* (Yukawa et al., 2006). There are tens of thousands of existing varieties and species of *N. tabacum*, though the relationships among them have not been fully established (Moon et al., 2009; Fricano et al., 2012; Sierro et al., 2014). The most widely used intraspecific classification of cultivated tobacco today is based on differences in plant morphology, as well as the quantitative and qualitative composition of their secondary metabolites. These indicators determine the key characteristics of tobacco raw materials, with this classification referred to as "market" (Lewis, Nicholson, 2007). There are eight market classes: Burley, cigar filler, cigar roll tobacco, dark air-cured tobacco, dark steam-cured tobacco, flue-cured tobacco, Maryland, and Oriental tobacco (Moon et al., 2009). Despite the high level of phenotypic variability among cultivated tobacco varieties (Lewis, Nicholson, 2007), the level of nucleotide variability revealed by restric-

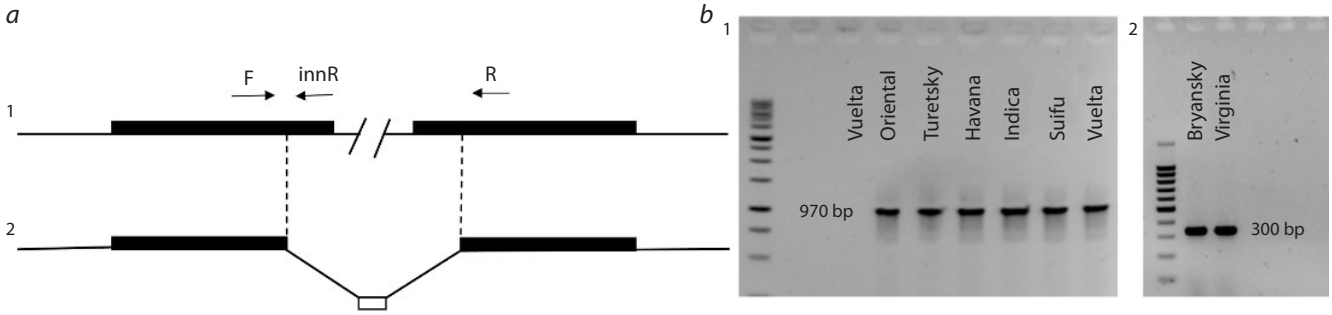


Fig. 2. TA cT-DNA in the genomes of *N. tabacum.*

a – the schematic representation of TA is shown, with the central part omitted. The *orf13* gene is indicated by black rectangles, and the deletion boundaries are marked with dotted lines. The primers selected in the work of G.V. Khafizova and T.V. Matveeva (2020) are also indicated; 1 – TA cT-DNA without deletion; 2 – TA cT-DNA with deletion, where the white rectangle marks a sequence of unknown origin that is 42 bp long; *b* – fragments obtained using primers F and innR (1), and F and R (2) in PCR analysis of eight *N. tabacum* varieties (according to (Khafizova, Matveeva, 2020)).

tion fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) markers, and genome-wide association studies (GWAS) is relatively low (Brandle, Bai, 1999; Ren, Timko, 2001; Rossi et al., 2001; Tong et al., 2020).

To conduct a phylogenetic analysis and construct a map that accurately reflects the relationships within the species, additional molecular markers are necessary. One such marker is cT-DNA. A study of cT-DNA in the whole-genome data of three *N. tabacum* varieties revealed an extended deletion in the central part of cT-DNA TA in the Basma/Xanti variety. In contrast, a TA sequence without a deletion was found in the K326 and TN90 varieties (Chen et al., 2014). Differences in the TA structure were previously demonstrated by PCR for the *N. tabacum* varieties Basma Drama 2, Samsoun, and Xanthi, the cT-DNA of which contains incomplete sequences of the *orf13* gene homologue, unlike the cT-DNA in the Wisconsin 38 and Havana 425 varieties, as well as in *N. tomentosiformis* (Mohajjel-Shoja et al., 2011). Later, eight more varieties of cultivated tobacco were analyzed using the PCR method: Vuelta Abajo, Suifu, Black Indian, Havana 307, Turetsky, Oriental, Bryansky 91, and Virginia × Burley 38 (Fig. 2). The analysis revealed the previously described deletion in the Bryansky 91 and Virginia × Burley 38 varieties, with the deletion localization site coinciding with those in the genomes of the Basma Drama 2, Samsoun, Xanthi, and Basma/Xanti varieties with nucleotide precision (Khafizova, Matveeva, 2020).

The use of the structural variant of cT-DNA as a molecular marker has allowed us to group the varieties belonging to the market class of oriental tobaccos based on the deletion in TA. It is hypothesized that the central part of TA was lost in the ancestral form of *N. tabacum*, which gave rise to this class. To date, the deletion in TA has been described in only five varieties of cultivated tobacco. Given the limited sample size, it is premature to conclude whether the presence of this deletion serves as a definitive marker for classifying a variety as oriental tobacco. Nevertheless, the results indicate that the structural polymorphism of cT-DNA in *N. tabacum* varieties can be utilized as one of the molecular markers for studying intraspecific relationships among varieties. As the list of sequenced ctDNA from different cultivated tobacco varieties expands, new structural differences in cT-DNA sequences may be discovered, potentially leading to the development of additional markers for this purpose.

Fine polymorphism of cT-DNA in phylogenetic studies

Fine polymorphism of cT-DNA with a common origin in the genome (sharing a common localization site) can be utilized to study interspecific variability and reconstruct phylogenetic relationships within a monophyletic group of descendant species from an ancient transformant. For a correct assessment of the completeness of the speciation process in plants, it is essential to evaluate and compare both intra- and interspecific variability for the studied markers. In this context, it has been proposed to reconstruct individual alleles of the studied markers, particularly for cross-pollinated species (Chen et al., 2022). First, let us focus on the methodology of allele separation.

Research methods

In scientific literature, allele phasing or haplotype phasing refers to obtaining sequences of DNA fragments located on one chromosome of a pair of homologous chromosomes or, in the case of polyploids, among homeologous chromosomes. Typically, "alleles" refer to shorter fragments, while "haplotypes" pertain to longer sections.

Haplotype phasing provides additional information compared to that derived solely from the consensus genome sequence. Exact haplotype sequences are valuable in a variety of studies, including phylogenetic reconstruction (Tiley et al., 2024) and hybrid studies (Sun et al., 2020).

Various approaches to haplotype phasing have been described (Snyder et al., 2015). For example, one can physically separate fragments of homologous chromosomes using molecular biology methods followed by separate sequencing of the fragments. Bacterial cloning is one such method. Another approach involves separating haplotypes based on genotyping data regarding nucleotide frequencies at polymorphic positions within a population, utilizing various statistical methods (Browning S.R., Browning B.L., 2011). However,

the most rapidly advancing approach relies on high-throughput sequencing data. Within this framework, methods can be categorized based on either the assembly of short reads or the mapping of short reads to a reference genome (Zhang et al., 2020). Let us examine the latter case in more detail.

If two polymorphic positions are located within the same sequenced DNA fragment (read from one or both short reads), they belong to the same haplotype. The haplotype sequence can be reconstructed as long as fragments connecting adjacent polymorphic positions are identified. Consequently, the higher the density of polymorphic positions, the greater the overlap between short reads, and the more increased the read depth, the more reliably and extensively haplotype sequences can be reconstructed. These parameters are particularly important when separating polyploid genomes, where more than two haplotypes must be distinguished, some of which may exhibit reduced variance over relatively long stretches (Schrinner et al., 2020).

The indicators mentioned above depend on the sequencing technology used. The best results can be achieved with Hi-Fi technology, which produces long reads of several tens of kilobase pairs (kb) in high quality, with a reading accuracy exceeding 99 % (Wenger et al., 2019). Hi-Fi sequencing enables the assembly of extended haplotype sections, potentially encompassing the entire genome (Tanaka et al., 2023). In contrast, Oxford Nanopore technology can generate very long reads, reaching several million base pairs (bp) in length, but with lower read quality – around 90 % accuracy (Wang Y. et al., 2021). With sufficient sequencing depth, these long reads can also effectively separate extended sections of the genome into haplotypes. Additionally, short reads obtained using Illumina technology can aid in haplotype separation; however, the resulting fragments are limited to conservative regions where read overlaps are insufficient to connect adjacent polymorphic positions.

A popular program for haplotype separation is WhatsHap (Martin et al., 2016). It is compatible with reads from all the aforementioned sequencing methods for both diploid and polyploid organisms. The program requires as input a reference genome sequence, a BAM file mapping reads to the reference genome, and a VCF file containing information on polymorphic positions that distinguish the mapped reads from the reference. The separation process yields a modified VCF file that includes information on the assignment of polymorphic positions to haplotypes. From this file, users can obtain data on the lengths and coordinates of the separated fragments (haplotype blocks), retrieve haplotype sequences in FASTA format, incorporate haplotype information into read mapping visualizations, and more.

Alleles can also be separated through the analysis of Sanger sequencing results. When sequencing DNA fragments from heterozygotes, two "peaks" corresponding to specific nucleotides appear on the chromatograms at the same position (Carr et al., 2009; Dehairs et al., 2016; Xie et al., 2019). To deduce the allele sequences of the gene under study, the sequences from each sample can be represented as a vector, where each cell corresponds to a polymorphic position in the gene for the species being examined. Each cell is filled according to the following rules: "11" if the sample contains the most common nucleotide at that position, "00" if it contains the least common nucleotide, and "10" if two different nucleotides occur at that position in the organism's genome (indicating a putative heterozygote). The resulting vectors can then be assigned to all possible combinations of heterozygous positions "10" while leaving the single-valued positions "11" and "00" untouched. Since there are homozygotes and samples with one allele among the samples, their sequences form a primary pool of alleles that can later be detected in the remaining samples. Each of these alleles in the diploid must correspond to an allele with alternative values in the polymorphic positions, allowing for the identification of a homologous pair for the primary allele (Zhidkin et al., 2023). By sorting through potential combinations of alleles and selecting those with a higher frequency of occurrence, it is possible to determine the genotypes of the samples.

The approaches described above were applied to study the intra- and interspecific variability of plants in the genera *Camellia*, *Vaccinium*, and *Arachis*. Let us explore these examples in more detail.

cT-DNA polymorphism in *Camellia* L. species

The genus *Camellia* L. belongs to the Theaceae family and includes several economically and culturally significant species. *C. sinensis* (L.) Kuntze comprises two main varieties: var. *sinensis* (primarily used for green tea production) and var. *assamica* (mainly used for black tea production). In the study by K. Chen et al. (2023), 72 species from 12 out of the 14 sections of the genus *Camellia* were analyzed, revealing at least 12 different cT-DNA insertions. These sequences span a total of 374 kb and contain 47 open reading frames. The identified genes can be categorized into four types: the first includes 19 *plast* genes, the second contains 6 opine synthesis genes, the third comprises 4 genes encoding tryptophan monooxygenase, and the fourth consists of all other genes with unknown functions. The protein sequences of the cT-DNA genes exhibit varying levels of similarity to known *Agrobacterium* sequences, with an average similarity of 73.8 % (standard deviation is 12.8 %). The minimum and maximum similarity values are 46 and 92 %, respectively.

Notably, some genes homologous to those found in camellias were also identified in various fungal species, including both ascomycetes and basidiomycetes. Internal inverted repeats are present in 7 of the 12 insertions, likely arising from the simultaneous insertion of multiple copies of T-DNA. The differences between these repeats range from 0.05 to 10 % across different fragments. Considering that the repeats were identical at the time of insertion and using the universal substitution rate of 6.5×10^{-9} per position per year (Gaut et al., 1996), along with the number of substitutions between repeats, it is possible to estimate the approximate time of fragment insertion (Haubold, Wiehe, 2001). This time ranges from 0.04 to 7.5 million years ago (Mya) (Chen et al., 2023).

As different phylogenetic lines of the genus evolved, the fragments were inserted at various stages, making them specific to modern taxa at different levels. It was demonstrated that some fragments were lost in certain lines, and the youngest fragments had not yet fully fixed in the populations of their respective species, being present only in some individuals.

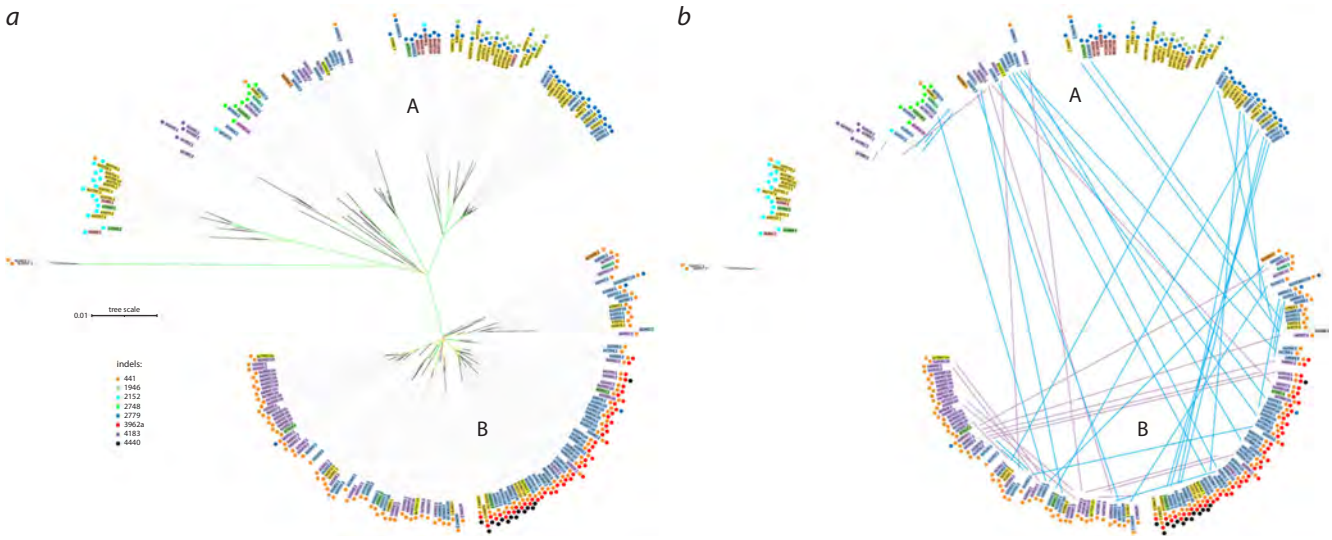


Fig. 3. Phylogenetic analysis of the section *Thea* using cT-DNA. *a* – phylogenetic tree of 225 clades representing nine species from the genus *Camellia*, section *Thea*; *b* – alleles identified in a single *C. sinensis* sample are connected by lines.

Two main clades are labeled with the Latin letters A and B in both figures. The color designations for the species are as follows: *C. sinensis* (including *C. sinensis* var. *sinensis, C. pubilimba, C. angustifolia*) – blue; *C. sinensis* var. *assamica* – purple; *C. tachangensis* – dark yellow; *C. crassicolumna* – green; *C. gymnogyna* – yellow; *C. taliensis* – pink; *C. leptophylla* – dark green; *C. kwangsiensis* – orange; *C. ptilophylla* – red; *C. fangchengensis* – no color. Indels are indicated by colored dots. Bootstrap values are color-coded: green – 85–100 %; yellow – 50–85 %; red – less than 20 %. The figure is based on illustrations from (Chen et al., 2022), published under a CC BY 4.0 license.

With known cT-DNA sequences, their presence or absence across different phylogenetic lines of the studied taxon is a convenient feature that can be easily assessed. Thus, the sequence of divergence among sections within a genus, based on the presence of different cT-DNAs, aligns with the phylogenetic tree derived from whole-exome sequencing (Wu et al., 2022). Variations among repeats can also help clarify both the relative and absolute timing of divergence.

In the *Thea* section, only one cT-DNA variety has been described, which was used to elucidate the phylogenetic relationships among the species in this section. This section comprises eleven species (Min, Bartholomew, 2007), including the tea bush *C. sinensis* (L.) O. Kuntze and various wild tea species. *C. sinensis* is further divided into the varieties *sinensis*, *assamica*, *pubilimba*, and *dehungensis*. A phylogeny was constructed for the nine species of the *Thea* section of the genus *Camellia* based on one of the aforementioned insertions (Chen et al., 2022). This insertion is 5.5 kb long, includes three genes: *acs*-like, *sus*-like, and a *rolB*-like fragment, and features inverted repeats of 1 kb each at the ends. Sequence analysis revealed that the *acs*-like and *sus*-like genes contain stop codons; however, it was noted that in some species, these stop codons arose independently at different positions. The estimated time of the fragment insertion is 7.5 Mya, which roughly coincides with the origin of section *Thea* at 6.7 Mya (Wu et al., 2022). Therefore, this insertion serves as a suitable marker for this group, which we anticipate to be free of potential biases associated with incomplete lineage sorting.

During the assembly of fragments from short reads, it was observed that many samples contained two alleles, with 2 to 7 % of positions exhibiting polymorphism. Following the analysis of 142 samples, 225 alleles were obtained, forming the basis for constructing the phylogenetic tree (Fig. 3). It was noted that the clades obtained often did not align with species boundaries, and many heterozygous samples contained alleles from different, distantly related clades. Particularly notable in some samples was the presence of alleles from clades separated by a significant evolutionary distance. The maximum divergence observed between alleles from different clades within a single sample reached about 4 %. Based on the premise that such allelic divergence would be unlikely within a single species population, it was suggested that alleles from the major clades evolved within individual ancestral species of the *Thea* section. This process likely led to the formation of modern species characterized by high allelic diversity due to introgressive crosses. Furthermore, if we assume that the number of major clades corresponds to the number of ancestral species, we can estimate the relationships among these species based on tree topology, divergence time derived from the number of substitutions between alleles, and the contributions of allele frequency to modern populations. For instance, two major pairs of phylogenetically distant clades can be identified, between which hybridization occurred. So, as a result of certain crosses, the species *C. sinensis* emerged, with one of the presumed ancestral species exhibiting twice as many alleles as the other. The 2.5 % divergence between alleles suggests that these species diverged approximately 1.9 Mya. In contrast, within the second pair, which diverged 2.9 Mya, the species *C. tachangensis* arose as a result of equally participatory crosses.

This work raises important questions regarding the revision of species boundaries, the intensity of interspecific crosses, and other aspects of the evolutionary dynamics within this section.

Polymorphism of cT-DNA in species of the genus *Vaccinium*

The first naturally transgenic species from the genus *Vaccinium* L. described in the literature was the large-fruited cranberry, *V. macrocarpon* Aiton (Matveeva, Otten, 2019). Using the BLAST algorithm, a cT-DNA sequence was identified in the complete genome of the cranberry, represented by a single, intact copy of the *rolB/C*-like gene. This gene belongs to the *plast* genes, and the amino acid sequence corresponding to it exhibits greater homology with the Plast-proteins of organisms such as *Laccaria bicolor* (Matre) Orton, *Nyssa sinensis* Oliver, and *Ensifer* sp., than with similar sequences from the genomes of known species of agrobacteria. This may suggest that these sequences originated from different transformation events involving the same unknown species of *Agrobacterium* (Matveeva, Otten, 2019). The second *Vaccinium* species found to contain a homologue of the studied gene was the highbush blueberry, *V. corymbosum* (Matveeva, 2021). Subsequently, we demonstrated the presence of a *rolB/C*-like gene in 26 additional *Vaccinium* species, as well as in *Agapetes serpens* (Wight) Sleumer, which belongs to the same family, Ericaceae (Zhidkin et al., 2023). The common localization site of the detected sequences in both *Vaccinium* and *A. serpens* is also noteworthy. This wide distribution of the transgene throughout the genus and the shared localization site suggest that the transformation occurred in a common ancestor of the studied species. Consequently, the *rolB/C*-like gene sequence can be utilized as a marker for reconstructing phylogenetic relationships among these species. This task is particularly relevant, as the phylogeny of the genus *Vaccinium* remains contentious (Becker et al., 2023).

In all studied species, the *rolB/C*-like gene is represented by a full-length sequence. The exception is the common cranberry, *V. oxycoccos* L., where most samples exhibited large deletions of varying lengths in the central part of the gene; however, some samples contained a full-length sequence of the *rolB/C*-like gene. In the remaining species, polymorphism of the *rolB/C*-like gene was characterized by single-nucleotide substitutions and indels that are multiples of three, preserving the open reading frame. Furthermore, the pattern of these nucleotide differences was species-specific.

Despite the long-standing use of cranberries, blueberries, bilberries, and lingonberries by humans for food and medicinal purposes, the selection of these crops began in the early 20th century (Wang H. et al., 2017; Vorsa, Zalapa, 2019; Sultana et al., 2020). This selection work led to the development of the genus system, the first version of which was established in 1945 (Camp, 1945). The classical system of the genus was based on various morphological features and divided it into sections. Over time, this system was regularly updated and modified, as interspecific hybridization and polyploidization are common in the genus (Camp, Gilly, 1943; Hancock, 2008). These characteristics made it difficult to clearly determine the phylogenetic relationships between species, prompting the development of molecular phylogenetics and the application of DNA barcoding methods to address these issues (Kron, 2002; Powell, Kron, 2003). The dendrogram obtained from the ITS (internal transcribed spacer) and *matK* (plastid gene of maturase K) sequences contradicted classical ideas about the division of the genus into sections and indicated the polyphyly of the genus *Vaccinium*. However, using classical phylogenetic markers in cladistic analysis for species where hybridization and polyploidization play significant roles can lead to errors (Soltis, 2002). In contrast, data obtained using SSR markers (simple sequence repeats — microsatellite DNA) (Zalapa et al., 2015; Schlautman et al., 2017), phylogenomics (Diaz-Garcia et al., 2019; Kawash et al., 2022), and chemotaxonomy (Leisner et al., 2017) showed fewer contradictions with classical concepts. Genome sequencing is labor-intensive and expensive, so it has been conducted only on economically significant species. The *rolB/C*-like gene, as a phylogenetic marker, allowed for the inclusion of more species in the analysis.

In the studied species, the intraspecific variability of the transgene (unlike in species from the *Camellia* section *Thea*) was lower than the interspecific variability, and the mosaicism of some clades could be attributed to hybridization events among the species within those clades. Phylogenetic analysis revealed the unification of representatives from the sections *Oxycoccus*, *Vaccinium*, *Myrtillus*, and *Conchophyllum* into distinct clades. In contrast, species from the section *Cyanococcus* did not form a monophyletic group, possibly due to its polyphyletic nature or hybridization events during the development of North American blueberry varieties. The remaining species studied are single representatives of the sections *Bracteata*, *Hemimyrtillus*, *Vitis-idaea*, *Oxycoccoides*, and *Praestantia*; therefore, further research is needed. Given the simplicity and low cost of the developed molecular marker, new species can easily be included in subsequent analyses. In other words, the phylogeny of the genus *Vaccinium*, determined from the sequences of the *rolB/C*-like gene, shows greater similarity to the traditional classification of the genus than to the phylogeny constructed based on the ITS and *matK* markers. Results similar to those obtained using the *rolB/C*-like gene marker have also been reported by other authors employing NGS (next generation sequencing) approaches (Diaz-Garcia et al., 2019; Kawash et al., 2022).

Polymorphism of cT-DNA in species of the genus *Arachis*

The genus *Arachis* L. comprises 80 species and is divided into nine taxonomic sections: *Arachis* (with genomes A, B, K), *Erectoides* (genome E), *Extranervosae* (genome EX), *Procumbentes* (genome PR), *Caulorrhizae* (genome C), *Heteranthae* (genome H), *Rhizomatosae* (genome R), *Trierectoides* (genome TE), and *Triseminatae* (genome T) (Stalker et al., 2017). Current understanding of the evolutionary relationships among representatives of the genus *Arachis* relies on morphological, geographical, molecular genetic, and cytogenetic data; however, many controversial issues regarding the structure of the genus remain (Krapovickas, Gregory, 2007; Koppolu et al., 2010; Stalker, 2017; Tian et al., 2021).

Initially, nGMOs were discovered within the *Arachis* section in the tetraploid cultivated peanut (*Arachis hypogaea* L.) and its ancestors: A. *duranensis* Krapov. & W.C. Greg, *A. ipaensis* Krapov. & W.C. Greg, the tetraploid species *A. monticola* Krapov. & Rigoni (Matveeva, Otten, 2019), and the diploid *A. stenosperma* Krapov. & W.C. Greg (Matveeva, Otten, 2021). The peanut and *A. monticola* contain ancestral genomes known as A and B. The A genome is also present in

A. duranensis and *A. stenosperma*, while *A. ipaensis* contains the B genome (Matveeva, Otten, 2019, 2021). The list of nGMOs was later expanded to include representatives of the *Erectoides*, *Extranervosae*, *Procumbentes*, *Caulorrhizae*, and *Heteranthae* sections (Bogomaz et al., 2024).

In total, 23 naturally transgenic species from this genus are currently known, and there is a high probability that this list will expand in the near future, as the studied nGMOs form a monophyletic group with a common ancestor that was transformed before the studied sections diverged. A homolog of the cucumopine synthase gene (*cus*) was found in all studied species. In addition to the *cus*-like gene, B-genome species contained remnants of the *mas2′* gene, PR-genome species contained remnants of the *mas1′* gene, *A. macedoi* contained remnants of the *ags* gene, and *A. pusilla* contained remnants of both the *mas2′* and *ags* genes (Bogomaz et al., 2024). All of these genes encode enzymes belonging to the same biosynthetic pathway, catalyzing the reactions that lead to the synthesis of agropin, and are found clustered together in the same *Agrobacterium* T-DNA (Ellis et al., 1984). The common ancestor of *Arachis* species was likely transformed by a strain containing all three genes; however, at some point, these genes ceased to provide selective advantages to their hosts, accumulated mutations, and were lost, remaining as separate fragments in representatives of different clades.

In contrast, the homologue of the cucumopine synthase gene has remained intact in most of the studied species. In cultivated peanut, which is a tetraploid with genomes A and B, the *cus*-like gene is present in both genomes; in genome A, it is intact, while in genome B, it is mutant (Matveeva, Otten, 2019). More detailed studies showed that among the 29 described alleles of the gene from genome A, only three contained mutations incompatible with its function (Bogomaz et al., 2024). Cultivated peanut is divided into two subspecies: *hypogaea* and *fastigiata* (Krapovickas, Gregory, 2007; Bertioli et al., 2011). The most common allele A of the *cus*-like gene has been identified in representatives of both subspecies, as well as in *A. duranensis*, confirming their close relationship. However, this allele has not been found in *A. monticola*, where its other alleles are evenly distributed among separate subclades within the clade containing the A alleles of the cultivated peanut genome and its relatives (Bogomaz et al., 2024). This finding supports previous descriptions of the close relationship between *A. monticola* and *A. hypogaea* (Tian et al., 2021). Based on the data obtained using the *cus*-like gene as a phylogenetic marker, it is possible that the genetic material of *A. paraguariensis* contributed to the formation of some varieties of *A. hypogaea*. Additional research confirms a close relationship between *A. paraguariensis* and *A. duranensis*, the ancestral species of cultivated peanut (Moretzsohn et al., 2013).

Meanwhile, phylogenetic studies of peanuts based on the *cus*-like gene indicate that the most distinct clade on the phylogenetic tree is represented by mutant alleles from the B genome. These sequences exhibit faster divergence and are more suitable for research than those subjected to stabilizing selection.

Consequently, studies of the genus *Arachis* illustrate some limitations in the use of cT-DNA for phylogenetic analysis.

Phylogenetic relationships of T-DNA genes in *Agrobacterium* s. lat and nGMOs

In the previous sections, we explored the use of T-DNA as a molecular marker for studying plant phylogeny. This marker can also be utilized to trace the relationships between nGMO cT-DNA and the T-DNA of currently known strains of *Agrobacterium* s. lat (Suzuki et al., 2002; Matveeva, Otten, 2021). In one of our group's studies (Matveeva, Otten, 2021), phylogenetic trees were constructed based on individual opine synthesis genes from all known nGMOs as of 2021, as well as from rhizobia strains characterized at that time. The results indicated that in *Parasponia andersonii* Planch., cT-DNAs containing homologues of the *susL* gene were obtained from various strains of *Agrobacterium* s. lat across different evolutionary stages. A similar pattern was observed in *Diospyros lotus* L. Conversely, cT-DNAs containing homologues of mikimopine synthase were most likely acquired by different species of tobacco (*Nicotiana* L.) and toadflax (*Linaria* Mill.) from a single or closely related strain (Matveeva, Otten, 2021). These findings enhance our understanding of the biodiversity of *Agrobacterium* s. lat and bring us closer to elucidating the mechanisms of host specificity, which is often linked to the structure and functioning of *vir* genes inherited with T-DNA as part of Ti(Ri) plasmids (Anderson, Moore, 1979). Addressing host specificity is crucial for optimizing plant genetic transformation protocols.

Conclusion

cT-DNA plays a crucial role in elucidating many controversial aspects of phylogenetic studies. The insertions of various T-DNAs mark significant evolutionary events, indicating groups of species that share a common ancestor. Inverted repeats provide insights into the age of this ancestor and help establish the sequence in which independent T-DNAs entered plant genomes. Analyzing the fine polymorphism of cT-DNA, while considering the allelic states of the markers, allows for tracking microevolutionary events and the consequences of hybridization during incomplete speciation. To date, molecular markers based on cT-DNA have been successfully employed in studying the genera *Nicotiana*, *Camellia*, *Vaccinium*, and *Arachis*. In the genus *Nicotiana*, the cT-DNA marker facilitated the identification and dating of major evolutionary stages within the section *Tomentisae*. In *Camellia*, the primary outcome was a clear demonstration of incomplete speciation within the section *Thea*. For *Vaccinium*, the marker helped confirm some classical ideas about the genus's system that conflicted with ITS-based data but aligned with NGS data from a small sample of species. In *Arachis*, the study of cT-DNA clearly illustrated the differing evolutionary fates of transgenes with and without stabilizing selection, highlighting some limitations in the marker's resolving power under strong stabilizing selection pressure.

References

Anderson A., Moore L. Host specificity in the genus *Agrobacterium*. *Phytopathology*. 1979;69(4):320-323. doi 10.1094/Phyto-69-320

Bahramnejad B., Naji M., Bose R., Jha S. A critical review on use of *Agrobacterium rhizogenes* and their associated binary vectors for plant transformation. *Biotechnol Adv.* 2019;37(7):107405. doi 10.1016/j.biotechadv.2019.06.004

Becker A.L., Crowl A.A., Luteyn J.L., Chanderbali A.S., Judd W.S., Manos P.S., Soltis D.E., Smith S.A., Goncalves D.J.P., Dick C.W., Weaver W.N., Soltis P.S., Cellinese N., Fritsch P.W. A global blueberry phylogeny: evolution, diversification, and biogeography of Vaccinieae (Ericaceae). *Mol Phylogenet Evol*. 2024;201:108202. doi 10.1016/j.ympev.2024.108202

Bertioli D.J., Seijo G., Freitas F.O., Valls J.F.M., Leal-Bertioli S.C.M., Moretzsohn M.C. An overview of peanut and its wild relatives. *Plant Genet Resour.* 2011;9(1):134-149. doi 10.1017/S1479262110000444

Bogomaz O.D., Bemova V.D., Mirgorodskii N.A., Matveeva T.V. Evolutionary fate of the opine synthesis genes in the *Arachis* L. genomes. *Biology* (*Basel*). 2024;13(8):601. doi 10.3390/biology13080601

Brandle J., Bai D. Biotechnology: uses and applications in tobacco improvement. In: Davis D.L., Nielsen M.T. (Eds) Tobacco: Production, Chemistry and Technology. Wiley, 1999;49-65

Browning S.R., Browning B.L. Haplotype phasing: existing methods and new developments. *Nat Rev Genet*. 2011;12(10):703-714. doi 10.1038/nrg3054

Camp W.H. The North American blueberries with notes on other groups of Vacciniaceae. *Brittonia*. 1945;5(3):203-275. doi 10.2307/2804880

Camp W.H., Gilly C.L. The structure and origin of species. *Brittonia*. 1943;4(3):323-385. doi 10.2307/2804896

Carr I.M., Robinson J.I., Dimitriou R., Markham A.F., Morgan A.W., Bonthron D.T. Inferring relative proportions of DNA variants from sequencing electropherograms. *Bioinformatics*. 2009;25(24):3244-3250. doi 10.1093/bioinformatics/btp583

Chen K., Dorlhac de Borne F., Szegedi E., Otten L. Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus *Nicotiana*. *Plant J*. 2014;80(4):669-682. doi 10.1111/tpj.12661

Chen K., Dorlhac de Borne F., Sierro N., Ivanov N.V., Alouia M., Koechler S., Otten L. Organization of the TC and TE cellular T-DNA regions in *Nicotiana otophora* and functional analysis of three diverged TE-*6b* genes. *Plant J*. 2018;94(2):274-287. doi 10.1111/tpj.13853

Chen K., Zhurbenko P., Danilov L., Matveeva T., Otten L. Conservation of an *Agrobacterium* cT-DNA insert in *Camellia* section *Thea* reveals the ancient origin of tea plants from a genetically modified ancestor. *Front Plant Sci*. 2022;13:997762. doi 10.3389/fpls.2022.997762

Chen K., Liu H., Blevins T., Hao J., Otten L. Extensive natural *Agrobacterium*-induced transformation in the genus *Camellia*. *Planta*. 2023;258(4):81. doi 10.1007/s00425-023-04234-9

Christey M.C. Use of Ri-mediated transformation for production of transgenic plants. *In Vitro Cell Dev Biol Plant*. 2001;37:687-700. doi 10.1007/s11627-001-0120-0

Dehairs J., Talebi A., Cherifi Y., Swinnen J.V. CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. *Sci Rep*. 2016;6:28973. doi 10.1038/srep28973

Diaz-Garcia L., Rodriguez-Bonilla L., Rohde J., Smith T., Zalapa J. Pacbio sequencing reveals identical organelle genomes between American cranberry (*Vaccinium macrocarpon* Ait.) and a wild relative. *Genes* (*Basel*). 2019;10(4):291. doi 10.3390/genes10040291

Doronina L., Feigin C.Y., Schmitz J. Reunion of Australasian possums by shared SINE insertions. *Syst Biol*. 2022;71(5):1045-1053. doi 10.1093/sysbio/syac025

Ellis J.G., Ryder M.M., Tate M.E. *Agrobacterium tumefaciens* T_R-DNA encodes a pathway for agropine biosynthesis. *Mol Gen Genet.* 1984;195:466-473. doi 10.1007/BF00341448

Fricano A., Bakaher N., Del Corvo M., Piffanelli P., Donini P., Stella A., Ivanov N.V., Pozzi C. Molecular diversity, population structure, and linkage disequilibrium in a worldwide collection of tobacco (*Nicotiana tabacum* L.) germplasm. *BMC Genet*. 2012;13:18. doi 10.1186/1471-2156-13-18

Furner I.J., Huffman G.A., Amasino R.M. An *Agrobacterium* transformation in the evolution of the genus *Nicotiana*. *Nature.* 1986;319:422-427. doi 10.1038/319422a0

Gaut B.S., Mortont B.R., McCaig B.C., Clegg M.T. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL* (orthology/nucleotide substitution rates/molecular clock). *Proc Natl Acad Sci USA.* 1996;93(19):10274-10279. doi 10.1073/pnas.93.19.10274

Gelvin S.B. Plant DNA repair and *Agrobacterium* T-DNA integration. *Int J Mol Sci.* 2021;22(16):8458. doi 10.3390/ijms22168458

Hancock J.F., Lyrene P., Finn C.E., Vorsa N., Lobos G.A. Blueberries and cranberries. In: Hancock J.F. (Ed.) Temperate Fruit Crop Breeding. Springer, 2008;115-150. doi 10.1007/978-1-4020-6907-9_4

Hansen G., Larribe M., Vaubert D., Tempe J., Biermann B.J., Montoya A.L., Chilton M.-D., Brevet J. *Agrobacterium rhizogenes* pRi8196 T-DNA: mapping and DNA sequence of functions involved in mannopine synthesis and hairy root differentiation. *Proc Natl Acad Sci USA.* 1991;88(17):7763-7767. doi 10.1073/pnas.88.17.7763

Haubold B., Wiehe T. Statistics of divergence times. *Mol Biol Evol*. 2001;18(7):1157-1160. doi 10.1093/oxfordjournals.molbev.a003902

Intrieri M.C., Buiatti M. The horizontal transfer of *Agrobacterium rhizogenes* genes and the evolution of the genus *Nicotiana*. *Mol Phylogenet Evol*. 2001;20(1):100-110. doi 10.1006/mpev.2001.0927

Jouanin L. Restriction map of an agropine-type Ri plasmid and its homologies with Ti plasmids. *Plasmid*. 1984;12(2):91-102. doi 10.1016/0147-619x(84)90055-6

Kawash J., Colt K., Hartwick N.T., Abramson B.W., Vorsa N., Polashock J.J., Michael T.P. Contrasting a reference cranberry genome to a crop wild relative provides insights into adaptation, domestication, and breeding. *PLoS One.* 2022;17(3):e0264966. doi 10.1371/journal.pone.0264966

Khafizova G.V., Matveeva T.V. Polymorphism in sequences of agrobacterial origin in *Nicotiana tabacum* cultivars. *Russ J Genet*. 2020;56(10):1215-1217. doi 10.1134/S1022795420100051

Khafizova G.V., Sierro N., Ivanov N.V., Sokornova S.V., Polev D.E., Matveeva T.V. *Nicotiana noctiflora* Hook. genome contains two cellular T-DNAs with functional genes. *Plants* (*Basel*). 2023;12(22):3787. doi 10.3390/plants12223787

Knapp S., Chase M.W., Clarkso J.J. Nomenclature changes and a new sectional classification in *Nicotiana* (Solanaceae). *Taxon*. 2004;53(1):73-82. doi 10.2307/4135490

Koppolu R., Upadhyaya H.D., Dwivedi S.L., Hoisington D.A., Varshney R.K. Genetic relationships among seven sections of genus *Arachis* studied by using SSR markers. *BMC Plant Biol.* 2010;10:15. doi 10.1186/1471-2229-10-15

Krapovickas A., Gregory W.C. Taxonomy of the genus *Arachis* (Leguminosae). *Bonplandia*. 2007;16(Supl.):1-205. doi 10.30972/bon.160158

Kron K., Powell E., Luteyn J. Phylogenetic relationships within the blueberry tribe (Vaccinieae, Ericaceae) based on sequence data from MATK and nuclear ribosomal ITS regions, with comments on the placement of *Satyria*. *Am J Bot.* 2002;89(2):327-336. doi 10.3732/ajb.89.2.327

Kuritzin A., Kischka T., Schmitz J., Churakov G. Incomplete lineage sorting and hybridization statistics for large-scale retroposon insertion data. *PLoS Comput Biol*. 2016;12(3):e1004812. doi 10.1371/journal.pcbi.1004812

Leisner C.P., Kamileen M.O., Conway M.E., O'Connor S.E., Buell C.R. Differential iridoid production as revealed by a diversity panel of 84 cultivated and wild blueberry species. *PLoS One.* 2017;12(6):e0179417. doi 10.1371/journal.pone.0179417

Leitch I.J., Hanson L., Lim K.Y., Kovarik A., Chase M.W., Clarkson J.J., Leitch A.R. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann Bot*. 2008;101(6):805-814. doi 10.1093/aob/mcm326

Lewis R.S., Nicholson J.S. Aspects of the evolution of *Nicotiana tabacum* L. and the status of the United States *Nicotiana* germplasm collection. *Genet Resour Crop Evol*. 2007;54:727-740. doi 10.1007/s10722-006-0024-2

Lynch M., Conery J.S. The evolutionary fate and consequences of duplicate genes. *Science*. 2000;290(5494):1151-1155. doi 10.1126/science.290.5494.1151

Martin M., Patterson M., Garg S., Fischer S., Pisanti N., Klau G.W., Schöenhuth A., Marschall T. WhatsHap: fast and accurate read-based phasing. *bioRxiv*. 2016. doi 10.1101/085050

Matveeva T. New naturally transgenic plants: 2020 update. *Bio Comm*. 2021;66(1):36-46. doi 10.21638/SPBU03.2021.105

Matveeva T.V. Why do plants need agrobacterial genes? *Ecol Genet*. 2021;19(4):365-375. doi 10.17816/ecogen89905

Matveeva T.V., Otten L. Widespread occurrence of natural genetic transformation of plants by *Agrobacterium*. *Plant Mol Biol*. 2019;101(4-5):415-437. doi 10.1007/s11103-019-00913-y

Matveeva T.V., Otten L. Opine biosynthesis in naturally transgenic plants: genes and products. *Phytochemistry*. 2021;189:112813. doi 10.1016/j.phytochem.2021.112813

Matveeva T.V., Pavlova O.A., Bogomaz D.I., Demkovich A.E., Lutova L.A. Molecular markers for plant species identification and phylogenetics. *Ecol Genet*. 2011;9(1):32-43. doi 10.17816/ecogen9132-43 (in Russian)

Meyer A.D., Ichikawa T., Meins F. Horizontal gene transfer: regulated expression of a tobacco homologue of the *Agrobacterium rhizogenes rolC* gene. *Mol Gen Genet*. 1995;249(3):265-273. doi 10.1007/BF00290526

Min T.L., Bartholomew B. Theaceae. In: Wu Z.Y., Raven P.H. (Eds) Flora of China. Vol. 12. St Louis: Science Press, Beijing & Missouri Botanical Garden Press, 2007;323-450

Mohajjel-Shoja H., Clément B., Perot J., Alioua M., Otten L. Biological activity of the *Agrobacterium rhizogenes*-derived *trolC* gene of *Nicotiana tabacum* and its functional relation to other *plast* genes. *Mol Plant Microbe Interact*. 2011;24(1):44-53. doi 10.1094/MPMI-06-10-0139

Moon H.S., Nifong J.M., Nicholson J.S., Heineman A., Lion K., Van der Hoeven R., Lewis R.S. Microsatellite-based analysis of tobacco (*Nicotiana tabacum* L.) genetic resources. *Crop Sci*. 2009;49(6):2149-2159. doi 10.2135/cropsci2009.01.0024

Moretzsohn M.C., Gouvea E.G., Inglis P.W., Leal-Bertioli S.C.M., Valls J.F.M., Bertioli D.J. A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Ann Bot*. 2013;111(1):113-126. doi 10.1093/aob/mcs237

Nester E.W. *Agrobacterium*: nature's genetic engineer. *Front Plant Sci*. 2015;5:730. doi 10.3389/fpls.2014.00730

Otten L. Natural *Agrobacterium*-mediated transformation in the genus *Nicotiana*. In: Ivanov N.V., Sierro N., Peitsch M.C. (Eds) The Tobacco Plant Genome. Compendium of Plant Genomes. Cham: Springer, 2020;195-209. doi 10.1007/978-3-030-29493-9_12

Powell E., Kron K. Molecular systematics of the northern Andean blueberries (Vaccinieae, Vaccinioideae, Ericaceae). *Int J Plant Sci*. 2003;164(6):987-995. doi 10.1086/378653

Ren N., Timko M.P. AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species. *Genome*. 2001;44(4):559-571. doi 10.1139/g01-060

Rossi L., Bindler G., Pijnenburg H., Isaac P.G., Giraud-Henry I., Mahe M., Gadani F. Potential of molecular marker analysis for variety identification in processed tobacco. *Plant Varieties Seeds*. 2001;14(2):89-102

Schell J., Van Montagu M. The Ti-plasmid of *Agrobacterium tumefaciens*, a natural vector for the introduction of *nif* genes in plants? In: Genetic Engineering for Nitrogen Fixation. Basic Life Sciences. Vol. 9. Springer, 1977;159-179. doi 10.1007/978-1-4684-0880-5_12

Schlautman B., Covarrubias-Pazaran G., Fajardo D., Steffan S., Zalapa J. Discriminating power of microsatellites in cranberry organelles for taxonomic studies in *Vaccinium* and Ericaceae. *Genet Resour Crop Evol*. 2017;64(3):451-466. doi 10.1007/s10722-016-0371-6

Schrinner S.D., Mari R.S., Ebler J., Rautiainen M., Seillier L., Reimer J.J., Usadel B., Marschall T., Klau G.W. Haplotype threading: accurate polyploid phasing from long reads. *Genome Biol*. 2020;21(1):252. doi 10.1186/s13059-020-02158-1

Shedlock A.M., Okada N. SINE insertions: powerful tools for molecular systematics. *BioEssays*. 2000;22(2):148-160. doi 10.1002/(SICI)1521-1878(200002)22:2<148::AID-BIES6>3.0.CO;2-Z

Sierro N., Battey J.N., Ouadi S., Bakaher N., Bovet L., Willig A., Goepfert S., Peitsch M.C., Ivanov N.V. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun*. 2014;5(1):3833. doi 10.1038/ncomms4833

Singer K. The mechanism of T-DNA integration: some major unresolved questions. *Curr Top Microbiol Immunol*. 2018;418:287-317. doi 10.1007/82_2018_98

Snyder M., Adey A., Kitzman J.O., Shendure J. Haplotype-resolved genome sequencing: experimental methods and applications. *Nat Rev Genet*. 2015;16(6):344-358. doi 10.1038/nrg3903

Soltis D.E., Mavrodiev E.V., Doyle J.J., Rauscher J., Soltis P.S. ITS and ETS sequence data and phylogeny reconstruction in allopolyploids and hybrids. *Syst Bot*. 2008;33(1):7-20. doi 10.1600/036364408783887401

Stalker H.T. Utilizing wild species for peanut improvement. *Crop Sci*. 2017;57(3):1102-1120. doi 10.2135/cropsci2016.09.0824

Sultana N., Menzel G., Heitkam T., Kojima K., Bao W., Serçe S. Bioinformatic and molecular analysis of satellite repeat diversity in *Vaccinium* genomes. *Genes* (*Basel*). 2020;11(5):527. doi 10.3390/genes11050527

Sun X., Jiao C., Schwaninger H., Chao C.T., Ma Y., Duan N., Khan A., Ban S., Xu K., Cheng L., Zhong G., Fei Z. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet*. 2020;52(12):1423-1432. doi 10.1038/s41588-020-00723-9

Suzuki K., Yamashita I., Tanaka N. Tobacco plants were transformed by *Agrobacterium rhizogenes* infection during their evolution. *Plant J*. 2002;32(5):775-787. doi 10.1046/j.1365-313X.2002.01468.x

Tanaka H., Hori T., Yamamoto S., Toyoda A., Yano K., Yamane K., Itoh T. Haplotype-resolved chromosomal-level assembly of wasabi (*Eutrema japonicum*) genome. *Sci Data*. 2023;10(1):441. doi 10.1038/s41597-023-02356-z

Tepfer D. Genetic transformation using *Agrobacterium rhizogenes*. *Physiol Plant*. 1990;79(1):140-146. doi 10.1111/j.1399-3054.1990.tb05876.x

Tian X., Shi L., Guo J., Fu L., Du P., Huang B., Wu Y., Zhang X., Wang Z. Chloroplast phylogenomic analyses reveal a maternal hybridization event leading to the formation of cultivated peanuts. *Front Plant Sci*. 2021;12:804568. doi 10.3389/fpls.2021.804568

Tiley G.P., Crowl A.A., Manos P.S., Sessa E.B., Solís-Lemus C., Yoder A.D., Burleigh J.G., Smith S. Benefits and limits of phasing alleles for network inference of allopolyploid complexes. *Syst Biol*. 2024;73(4):666-682. doi 10.1093/sysbio/syae024

Tong Z., Fang D., Chen X., Jiao F., Zhang Y., Li Y., Xiao B. Genome-wide association study of leaf chemistry traits in tobacco. *Breed Sci*. 2020;70(3):253-264. doi 10.1270/jsbbs.19067

Tzfira T., Li J., Lacroix B., Citovsky V. *Agrobacterium* T-DNA integration: molecules and models. *Trends Genet*. 2004;20(8):375-383. doi 10.1016/j.tig.2004.06.004

Vorsa N., Zalapa J. Domestication, genetics, and genomics of the American cranberry. In: Goldman I. (Ed.) Plant Breeding Reviews. Vol. 43. John Wiley & Sons, Inc., 2019;279-315. doi 10.1002/9781119616801.ch8

Wang H., Guo X., Hu X., Li T., Fu X., Liu R. Comparison of phytochemical profiles, antioxidant and cellular antioxidant activities of different varieties of blueberry (*Vaccinium* spp.). *Food Chem*. 2017;217:773-781. doi 10.1016/j.foodchem.2016.09.002

Wang Y., Zhao Y., Bollas A., Wang Y., Au K.F. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol*. 2021; 39(11):1348-1365. doi 10.1038/s41587-021-01108-x

Wenger A.M., Peluso P., Rowell W.J., Chang P., Hall R.J., Concepcion G.T., Ebler J., … Li H., Koren S., Carroll A., Rank D.R., Hunkapiller M.W. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol*. 2019;37(10):1155-1162. doi 10.1038/s41587-019-0217-9

White F.F., Ghidossi G., Gordon M.P., Nester E.W. Tumor induction by *Agrobacterium rhizogenes* involves the transfer of plasmid DNA to the plant genome. *Proc Natl Acad Sci USA*. 1982;79(10):3193-3197. doi 10.1073/pnas.79.10.3193

White F.F., Garfinkel D.J., Huffman G.A., Gordon M.P., Nester E.W. Sequence homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. *Nature*. 1983;301:348-350. doi 10.1038/301348a0

White F., Taylor B., Huffman G., Gordon M., Nester E. Molecular and genetic analysis of the transferred DNA regions of the root-inducing plasmid of *Agrobacterium rhizogenes*. *J Bacteriol*. 1985;164(1): 33-44. doi 10.1128/jb.164.1.33-44.1985

Wu Q., Tong W., Zhao H., Ge R., Li R., Huang J., Li F., Wang Y., Mallano A.I., Deng W., Wang W., Wan X., Zhang Z., Xia E. Comparative transcriptomic analysis unveils the deep phylogeny and secondary metabolite evolution of 116 *Camellia* plants. *Plant J*. 2022;111(2): 406-421. doi 10.1111/tpj.15799

Xie X., Ma X., Liu Y.G. Decoding Sanger sequencing chromatograms from CRISPR-induced mutations. *Methods Mol Biol*. 2019;1917: 33-43. doi 10.1007/978-1-4939-8991-1_3

Yukawa M., Tsudzuki T., Sugiura M. The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*. *Mol Genet Genom*. 2006; 275(4):367-373. doi 10.1007/s00438-005-0092-6

Zalapa J.E., Bougie T.C., Bougie T.A., Schlautman B.J., Wiesman E., Guzman A., Fajardo D.A., Steffan S., Smith T. Clonal diversity and genetic differentiation revealed by SSR markers in wild *Vaccinium macrocarpon* and *Vaccinium oxycoccos*. *Ann Appl Biol*. 2015; 166(2):196-207. doi 10.1111/aab.12173

Zhang X., Wu R., Wang Y., Yu J., Tang H. Unzipping haplotypes in diploid and polyploid genomes. *Comput Struct Biotechnol J*. 2020; 18:66-72. doi 10.1016/j.csbj.2019.11.011

Zhidkin R., Zhurbenko P., Bogomaz O., Gorodilova E., Katsapov I., Antropov D., Matveeva T. Biodiversity of *rolB/C*-like natural transgene in the genus *Vaccinium* L. and its application for phylogenetic studies. *Int J Mol Sci*. 2023;24(8):6932. doi 10.3390/ijms24086932

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

Received December 4, 2024. Revised May 19, 2025. Accepted May 26, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):868-882

doi 10.18699/vjgb-25-94

A new molecular marker including parts of conservative histone H3 and H4 genes and the spacer between them for phylogenetic studies in dragonflies (Insecta, Odonata), extendable to other organisms

A.V. Mglinets [1], V.S. Bulgakova[1], O.E. Kosterin [1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
kosterin@bionet.nsc.ru

Abstract. In spite of recent substantial progress in genomic approaches, there is still a need for molecular markers convenient for Sanger sequencing and providing good phylogenetic reconstructions at short evolutionary distances. A new molecular marker, the histone H3–H4 region, containing partial coding sequences of the genes for histones H3 and H4 and the non-coding spacer between them, is proposed. This marker is potentially useful for molecular phylogenetic studies at the generic, species, and even intra-species level in insects and some other organisms, even from other phyla. The highly conserved histone-coding sequences ensure the universality of primers and the ease of primary alignment, while the highly variable non-coding spacer provides enough variation for analyses at short evolutionary distances. In insects, the histone genes reside in the histone repeat which is tandemly repeated in dozens to hundred copies forming the so-called histone cluster. This ensures a high concentration of the template for the marker in genomic DNA preparations. However, the order and orientation of the histone genes in the histone repeat is variable among orders, which puts some limitations on the use of the proposed marker. The marker efficacy is hereby shown for Odonata (dragonflies and damselflies), where it provided good resolution at the family, genus and species levels. The new marker also provided an interesting pattern in the relationship of two *Sympetrum* species, *S. croceolum* and *S. uniforme*, showing the sequences of the latter as a branch nested among those of the former. The same combination of the proposed original primers should also work in Diptera.

Key words: histone repeat; histone H3; histone H4; intergenic spacer; Odonata; dragonflies; insects; molecular marker; phylogenetic studies

For citation: Mglinets A.V., Bulgakova V.S., Kosterin O.E. A new molecular marker including parts of conservative histone H3 and H4 genes and the spacer between them for phylogenetic studies in dragonflies (Insecta, Odonata), extendable to other organisms. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):868-882. doi 10.18699/vjgb-25-94

Funding. The work was supported by the project FWNR-2022-0019 of the Institute of Cytology and Genetics SB RAS. DNA sequencing was performed at the SB RAS Genomics Core Facility, Novosibirsk. Assembly of sequences from the SRA archive was performed at the Siberian SuperComputer Computational Facility, Novosibirsk.

Acknowledgement. The authors are grateful to V.V. Dubatolov for providing the specimens of *S. croceolum* and *S. uniforme* from the Far East of Russia and to the reviewers for their valuable comments.

Новый молекулярный маркер для филогенетических исследований стрекоз (Insecta, Odonata), включающий части консервативных генов гистонов H3 и H4 и спейсер между ними, применимый и к другим организмам

А.В. Мглинец [1], В.С. Булгакова[1], О.Э. Костерин [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
kosterin@bionet.nsc.ru

Аннотация. Несмотря на значительный прогресс в геномных исследованиях, до сих пор существует потребность в молекулярных маркерах, удобных для секвенирования по Сэнгеру и позволяющих делать филогенетические реконструкции на коротких эволюционных дистанциях. Предложен новый молекулярный маркер, район генов гистонов H3 и H4, содержащий части кодирующих последовательностей генов гистонов H3 и H4 и некодирующий спейсер между ними. Он может быть удобен для молекулярно-филогенетических исследований на родовом, видовом и даже внутривидовом уровне у насекомых и некоторых других групп, в том числе из других типов. Высококонсервативные кодирующие последовательности гистоновых генов обеспечивают уни-

версальность праймеров и однозначность первичного выравнивания, тогда как вариабельный некодирующий спейсер содержит достаточно изменчивости для анализа на коротких эволюционных дистанциях. У насекомых гены гистонов находятся в гистоновом повторе, тандемно повторенном в десятках или сотнях копий в так называемом гистоновом кластере. Это обеспечивает высокую концентрацию матрицы для нового маркера в препаратах геномной ДНК. Однако порядок и ориентация конкретных генов у разных отрядов изменчива, что налагает определенные ограничения на использование предлагаемого маркера. В нашей работе эффективность маркера продемонстрирована на материале отряда стрекоз (Odonata), где он хорошо разрешил семейства, роды и виды. Новый маркер также выявил интересный паттерн взаимоотношения двух видов рода *Sympetrum*, *S. croceolum* и *S. uniforme*, показав последовательности последнего в качестве ветви среди последовательностей первого. Та же комбинация разработанных нами оригинальных праймеров будет работать и в отряде двукрылых (Diptera).

Ключевые слова: гистоновый повтор; гистон H3; гистон H4; межгенный спейсер; Odonata; стрекозы; насекомые; молекулярный маркер; филогенетические исследования

Introduction

Analysis of DNA variation is a powerful tool in reconstructing phylogenetic history of living creatures, so the use of molecular methods has provided a profound progress in phylogenetic analysis, with applications in taxonomy, paleobiology, paleogeography and evolutionary theory. Particular sequences used for this purpose, traditionally called 'phylogenetic markers', differ in their rate of fixation of mutations thus permitting phylogenetic resolution at different time scales, with resolution of most recent divergences being possible with most variable markers; for those applied to Odonata see Y.C. Cheng et al. (2018).

Mitochondrial DNA in animals is on average more variable than nuclear DNA, so the popular mitochondrial marker cytochrome oxydase I (*COI*) is widely used for barcoding of animals (Ballard, Whitlock, 2004; Avise, 2009). However, the phylogenies reconstructed from mitochondrial markers quite frequently meet serious problems. They often contradict both phylogenies reconstructed from nuclear markers and traditional taxonomy based on morphology. In some groups of organisms (Cheng Z. et al., 2023), including the family Coenagrionidae in Odonata (Dow et al., 2019; Deng et al., 2021; Galimberti et al., 2021; Geiger et al., 2021), mitochondria seem to 'live a life of their own'. The most drastic patterns were revealed in *Coenagrion* Kirby, 1890, where mitochondrial haplotypes cross species barriers but seem unable to cross the Hybraltar Strait (Ferreira et al., 2016; Galimberti et al., 2021, Geiger et al., 2021), and in *Ischnura elegans* (Vander Linden, 1820) where the Japanese population appeared to be strongly divergent from the continental ones (Deng et al., 2021). Such drastic patterns can hardly be explained by incomplete lineage sorting or introgressive hybridisation, the phenomena usually supposed to explain such cases. The reason could be co-selection of mitochondria with strains of the intra-cellular bacterial endosymbiont *Wolbachia* Hertig, 1826 (Deng et al., 2021). Horizontal transfer of mitochondria via an unknown agent is also postulated but not proved (Gurdon et al., 2016). Another problem of using markers based on mitochondrial DNA is the sporadic occurrence of mtDNA fragments adopted by the nuclear genome (NUMT), which are variably divergent from the actual mtDNA and may result in false phylogenetic results; for examples in Odonata, see Ožana et al. (2022) or Lorenzo-Carballa et al. (2022). For the purpose of revealing evolutionary history at short time distances, microsatellite or SSR markers have been popular for a long time and applied also to Odonata (e. g. Lowe et al., 2008). These are tracts of very short (one to few nucleotides) repeats with the number of copies frequently changing because of slippage mispairing and unequal crossing over. Such markers suffer from a high rate of homoplasy, when different events of copy number change result in the same alleles, here understood as particular numbers of repeats, but are still the marker of choice for population genetic studies.

The modern next-generation high-throughput sequencing and genomic approaches offer ample phylogenetic data (for examples in Odonata, see Futahashi et al., 2015; Bybee et al., 2021; Kohli et al., 2021), potentially useful for analysis even at short evolutionary distances, but are expensive and more demanding in sample preparation. Although the prices are getting lower, these technologies still remain unaffordable for many researchers in countries which harbour the richest biodiversity. Therefore, there is still a need for easily analysed and cheap nuclear markers based on Sanger sequencing which would provide good resolution at the species level and could be useful at least for fast preliminary tests revealing the evolutionary history of populations, subspecies and closely related species. Besides, such markers make it possible to analyse regular collection specimens, not specially collected for DNA analysis.

Non-coding sequences, the variability of which is mainly determined by physical properties of DNA replication, are useful candidates for variable phylogenetic markers. Their use is limited by possibility of working out universal primers, which could be achieved by involvement of bordering conserved sequences. Besides, non-coding sequences demonstrate a high rate of indels, which bring about difficulties as to their alignment. Of such markers, the so-called *ITS* region including internal spacers between the conserved ribosomal RNA genes, *ITS1* and *ITS2*, is the most popular among nuclear markers; for its use in Odonata, see Hovmöller and Johansson (2004), Dumont et al. (2010), Karube et al. (2012), Schneider et al. (2023). The highly repetitive nature of the nucleolus organiser provides an advantage of high concentration of the template in preparations of genomic DNA and a disadvantage of possible heterozygosity as well as cis-heterogeneity between individual repeat copies (Hovmöller, Johansson, 2004).

Recently, a useful approach has been proposed, focusing on introns of nuclear genes (Ferreira et al., 2014). The primary structure of introns has scarce adaptive constraints save mutations affecting splicing. At the same time, the bordering exons are usually conserved enough to allow for universal primers design. These markers have a disadvantage of low concen-

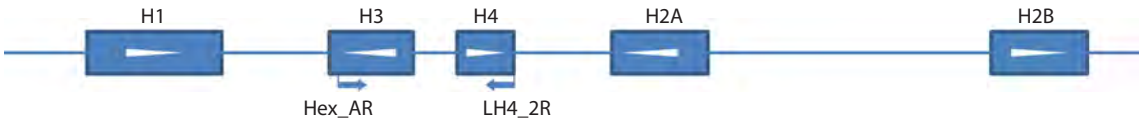


Fig. 1. Positions of the primers used to amplify the histone H3–H4 region in the histone repeat as exemplified by a fragment of the assembly of the *Ischnura elegans* genome (NW_025791746). White arrowheads indicate the direction of transcription.

tration of template genomic DNA, since the genes involved are unique and present in the genome only in two copies, so they may be less readily amplifiable from specimens with somewhat degraded DNA as compared to the highly repeated sequences of the *ITS* region.

In the present work, we propose and test the usability of a new phylogenetic marker, the spacer between the genes of the conservative core histones H3 and H4 and partial coding sequences of these histones, which we designate as *the histone H3–H4 region*. It resembles the popular ITS region mentioned above.

In animals, the genes for five histones (H1, H2A, H2B, H3, H4) are included into the so-called histone repeat, tandemly repeated copies of which form the histone cluster (Eirín-López et al., 2009). Several important circumstances should be noted in this respect:

(i) Histones H3 and H4 are among the most conserved proteins in eukaryotes (Stein et al., 1984; Doenecke et al., 1997; Eirín-López et al., 2009), so their coding sequences allow primers of very broad applicability across taxonomic groups.

(ii) In insects, genes of these two histones are disposed in the histone repeat relatively close to each other (Eirín-López et al., 2009). The order of histone genes and orientation of their reading frames is variable among different insect orders. For instance, their order is H1, H3, H4, H2A, H2B in *Drosophila melanogaster* (Goodenough, 1984: p. 304; Eirín-López et al., 2009) and in all Odonata species tested in this work (Fig. 1). Although this variation may also take place at lower taxonomical levels and is sometimes observed even in different copies of the histone repeat in the same chromosome, in all species of Odonata tested by us (see below) we obtained a PCR product using the same primer pair matching histones H3 and H4. So the original primers we suggest are useful at least across dragonflies and damselflies.

(iii) The spacer between H3 and H4 histone genes is non-coding and therefore is expected to undergo neutral evolution, hence being a kind of molecular clock.

(iv) Insects have hundreds of copies of the histone repeat (Stein et al., 1984; Solovyev et al., 2022), which facilitates amplification from total genomic DNA preparations, but may also bring about problems related to cis- (within-cluster) and trans- (allelic, between homologous chromosomes) heterogeneity.

(v) Most insects have only one histone cluster, whereas in some of them and in other animal groups there is a number of paralogous clusters (Eirín-López et al., 2009). A single histone cluster is an advantage since this avoids trans-cluster heterogeneity which would complicate an analysis.

To develop and test the marker, we chose the order Odonata and designed original primers which worked in all tested species. We tested its resolution at different taxonomical levels, by sequencing amplicons from representatives of different families (Calopterygidae, Coenagrionidae, Aeshnidae, Gomphidae, Corduliidae, and Libellulidae), from several species of some genera and from a number of specimens of some *Sympetrum* spp. The latter involved a series of three species from the same *danae* species group (Pilgrim, Dohlen, 2012), namely *S. croceolum* (Selys, 1883), *S. danae* (Sulzer, 1776) and *S. uniforme* (Selys, 1883), including those collected in the same populations.

Materials and methods

Material. The specimens of *S. croceolum*, *S. danae*, *S. flaveolum* (Linnaeus, 1758), and *S. uniforme* were preserved in 96 % ethanol, other specimens were treated overnight with acetone and then dried out. The species and specimens from which the histone H3–H4 region was sequenced in the course of this study and the GenBank accession numbers of these sequences are enumerated (in parentheses) in the Table.

Sequences from public databases. To expand our sample, we downloaded sequences of the H3–H4 region from the Whole Genome Sequence datasets available in public databases of 16 more Odonata species: *Hetaerina americana* (Fabricius, 1798), *H. titia* (Drury, 1773) (Calopterygidae), *Argia fumipennis* (Burmeister, 1839), *Ceriagrion tenellum* (De Villers, 1789), *Ischnura elegans*, *I. senegalensis* (Rambur, 1842) *Pseudagrion microcephalum* (Rambur, 1842), *Pyrrhosoma nymphula* (Sulzer, 1776) (Coenagrionidae), *Platycnemis pennipes* (Pallas, 1771), *Prodasineura notostigma* (Selys, 1860) (Platycnemididae), *Tanypteryx hageni* (Selys, 1879), *Tachopteryx thoreyi* (Selys, 1889), *Uropetala carovei* (White in Dieffenbach, 1843) (Petaluridae), *Brachytron pratense* (Müller, 1764) (Aeshnidae), *Pachydiplax longipennis* (Burmeister, 1839), *Pantala flavescens* (Fabricius, 1798), (Libellulidae).

Besides, the histone H3–H4 region was assembled from SRA archives available at GenBank, with the use of the MIRA software (Chevreux et al., 1999) for the following 20 species: *Archilestes grandis* (Rambur, 1842) (Lestidae), *Calopteryx splendens* (Harris, 1780), *Hetaerina vulnerata* (Hagen in Selys, 1853), *Mnais tenuis* (Oguma, 1913), *Neurobasis kaupi* (Brauer, 1867) (Calopterygidae), *Agriocnemis femina* (Brauer, 1868) (Coenagrionidae), *Anax parthenope* (Selys, 1839), *A. strenuus* (Hagen, 1867) (Aeshnidae), *Gomphus vulgatissimus* (Linnaeus, 1758), *Lanthus parvulus* (Selys, 1854), *Onychogomphus forcipatus* (Linnaeus, 1758), *Ophiogomphus mainensis* (Packard in Walsh, 1863) (Gomphidae), *Cordulegaster boltonii* (Donovan, 1807) (Cordulegastridae), *Macromia manchurica* (Asahina, 1964) (Macromiiddae), *Ladona fulva* (Müller, 1764), *Leucorrhinia albifrons* (Burmeister, 1839), *Libellula angelina* (Selys, 1883), *L. quadrimaculata* (Linnaeus, 1758), *Nannophya pygmaea* (Rambur, 1842), *Orthetrum coerulescens* (Fabricius, 1798) (Libellulidae). The

Species (by families) and specimens sequenced for the histone H3–H4 region, their origin and the GenBank accession numbers of the sequences. Coordinates are given in decimal degree format

Species name	Locality	Latitude (N)	Longitude (E)	Date	Specimens, their codes (if any) and accession numbers of their sequences
		Calopterygidae			
Atrocalopteryx atrata (Selys, 1853)	Russia, Khabarovskiy Kray, 44 km SE of Khabarovsk, Chirki	48.19	134.68	15.07.2014	1 ♂ (PQ498535)
		Coenagrionidae			
Coenagrion johanssoni (Wallengren, 1894)	Russia, Khabarovskiy Kray, 44 km SE of Khabarovsk, Chirki	48.19	134.68	15.07.2014	1 ♂ (PQ498539)
		Aeshnidae			
Aeshna juncea (Linnaeus, 1758)	Russia, Khabarovskiy Kray, 28 km SE of Khabarovsk, Bychikha village	48.30	134.83	9.07.2014	1 ♂ (PQ498536)
		Gomphidae			
Nihonogomphus ruptus (Selys in Hagen, 1858)	Russia, Khabarovskiy Kray, 44 km SE of Khabarovsk, Chirki ranger station	48.19	134.68	15.07.2014	1 ♂ (PQ498542)
		Corduliidae			
Cordulia aenea (Linnaeus, 1758)	Russia, Novosibirsk, Academy Town	54.8473	83.1072	1.07.2015	1 ♂ (PQ498537)
Somatochlora arctica (Zetterstedt, 1840)	Russia, Novosibirsk, Academy Town, Zvezda Beach	54.8314	83.0538	30.06.2015	1 ♀ (PQ498546)
Somatochlora graeseri Selys, 1887	Russia, Primorskiy Kray, Khanka District, Lake Khanka W bank 2 km N of Platono-Aleksandrovskoe village	45.062	131.995	4.07.2014	1 ♀ (PQ498561)
		Macromiidae			
Macromia amphigena fraenata Martin, 1906	Russia, Khabarovskiy Kray, 44 km SE of Khabarovsk, Chirki	48.19	134.68	15.07.2014	1 ♂ (PQ498540)
		Synthemistidae *sensu lato*			
Macromidia rapida Martin, 1907	Cambodia, Mondulkiri Province, a left tributary of the Monorom Waterfall River, 3.5 km SE of Sen Monorom	12.441	107.159	13.06.2014	1 ♂ (PQ498541)
		Cordulegastridae			
Cordulegaster picta Selys, 1854	Russia, Krasnodarskiy Kray, Kabardinka village, the Krasnaya Shchel' Valley	44.671	37.918	30.07.2015	1 ♀ (PQ498538)
		Libellulidae			
Orthetrum albistylum (Selys, 1848)	Russia, Primorskiy Kray, Khanka District, Lake Khanka W bank 2 km N of Platono-Aleksandrovskoe village	45.062	131.995	4.07.2014	1 ♀ (PQ498543)
Orthetrum glaucum (Brauer, 1865)	Cambodia, Mondulkiri Province, a brook downstream of the Buu Sraa Waterfall	12.570	107.418	10.06.2014	1 ♂ (PQ498544)
Rhyothemis phyllis (Sulzer, 1776)	Cambodia, Kampot Province, a left oxbow of the Kampot River downstream of Tek Chhou Rapids	10.672	104.137	21.08.2011	1 ♀ (PQ498545)
Sympetrum arenicolor Jödicke, 1994	Iran, Lorestan, Khorramabad County, 3.2 km NNW Pasil village	33.33	48.88	26.05.2017	1 ♂ (PQ498547)
Sympetrum cordulegaster	Russia, Khabarovskiy Kray, 28 km SE of Khabarovsk, a pond S of Bychikha village	48.292	134.829	21-22.08.2016	1 ♂ (PQ498588)

Table (end)

Species name	Locality	Latitude (N)	Longitude (E)	Date	Specimens, their codes (if any) and accession numbers of their sequences
Sympetrum croceolum (Selys, 1883)	Russia, Khabarovskiy Kray, Amursk District, Bolon'skiy Nature Reserve, Kirpu ranger station	49.506	136.028	15.09.2016	10 ♂♂ (Sc-1 – Sc-10; PQ498574–PQ498582)
	Russia, Khabarovskiy Kray, 28 km SE of Khabarovsk, Bychikha village	48.292	134.829	21-22.08.2016	2 ♀♀ (Sc-31, Sc-32; Q498583, PQ498586)
	Russia, Altai Republic, Mayma District, Lake Manzherok SE bank	51.82	85.81	16.09.2016	2 ♂♂ (Sc-41, Sc-42; PQ498584, PQ498585)
Sympetrum danae (Sulzer, 1776)	Russia, Novosibirsk, Academy Town, Botanical Garden	54.825	83.114	8.10.2015	12 ♂♂ (Sd-1, Sd-3 – Sd-12; PQ498549–PQ498557)
Sympetrum flaveolum (Linnaeus, 1758)	Russia, Novosibirsk, Academy Town, Botanical Garden	54.825	83.114	8.10.2015	1 ♂ (Sf-1; PQ498558), 1 ♀ (Sf-2; PQ498559)
	Russia, Primorskiy Kray, Lake Khanka W bank just N of Platono-Aleksandrovskoe village env.	45.06	131.99	4.07.2014	1♀ (Sf-3; PQ498560)
Sympetrum fonscolombii (Selys, 1840)	Iran, Esfahan Province, Golpayegan City, the Ghomrood (Anaarbar) River	33.46	50.26	21.05.2017	1 ♂ (No. 1)
	Russia, Krasnodarskiy Kray, Bol'shoy Utrish village env.	44.76	37.39	5.08.2015	1 ♂ (No. 2; PQ498589)
Sympetrum pedemontanum (Müller in Allioni, 1766)	Russia, Novosibirsk, a glade in a pine forest between Nizhnyaya El'tsovka and Pravye Chyomu estates	54.864	83.051	4.08.2018	1 ♂ (PQ498591)
Sympetrum sanguineum (Müller, 1764)	Russia, Novosibirsk Province, Kyshtovka District, a roadside swamp in the Kyshtovka village NW suburbs	56.579	76.603	14.07.2015	1 ♂ (Ss-1, PQ498590)
	Russia, Krasnodarskiy Kray, Lake Krugloe	44.677	37.594	10.06.2016	1 ♂ (Ss-2)
Sympetrum striolatum (Charpentier, 1840)	Iran, Lorestan, Borujerd County, Chenar Chashmah	33.798	48.905	25.05.2017	1 ♂ (PQ498592)
Sympetrum vulgatum (Linnaeus, 1758)	Russia, Krasnodarskiy Kray, Kabardinka village env., Krasnaya Shchel' Valley	44.672	37.922	6.07.2016	1 ♂ (PQ498593)
Sympetrum uniforme (Selys, 1883)	Russia, Khabarovskiy Kray, 28 km SE of Khabarovsk, Bychikha village	48.292	134.829	21-22.08.2016	5 ♂♂, 4 ♀♀ (Su-21–Su-29; PQ498565–PQ498573
	Russia, Khabarovskiy Kray, Amursk District, Bolon'skiy Nature Reserve, Kirpu ranger station	49.506	136.028	15.09.2016	1 ♂ (Su-1; PQ498563), 1 ♀ (Su-2; PQ498564)

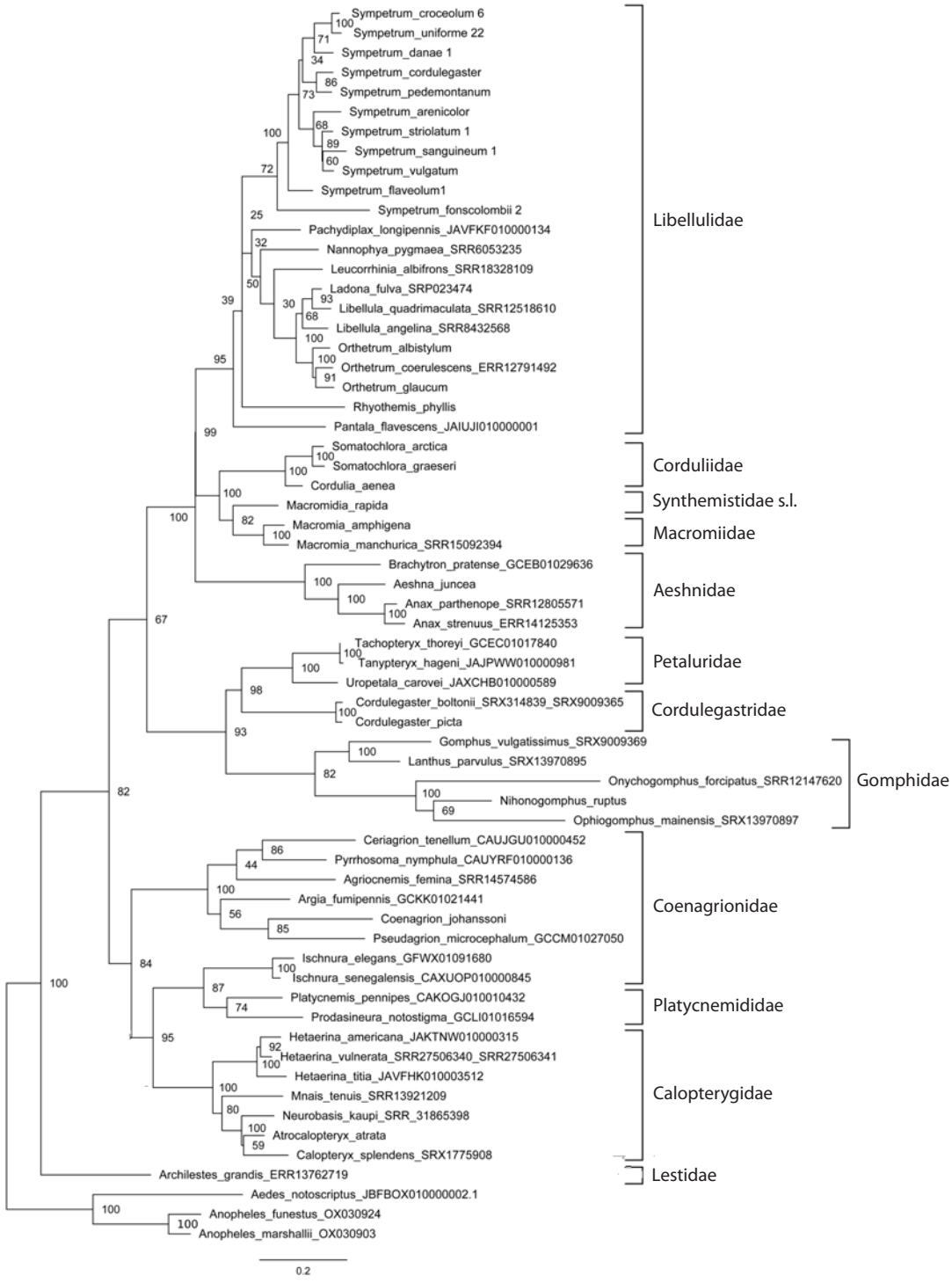


Fig. 2. Phylogenetic tree of the studied species of Odonata reconstructed with Maximum Likelihood method from the histone H3–H4 region sequences. Bootstrap values are shown at the respective nodes. Three species of Diptera, Culicidae serve as the outgroup.

accession numbers of the database entries used as sources of these sequences are indicated at species names in Figures 2 and 3.

Primer design. We designed 14 original primers to match different parts of insect histone genes coding for H1, H2B, H3, H4, comprising the histone gene cluster. At the start of the present work we did not know the precise order and orientation of the histone genes in the cluster, so we tested different primer combinations to select primer pair(s) which would produce an amplification product containing the fragments of the genes of histone H3 and H4 and the spacer between them. We found out that the pair of primers Hex_AR matching the 3′ portion of the H3 gene (in the orientation opposite to that of transcription) and LH4_2R matching the 3′ portion of the

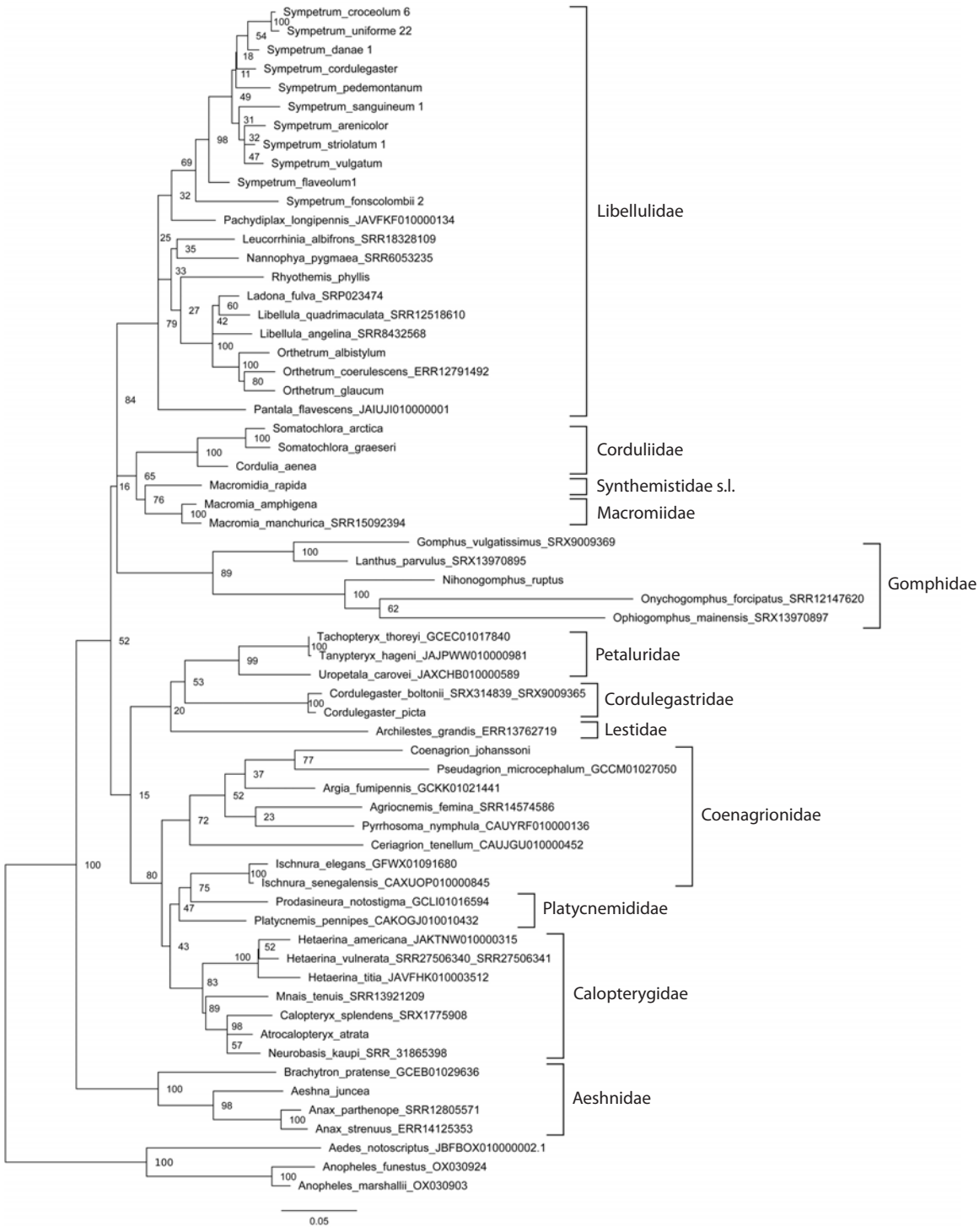


Fig. 3. Phylogenetic tree of the studied Odonata species reconstructed with the Maximum Likelihood method from fragments of the coding sequences of histone H3 and histone H4 genes involved into the proposed 'histone H3–H4 region' marker. Bootstrap values are shown at the respective nodes.

H4 gene (also in the orientation opposite to that of transcription) (Fig. 1) produced the expected product, indicating that the H3 and H4 genes were oriented anti-parallel as to their reading frames, with 5′ ends of their coding chains oriented towards each other (Fig. 1). The primer sequences are as follows:

Hex_AR: 5′ atatccttgggcatgatggtgac (forward)
LH4_2R: 5′ ttaaccgccgaaaccgtacagggt (reverse).

The primer LH4_2R, matching the coding region of the histone H4 of the moth *Bombyx mori* (Linnaeus, 1758) (Lepidoptera: Bombycidae) (GenBank accession AADK01010708), was worked out in the course of our previous study of the variation of the histone H1 gene in some Lepidoptera (Solovyev et al., 2015), although this particular primer was not mentioned in the cited work and is published here for the first time.

The Hex_AR primer was worked out to match the coding sequences of the histone H3 gene of *Ophiogomphus severus* (Hagen, 1874) (Odonata: Gomphidae) taken from GenBank (accession AY125228).

The coding sequences of histones H3 and H4 are well conserved (Stein et al., 1984; Doenecke et al., 1997; Eirín-López et al., 2009), so the primers worked out to match a particular sequence, one of which was from Odonata and the other from Lepidoptera, worked well for all tested species of Odonata.

DNA isolation, sequencing and analysis. Dragonfly legs were homogenized in a mortar in 0.2 ml of isolation buffer (0.1 Tris-HCl, pH 8.0; 0.05 M EDTA; 1.25 % SDS; 0.5 M NaCl) with Al_2O_3 as grinding particles, then mixed with 0.8 ml of the same buffer. The mixture was incubated for 1 h at 55 °C, then 350 µl of 5 M potassium acetate was added, the mixture was incubated for 30 min on ice and centrifuged at 16.1 g for 10 min. The supernatant was transferred to fresh tubes, mixed with 0.6 ml of isopropanol, incubated at room temperature for 1 h and centrifuged at 12.2 g for 10 min. The precipitate was washed twice with 0.1 ml 70 % ethanol with subsequent centrifugation at 12.2 g for 5 min, dried at 50 °C for 5 min and dissolved in 50 µl of deionized water.

PCR reaction was carried out in a volume of 20 µl with 2 µl of 10× ammonium sulphate buffer, 2 µl of 25 mM $MgCl_2$, 0.3 µl of the Hot Start Taq polymerase produced by SIBENZYME company, Novosibirsk (5 U/µl), 0.15 µl BSA (10 mg/ml), 1 µl of forward and reverse primers (10 pM) each, 2 µl of 2 mM dNTPs, 2 µl of diluted DNA (20–60 ng) and 10.55 µl of deionised water. For PCR, BIO-RAD MyCycler thermal cycler was used, with the reaction parameters as follows: denaturation at 95 °C for 3 min followed by 32 cycles including denaturation at 94 °C for 30 s, annealing at 55 °C for 25 s, elongation at 72 °C for 45 s. The PCR products were purified with Invisorb® Spin Filter PCRapid Kit and Sanger sequenced using Big Dye Terminators version 3.0 or 1.1 at SB RAS Genomic Core Facility.

Raw trace files were visualized and translated into nucleotide sequences with the use of the Gap4 software (Staden et al., 2003). The sequences were aligned with ClustalW (Larkin et al., 2007) using the MEGA 6.0 software package (Tamura et al., 2013) with default parameters. For a separate analysis of the non-coding spacer between the histone H3 and H4 genes, the relevant part of the alignment of the entire histone H3–H4 region was used, since separate alignment of the spacer sequences as such is less reliable.

The phylogenetic relationships were reconstructed with the Maximum Likelihood method using MEGA 6.0, with Kimura 2-parameter substitution model, as default in the package; rate among sites: gamma-distributed with invariant sites. Bootstrap values from 100 replications were calculated. The sequences of the histone H3–H4 region of three species of Diptera, *Aedes notoscriptus* (Skuse, 1889), *Anopheles funestus* (Giles, 1900) and *A. marshalli* (Theobald, 1903) from GenBank, were used as the outgroup for the order Odonata-wide phylogenetic reconstruction (Figs 2–4), because in Diptera we found the same order and orientation of the genes of histones H3 and H4.

The uncorrected p-distances between different alleles of the histone H3–H4 regions within species of *Sympetrum* were calculated with the MEGA 6.0 software (the entire matrix is not shown).

Results

The histone H3–H4 region was successfully amplified with the above suggested primer pair and sequenced from DNA isolated from specimens of Odonata enumerated in the Table, 59 individuals of 24 species. Together with the 36 sequences adopted from public databases, this comprised a sample of 95 sequences of 60 species.

The sequences of the histone H3–H4 region contained parts of the conservative coding sequences of the genes of histones H3 (351 b.p.) and H4 (288 b.p.) and the spacer between them of a variable length of about 250 b.p. All substitutions revealed in the coding sequence fragments were synonymous except for the substitution T > A in the first position of the second codon of the histone H4 gene, which changes threonine to serine, in both sequenced specimens of *S. fonscolombii* (Selys, 1840) (not shown). As expected, the sequences of bordering coding sequences were unambiguously aligned. At the same time the spacer is expectedly hyper-variable and exhibits a high rate of indels, so its alignment was much less certain and retained some ambiguity. The alignment involving one sequence per each studied species, including the outgroup, used for reconstruction of the phylogenetic tree of Figure 2, was 1019 b.p. long and had 583 (57 %) variable sites, 529 (52 %) parsimoniously informative sites and 42 (4 %) singletons. These numeric estimates, however, are conventional and should be taken with caution because of uncertainty of the alignment of the non-coding sequence of evolutionary distant species.

One specimen of *S. sanguineum* (Müller, 1764) (Ss-2), one specimen of *S. fonscolombii* (No. 1), and one specimen of *S. uniforme* (Su-23) appeared to be heterogeneous containing reads with and without deletion of a number of nucleotides in the spacer. One of those indels found in *S. sanguineum* (Ss-2) concerned just one base pair; so we were able to infer both sequence variants from the chromatogram but used for further analysis only one of them, chosen randomly. Indels found in the other two species were longer, about 5 and 10 b.p. Although the sequences beyond the deleted region could be reconstructed, we preferred to exclude these specimens from further analyses.

In some positions, the chromatograms revealed two peaks of comparable height suggesting within-specimen heterogeneity for nucleotides occupying these positions. Those positions reflected either heterozygosity for different alleles or cis-heterogeneity for the histone repeat along a histone cluster, quite expectable in the case of repeated units. Such positions made uncertain the exact number of unique alleles found in a species. A few chromatograms did not resolve nucleotides in a number of positions adjacent to the primers; we nevertheless involved the shortened, well resolved sequences into phylogenetic reconstructions.

First, we reconstructed a phylogenetic tree based on the sequence of the H3–H4 region from one representative of each involved species, both newly sequenced and available or reconstructed from public databases (Fig. 2). The tree was rooted with the sequences of mosquitoes *Anopheles* and *Aedes* used as the outgroup. The overall tree topology well corresponded to the family system of Odonata (Dijkstra et al., 2013), except for the odd position of the genus *Ischnura*, which is attributed to Coenagrionidae but clustered with Platycnemididae, although with a weak bootstrap support of 83.

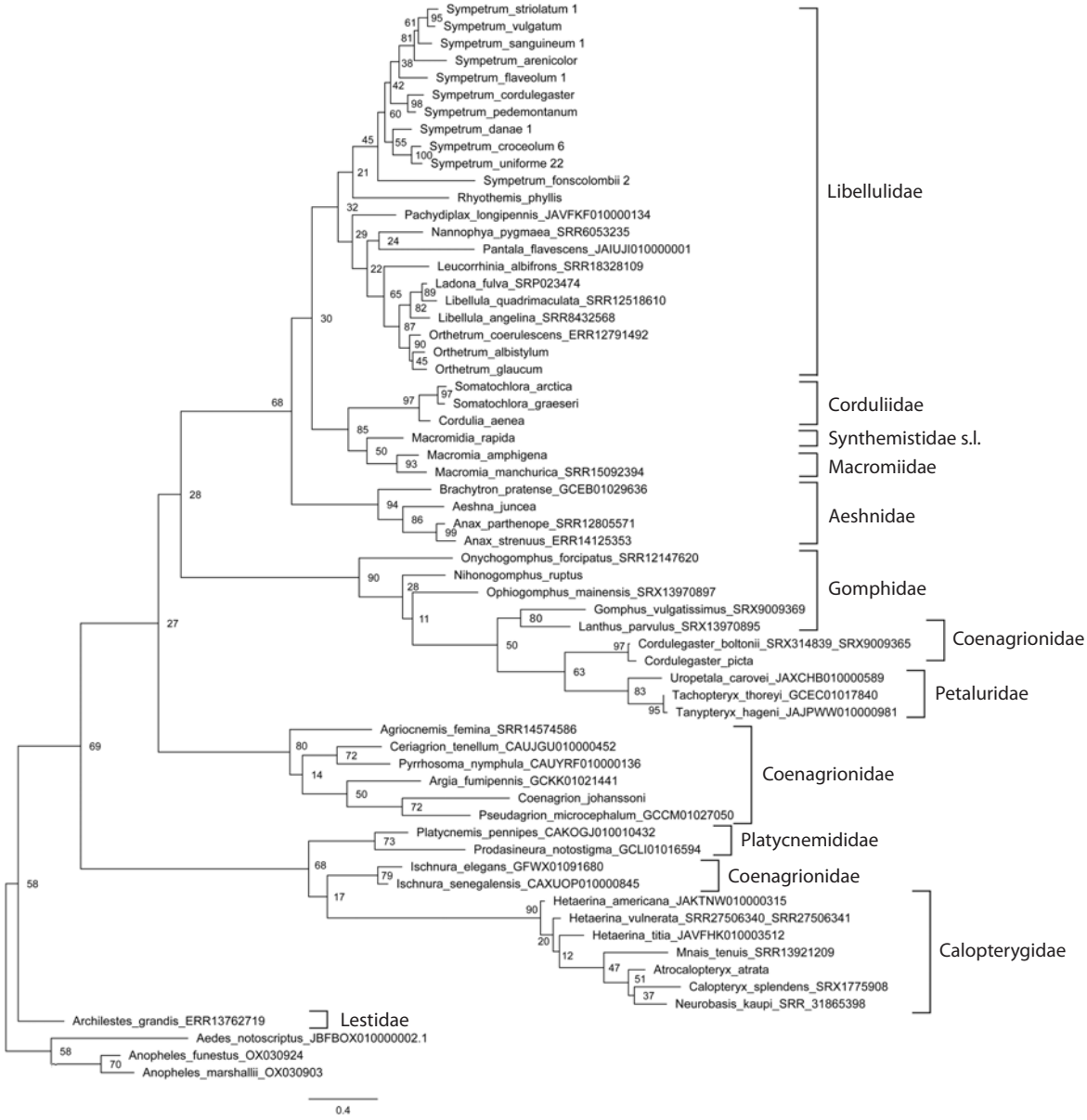


Fig. 4. Phylogenetic tree of the studied Odonata species reconstructed with the Maximum Likelihood method from the intergenic spacer between histone H3 and histone H4 genes. Bootstrap values are shown at the respective nodes.

Our tree included 10 currently recognised families of Odonata (Dijkstra et al., 2013) represented by several species. Seven of them were revealed as monophyletic clades well supported by high bootstrap values: Libellulidae (95), Corduliidae (100), Macromiidae (100), Aeshnidae (100), Petaluridae (100), Cordulegastridae (100) and Calopterygidae (100). Two families appeared monophyletic with weak support: Gomphidae (82) and Platycnemididae (74). The cluster of Coenagrionidae without *Ischnura* had the highest support of 100. Even representatives of the three families (Corduliidae, Macromiidae and Synthemistidae s. l.), previously considered in the family Corduliidae in the broad sense, also grouped in a cluster with the maximum support of 100.

Archilestes grandis is the only involved representative of Lestidae, the family considered to retain most plesiomorphic characters among Odonata (Dijkstra et al., 2013). Hence its position as the most basal branch of Odonata was rather expected. If we were to exclude this branch formally attributed to Zygoptera, both suborders Anisoptera and Zygoptera would appear monophyletic but weakly supported (67 and 84, respectively).

The tree of Figure 2 includes nine genera represented by more than one species. Seven of them (*Orthetrum* Newmann, 1833, *Somatochlora*, *Macromia* Rambur, 1842, *Anax* Leah in Breuster, 1815, *Cordulegaster* Leah in Breuster, 1815, *Ischnura* Charentier, 1840, *Hetaerina* Hagen in Selys, 1853) had the highest support of 100. The genus *Sympetrum*, represented by 11 species, formed a monophyletic cluster with weak support (72). However, if we were to exclude the problematic (see below) divergent species *S. fonscolombii*, the remained 10 species would cluster together with the maximum support of 100. The genus *Libellula* Linnaeus, 1758 would

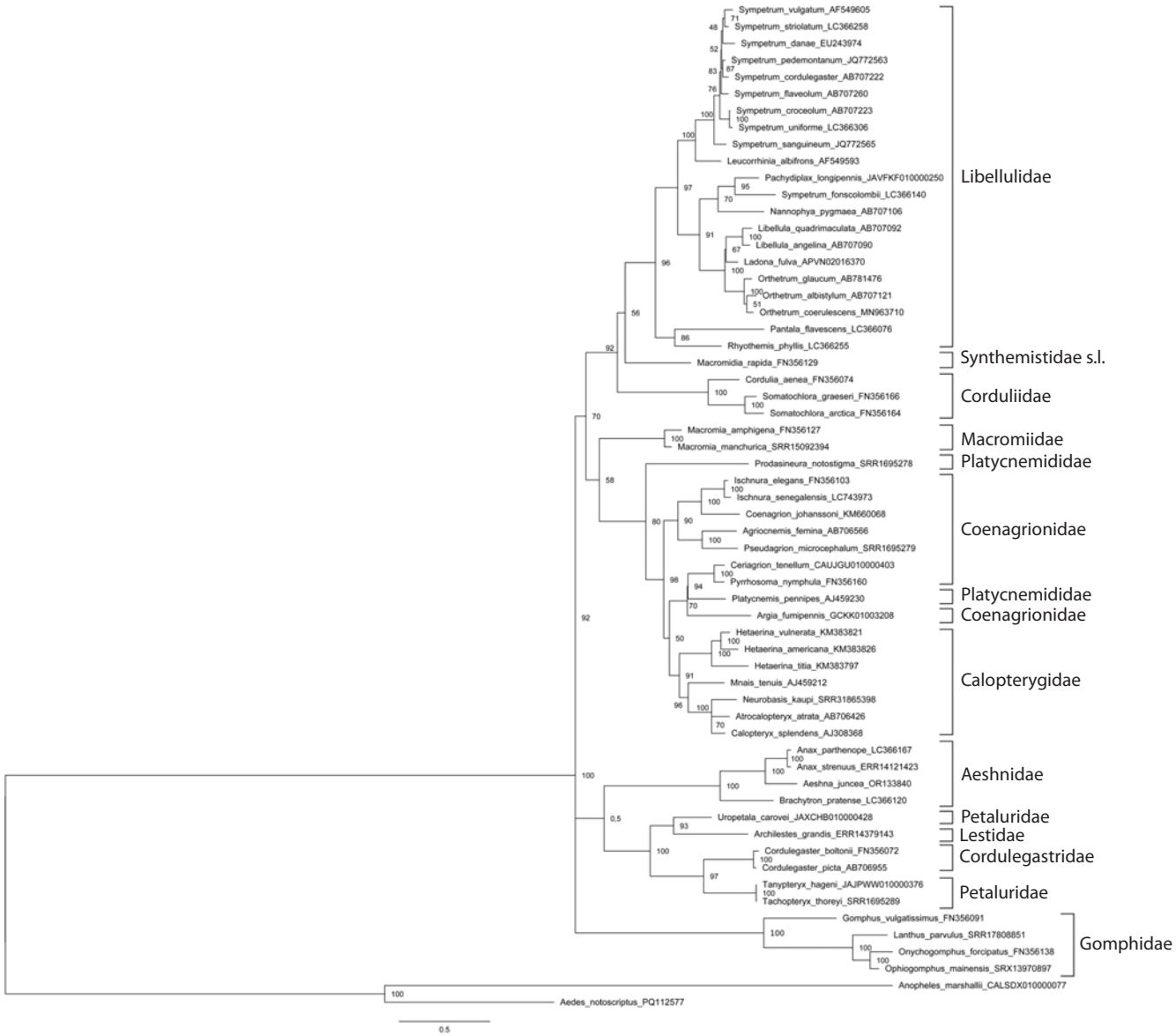


Fig. 5. Phylogenetic tree of the species of Odonata as in Figure 2 (with two omissions) reconstructed with the Maximum Likelihood method from the *ITS* region sequences adopted from GenGank. Bootstrap values are shown at the respective nodes. Two species of Diptera, Culicidae serve as the outgroup.

become monophyletic but weakly supported (68) if we were to assume *Ladona* Needham, 1897 to be its synonym, as it is often considered.

It was interesting to evaluate separate inputs into this phylogenetic resolution of the histone coding sequences and spacer, so we reconstructed phylogenetic trees based on these two components separately (Figs 3, 4). In both trees, terminal branches uniting close species or genera are mainly well supported. The support of families is somewhat lower than in the tree based on the entire histone H3–H4 region (Fig. 2), with the values in the tree based on the concatenated coding sequences of both histone genes (Fig. 3) being in general higher than in the tree based on the non-coding spacer (Fig. 4). The principal topology of the tree based on the spacer sequences (Fig. 4) remained similar to that of the tree based on the entire H3–H4 region (Fig. 2), but is not supported. The overall topology with respect to positions of families of the tree based on the coding sequences (Fig. 3) is different, does not reflect dichotomy for the two suborders and is even less supported than in the spacer tree (Fig. 4). This can be attributed to saturation of conservative histone gene sequences by synonymous substitutions at long evolutionary distances. Altogether, we may conclude that both parts of the histone H3–H4 region have their input into its resolving power, but the best result is produced by the two parts taken together.

As stated above, the phylogenetic marker proposed here, the histone H3–H4 region, is similar to the popular *ITS* region containing rRNA genes and two non-coding spacers between them. To compare phylogenetic resolution of these two markers, we reconstructed a phylogenetic tree based on the *ITS* region sequences adopted from GenBank, which contains the same species except for *N. ruptus* and *S. arenicolor* (Fig. 5). The alignment of these sequences was 1048 b.p. long and had 878 (84 %) variable sites, 529 (71 %) parsimoniously

Fig. 6. Phylogenetic tree of the studied *Sympetrum* species reconstructed with the Maximum Likelihood method from the histone H3–H4 region sequences. Bootstrap values are shown at the respective nodes. *Orthetrum albistylum*, *O. glaucum* and *Rhyothemis phyllis* serve as outgroup.

informative sites and 118 (11 %) singletons. These data are also affected by the ambiguity of the alignments of the non-coding spacers as those for the histone H3–H4 region and should be taken with caution. The *ITS* region contained a somewhat greater share of parsimoniously informative sites than the histone H3–H4 region, 71 vs 52 %. However, the *ITS* marker appeared substantially inferior in resolving odonate families as compared to the histone H3–H4 region, as seen in the phylogenetic tree reconstructed from the ITS region (Fig. 5). This tree contains a number of awkwardly placed species. The zygopteran *Archilestes grandis* occurs among Anisoptera where it clusters with *U. carovei*, which in turn does not cluster with the two other Petaluridae. *P. pennipes* does not cluster with the second Platycnemididae, *P. notostigma*, but occurs among representatives of Coenagrionidae. The branch of Macromiidae does not cluster with other Anisoptera. *S. fonscolombii* is far decoupled from other *Sympetrum* spp. We may conclude that the *ITS* region is unable to adequately resolve the phylogeny at the level of Odonata families.

To test the applicability of the proposed marker, the histone H3–H4 region, to evaluating intra-generic and intra-species variation, we estimated its variation and reconstructed a phylogenetic tree for 45 specimens belonging to 11 species of the genus *Sympetrum* involved, using the sequences of *Rhyothemis phyllis* (Sulzer, 1776), *Orthetrum albistylum* (Selys, 1848) and *O. glaucum* (Brauer, 1865) as the outgroup (Fig. 6). The magnitude of intra-species variation of the histone H3–H4 region sequence appeared quite substantial. For the three species represented by 10 to 14 specimens, *S. croceolum*, *S. uniforme* and *S. danae*, the maximum uncorrected p-distances (that is, the share of variable positions among all positions) between different alleles within a species appeared to be respectively 0.0216, 0.0037 and 0.0011, that is ca 2.1, 0.4 and 0 %. The averaged differences between any two sequences within *S. croceolum*, *S. uniforme* and *S. danae* were 0.0165, 0.0009 and 0.0003, respectively.

In the reconstructed phylogenetic tree (Fig. 6), ten sequences of *S. uniforme*, ten sequences of *S. danae* and three sequences of *S. flaveolum* expectedly clustered with the maximum bootstrap support of 100. Strikingly, the cluster of *S. uniforme* appeared to be nested inside that of *S. croceolum*, with the united cluster of these two species also having the support of 100. *S. pedemontanum* (Müller in Allioni, 1766) and *S. cordulegaster* clustered with a support of 76. At the same time, *S. fonscolombii* showed a very deep divergence from the rest of *Sympetrum*. We may conclude that sequences

of different specimens of a species clustered together with the maximum support or nearly so, and cases of tight clustering of different species corresponded to the notion of their relatedness based on morphology.

It is noteworthy that the sequences of two specimens of *S. croceolum* from its West Siberian isolate (specimens Sc-41 and Sc-42) (Kosterin, 2002) did not show divergence from those from the main Far Eastern range of the species (the rest of the specimens) but were nested among them.

Discussion

The marker proposed

We may conclude that the phylogenetic information provided by the proposed marker well resolved the overall phylogenetic relationships of Odonata at taxonomic levels of families, genera and species (Fig. 2), with few notable exceptions, which could actually reflect weak points of the currently accepted taxonomic system (Dijkstra et al., 2013).

The high conservation of the histone H3 and H4 proteins is paralleled by the high conservation of their coding sequences, the variation of which is nearly confined to synonymous substitutions (Stein et al., 1984; Doenecke et al., 1997; Eirín-López et al., 2009). This allowed us to design primers highly specific to these particular genes but of a very broad applicability to biological objects. It is noteworthy that the substitutions in the histone coding sequence, the overwhelming majority of which are synonymous, provide enough variation for satisfactory resolution of phylogenetic relations between the studied species (Fig. 3). Because of this, the histone H3 gene coding sequence has been broadly used as a phylogenetic marker for short evolutionary distances, e. g. in Odonata by Carle et al. (2015), with conserved positions permitting universal primers whereas the phylogenetic signal being mostly provided by synonymous substitutions.

The marker proposed here, the histone H3–H4 region, has an advantage of possibility to design highly universal primers matching the most conserved eukaryotic coding regions, those of histones H3 and H4. Note that we used the LH4_2R primer designed to match the sequence of a lepidopteran, *B. mori*. The amplified fragment contains most of the coding region of histone H4 and about a half of that of histone H3, and ca 250 b.p. long spacer between them. No significant adaptive constraint is expected for variation of the spacer, except for the origins of transcription of both genes (in opposite directions, with the transcribed sequences not overlapping), which the spacer contains judging from transcriptome data in public databases (not shown). Hence, the spacer enjoys mostly a neutral regime of evolution and may serve as a molecular clock.

The histone H3–H4 region includes both highly conserved coding sequences and a neutrally variable non-coding spacer, and is tandemly repeated; this makes the proposed phylogenetic marker similar in biological and technical respects to such a popular nuclear marker as the ITS region of the nucleolus organiser (for its use in Odonata, see Hovmöller, Johansson, 2004; Dumont et al., 2010; Schneider et al., 2023) including the internal spacers *ITS1* and *ITS2* and the functional 5.8 S rRNA between them. The length of ca 250 b.p. of the spacer in the marker proposed here is comparable to ca 200 b.p. of *ITS1* and ca 160 b.p. of *ITS2* (these figures are for Odonata). Both markers, the *ITS* and the histone H3–H4 region, have comparable lengths (ca 900 b.p.) and suffer from the same drawback of certain ambiguity of alignment because of frequent indels. Both are encoded by the nuclear genome but functionally unrelated. Hence the histone H3–H4 region can be used for the same purposes as the *ITS* region. Moreover, comparison of phylogenetic resolution of Odonata at the family level (Figs 2, 5) showed that the H3–H4 region adequately resolved the phylogeny of the Odonata families (Fig. 2) while the *ITS* region rather failed to do this (Fig. 5), so the use of the former is preferable at this level.

Therefore, the use of the histone H3–H4 region can update the traditional analysis of the *ITS* region with about the same amount of independent phylogenetic information of the same nature. A joint analysis of both similar but unrelated nuclear markers, the ITS and H3–H4 regions (by their concatenation or, better, involving software specially designed for simultaneous analysis of different markers), is expected to provide a more robust phylogenetic inference than the analysis of ITS alone. Judging by the phylogenetic trees obtained (Figs 2, 6), the use of the histone H3–H4 region as a phylogenetic marker is highly recommendable at the levels of species and genera. Since it correctly resolves the family structure of the order, with few exceptions, it could also be useful at the level of families as well, but better as an additional marker analysed together with other phylogenetic markers.

Applicability of the new marker beyond Odonata

Because of conservativeness of the histone H3 and H4 genes, the new marker can be used with the primers provided herein for any Odonata and other insects with the same order and orientation of the histone H3 and H4 genes in the histone repeat. Our investigation of public databases revealed the same order and orientation of these two genes in the histone repeat in a number of species of Diptera. In the present study, they are exemplified by such genera as *Aedes* Meigen, 1818 and *Anopheles* Meigen, 1818, used as the outgroup in the phylogenetic trees of Figures 2–4. The same order was found in *D. melanogaster* (Goodenough, 1984), which represents another suborder of Diptera. Besides, the same was found in the *Formica* Linnaeus, 1758 ants representing Hymenoptera.

For the use of the marker proposed here in insects with other order or orientations of these genes, other relevant primers have to be worked out. For example, in Lepidoptera, where the genes of histones H3 and H4 have parallel orientation, the same LH4-2R primer can be used in combination with a primer of the sequence which is a reverse complement of that of the Hex-AR primer. In this case, the portion of the coding sequence of the H3 gene will be smaller, 37 b.p., and the spacer will be somewhat longer – about 900–1400 b.p.

Histone genes are organised in tandem repeats in a broad range of large groups of organisms such as amphibians, fish, echinoderms, arthropods and annelids (Eirín-López et al., 2009). Among the examples given by Eirín-López et al. (2009), beyond insects, the genes of histones H3 and H4 are adjacent in the histone repeat in the rainbow trout (fish), *Xenopus* spp. (amphibians), starfish (species not indicated, echinoderms), *Asellus aquaticus* (Linnaeus, 1758) (crustacean) and three species from different genera of annelids (Eirín-López et al., 2009: Fig. 8.2). All these groups are potential targets

Fig. 7. Males of *Sympetrum croceolum* (top – Russia, Novosibirsk Academy Town env., 24.08.2023) and *S. uniforme* (bottom – Russia, Primorye, Gornye Klyuchi env., Draguchina Arm of the Ussuri River, 30.07.2020) in nature. Photos by O. Kosterin.

for the use of versions of the marker proposed here based on the coding sequences of the histone H3 and H4 genes and the spacer between them, taking into account orientation of the two genes. Moreover, among the organisms mentioned, the crustacean *A. aquaticus* and the annelids *Platynereis dumerilii* (Audouin, Milne-Edwards, 1834) and *Chaetopterus variopedatus* (Reiner, 1804) have the same orientation of the two genes as Odonata (Eirín-López et al., 2009: Fig. 8.2). Taking into account the high conservativeness of the histone H3 and H4 genes, it is not excluded that the marker proposed here is applicable to these objects with the same primers.

Phylogenetic relationships among *Sympetrum* spp.

The analysis of the new marker yielded two rather unexpected results at once, both concerning the species *S. croceolum* and *S. uniforme* (Fig. 6). First, *S. uniforme* appeared to be an inner branch nested inside *S. croceolum*. This result appeared robust regardless of the methods and models of phylogenetic reconstructions (not shown). These species are obviously related but well distinguishable by the morphology of the male genitalia and female vulvar scale, wing coloration (dull, complete but gradually changing in intensity in *S. uniforme* versus bright but with gaps in *S. croceolum*) (Fig. 7) and size (in the former it is somewhat larger). In East Asia, they usually co-exist in a wide range of lentic habitats (Onishko, Kosterin, 2021), while *S. croceolum* also has an isolated range fragment in the southern West Siberia (Kosterin, 1987; 2002; Popova, Haritonov, 2020). It should be stressed that specimens of *S. uniforme*, identified by external characters, formed a highly supported cluster (Fig. 6). There is no doubt that *S. croceolum* and *S. uniforme* are *bona species*. The phylogenetic pattern obtained, where the sequences of *S. uniforme* are nested inside those of *S. croceolum*, suggests *S. uniforme* being a phyletic descendant of *S. croceolum*, which hence appeared paraphyletic. This contradicts the cladistic approach in systematics and the phylogenetic concept of species. At the same time, this pattern fits well the so-called punctuated equilibria mode of speciation (Eldredge, Gould, 1972), suggesting that speciation takes place for short time periods in evolutionary scale (tens to hundreds of thousand years) in small, isolated populations in the periphery of parental species' ranges, while species exist almost unchanged for millions of years (evolutionary stasis). This concept better fits the basics of evolutionary genetics (Mayr, 1963; Berdnikov, 1999) than the earlier prevailing model of gradual divergent evolution. In the punctuated evoltion point of view, species 'propagate' as if being individuals, with younger species often co-existing with their parental species.

In the phylogenetic tree based on the histone H3–H4 region (Fig. 6), two analyzed specimens of *S. croceolum* from its West Siberian isolate (specimens Sc-41 and Sc-42 from Lake Manzherok) lacked supported divergence from specimens from the main Far Eastern range of the species (the rest of the specimens), which is quite remarkable. Specimens from the West Siberian isolate (Kosterin, 2002) differ from the Far Eastern specimens by a much more developed wing amber colour and an appearance of a brown infumation in the wing apical parts and so were for a long time supposed to represent a separate subspecies (Kosterin, 1987; Popova, Haritonov, 2020), which, however, has not been named yet. This lack of divergence at the molecular level suggests the West Siberian isolate to be very young in evolutionary time scale and well fits its hypothetic Holocene age, implying the range split after Atlantic time (Kosterin, 2002), as well as in some nemoral species of Lepidoptera (Dubatolov, Kosterin, 2000; Solovyev et al., 2015, 2022). This, however, does not exclude a subspecies rank of the West Siberian population(s), since subspecies are entities of well-defined geographical variation for some phenotype characters, which implies specific divergence of the genes determining these characters rather than those of the entire genome.

In all phylogenetic reconstructions from sequences of the histone H3–H4 region (Figs 2, 6), *S. fonscolombii* is strongly diverged from the rest of the genus *Sympetrum*. The same result was earlier obtained by Pilgrim and von Dohlen (2012) who undertook a molecular phylogenetic study of *Sympetrum* and related genera based on the joint analysis of the nuclear marker *EF-1α* and *ITS2*, and the mitochondrial genes *16S*, *tRNA valine*, *12S*, and *CO*I. This divergent position of *S. fonscolombii* has long ago been recognised at the level of phenotype, resulting in a suggestion to move *S. fonscolombii* to

the genus *Tarnetrum* Needham and Fischer, 1936 (Schmidt, 1987). This genus was erected for two Nearctic species, *Mesothemis corrupta* Hagen 1861, and *M. illota* Hagen, 1861 (mentioned in (Schmidt, 1987) as well as presently considered in combinations *S. illotum* and *S. corruptum*), with *M. illota* (sub. *S. illotum*) indicated as the type species (Needham, Fischer, 1936). However, according to Pilgrim and von Dohlen (2012), *S. corruptum* is quite closely related to *S. fonscolombii* (together with *S. villosum* Ris, 1911 and *Nesogonia blackburni* (McLachlan, 1883)) while *S. illotum* is not. Since the type species of the genus *Tarnetrum* is not closely related to *S. fonscolombii*, this genus is not suitable for the latter species (Dijkstra, Kalkman, 2015). Therefore, the genus *Sympetrum* in the current sense deserves further reconsideration, maybe with erection of a new genus at least for the *fonscolombii*-group *sensu* Pilgrim et al. (2012).

References

Avise J.C. Phylogeography: retrospect and prospect. *J Biogeogr.* 2009; 36:3-15. doi 10.1111/j.1365-2699.2008.02032.x

Ballard J.W.O., Whitlock M.C. The incomplete natural history of mitochondria. *Mol Ecol.* 2004;13:729-744. doi 10.1046/j.1365-294x.2003.02063.x

Berdnikov V.A. Evolution and Progress. Sofia; Moscow: Pensoft, 1999

Bybee S.M., Kalkman V.J., Erickson R.J., Frandsen P.B., Breinholt J.W., Suvorov A., Dijkstra K.B., … Abbott J.C., Sanchez Herrera M., Lemmon A.R., Moriarty Lemmon E., Ware J.L. Phylogeny and classification of Odonata using targeted genomics. *Mol Phyl Evol.* 2021;160:107115. doi 10.1016/j.ympev.2021.107115

Carle F.L., Kjer K.M., May M.L. A molecular phylogeny and classification of Anisoptera. *Arthropod Syst Phylo.* 2015;73(2):281-301. doi 10.3897/asp.73.e31805

Cheng Y.-C., Chen M.-Y., Wang J.-F., Liang A.P., Lin C.-P. Some mitochondrial genes perform better for damselfly phylogenetics: species- and population-level analyses of four complete mitogenomes of *Euphaea* sibling species. *Syst Entomol.* 2018;43(4):702-715. doi 10.1111/syen.12299

Cheng Z., Li Q., Deng J., Liu Q., Huang X. The devil is in the details: Problems in DNA barcoding practices indicated by systematic evaluation of insect barcodes. *Front Ecol Evol.* 2023;11:1149839. doi 10.3389/fevo.2023.1149839

Chevreux B., Wetter T., Suhai S. Genome sequence assembly using trace signals and additional sequence information. In: Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB). 1999;99:45-56

Deng J., Assandri G., Chauhan P., Futahashi R., Galimberti A., Hansson B., Lancaster L.T., Takahashi Y., Svensson E.I., Douploy A. *Wolbachia*-driven selective sweep in a range expanding insect species. *BMC Ecol Evol.* 2021;21:181. doi 10.1186/s12862-021-01906-6

Dijkstra K.D.B., Bechly G., Bybee S.M., Dow R.A., Dumont H.J., Fleck G., Garrison R.W., … Theischinger G., Trueman J.W.H., Van Tol J., Von Ellenrieder N., Ware J. The classification and diversity of dragonflies and damselflies (Odonata). *Zootaxa*. 2013;3703(1): 36-45. doi 10.11646/zootaxa.3703.1.9

Dijkstra J.P., Kalkman V. Taxonomy. In: Boudot J.P., Kalkman V.J. (Eds). Atlas of the European dragonflies and damselflies. The Netherlands: KNNNV Publishing, 2015;15-25

Doenecke D., Albig V., Bode C., Drabent B., Franke K., Gavenis K., Witt O. Histones: genetic diversity and tissue-specific gene expression. *Histochem Cell Biol.* 1997;107:1-10. doi 10.1007/s004180050083

Dow R.A., Butler S.G., Reels G.T., Steinhoff O.M., Stokvis F., Unggang L. Previously unpublished Odonata records from Sarawak, Borneo, part IV: Bintulu Division including the Planted Forest Project and Similajau National Park. *J Int Dragonfly Fund.* 2019;27: 1-66

Dubatolov V.V., Kosterin O.E. Nemoral species of Lepidoptera (Insecta) in Siberia: a novel view on their history and the timing of their range disjunctions. *Entomol Fennica.* 2000;11(3):141-166. doi 10.33338/ef.84061

Dumont H.J., Vierstraete A., Vanfleteren J.R. Molecular phylogeny of the Odonata (Insecta). *Syst Entomol.* 2010;35(1):6-18. doi 10.1111/j.1365-3113.2009.00489.x

Eirín-López J.M., González-Romero R., Dryhurst D., Méndez J., Ausio J. Long-term evolution of histone families: old notions and new insights into their mechanisms of diversification across eukaryotes. In: Pontarotti P. (Ed.). Evolutionary Biology: Concept, Modeling, and Application. Heidelberg: Springer, 2009;139-162. doi 10.1007/978-3-642-00952-5_8

Eldredge N., Gould S.L. Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf T.J.M. (Ed.). Models in Palaeobiology. San Francisco: Freeman, Cooper & Co., 1972;82-115

Ferreira S., Lorenzo-Carballa M.O., Torres-Cambas Y., Cordero-Rivera A., Thompson D.J., Watts P.C. New EPIC nuclear DNA sequence markers to improve the resolution of phylogeographic studies of coenagrionids and other odonates. *Int J Odonatol.* 2014;17(2-3): 135-147. doi 10.1080/13887890.2014.950698

Ferreira S., Boudot J.-P., El Haissoufi M., Alves P.C., Thompson D.J., Watts P.C. Genetic distinctiveness of the damselfly *Coenagrion puella* in North Africa: an overlooked and endangered taxon. *Conserv Genet.* 2016;17:985-991. doi 10.1007/s10592-016-0826-5

Futahashi R., Kawahara-Miki R., Kinoshita M., Yoshitake K., Yajima S., Arikawa K., Fukatsu T. Extraordinary diversity of visual opsin genes in dragonflies. *Proc Natl Acad Sci USA.* 2015;112(11): E1247-E1256. doi 10.1073/pnas.1424670112

Galimberti A., Assandri G., Maggioni D., Ramazotti F., Baroni D., Bazzi G., Chiandetti I., … Ramellini S., Santinelli R., Soldato G., Surdo S., Casiraghi M. Italian odonates in the Pandora's box: a comprehensive DNA barcoding inventory shows taxonomic warnings at the Holarctic scale. *Mol Ecol Resour.* 2021;21(1):183-200. doi 10.1111/1755-0998.13235

Geiger M., Koblmüller S., Assandri G., Chovanec A., Ekrem T., Fischer I., Galimberti A., … Sittenthaler M., Stur E., Tończyk G., Zangl L., Moriniere J. Coverage and quality of DNA barcode references for Central and Northern European Odonata. *PeerJ.* 2021; 9:e11192. doi 10.7717/peerj.11192

Goodenough U. Genetics. Hong Kong: CBS College Publishing, 1984

Gurdon C., Svab Z., Feng Y., Kumar D., Maliga P. Cell-to-cell movement of mitochondria in plants. *Proc Natl Acad Sci USA.* 2016; 113(12):3395-3400. doi 10.1073/pnas.1518644113

Hovmöller R., Johansson F. A phylogenetic perspective on larval spine morphology in *Leucorrhinia* (Odonata: Libellulidae) based on ITS1, 5.8s and ITS2 rDNA sequences. *Mol Phyl Evol.* 2004;30(3): 653-662. doi 10.1016/S1055-7903(03)00226-4

Karube H., Futahashi R., Sasamoto A., Kawashima I. Taxonomic revision of Japanese Odonata species, based on nuclear and and mitochondrial gene genealogies and morphological comparison with allied species. Part I. *Tombo.* 2012;54:75-106

Kohli M., Letsch H., Greve C., Bethoux O., Deregnaucourt I., Liu S., Zhou X., … Rust J., Wappler T., Yu X., Misof B., Ware J. Evolutionary history and divergence times of Odonata (dragonflies and damselflies) revealed through transcriptomics. *iScience* 2021;24(11): 103324. doi 10.1016/j.isci.2021.103324

Kosterin O.E. Discovery of East-Asiatic dragonfly (Odonata, Libellulidae) at the Manzherock Lake (Altay). In: Cherepanov A.I. (Ed.). Nasekomye, kleshchi i gel'minty. Novye i maloizvestnye vidy fauny Sibiri. Novosibirsk: Nauka, 1987;57-63 [in Russian]

Kosterin O.E. Western range limits and isolates of eastern odonate species in Siberia and their putative origins. *Odonatologica.* 2002; 34(3):219-242

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947-2948. doi 10.1093/bioinformatics/btm404

Lorenzo-Carballa M.O., Sanmartín-Villar I., Cordero-Rivera A. Molecular and morphological analyses support different taxonomic units for Asian and Australo-Pacific forms of *Ischnura aurora* (Odonata, Coenagrionidae). *Diversity.* 2022;14(8):606. doi 10.3390/d14080606

Lowe C.D., Harvey I.F., Thompson D.J., Watts P.C. Strong genetic divergence indicates that congeneric damselflies *Coenagrion puella* and *C. pulchellum* (Odonata: Zygoptera: Coenagrionidae) do not hybridise. *Hydrobiologia.* 2008;605:55-63. doi 10.1007/s10750-008-9300-9

Mayr E. Zoological species and evolution. Cambridge, Massachusetts: The Belknap Press of Harward University Press, 1963

Needham J.G., Fischer E. The nymphs of North American libelluline dragonflies. *Trans Am Entomol Soc.* 1936;62(2):107-116

Onishko V., Kosterin O. Dragonflies of Russia. Illustrated Photo Guide. Moscow: Phyton XXI, 2021 [in Russian]

Ožana S., Dolný A., Pánek T. Nuclear copies of mitochondrial DNA as a potential problem for phylogenetic and population genetic studies of Odonata. *Syst Entomol.* 2022;47(4):591-602. doi 10.1111/syen.12550

Pilgrim E.M., von Dohlen C.D. Phylogeny of the dragonfly genus *Sympetrum* (Odonata: Libellulidae). *Org Divers Evol.* 2012;12:281-295. doi 10.1007/s13127-012-0081-7

Popova O.N., Haritonov A.Y. On the distribution of *Sympetrum croceolum* in the Russian part of its range (Odonata: Libellulidae). *Odonatologica.* 2020;49(1/2):29-49. doi 10.5281/zenodo.3823325

Schmidt E. Generic reclassification of some Westpalearctic Odonata taxa in view of their Nearctic affinities. *Adv Odonatol.* 1987;3(1):135-145

Schneider T., Vierstraete A., Kosterin O.E, Ikemeyer D., Hu F.-S., Snegovaya N., Dumont H.J. Molecular phylogeny of Holarctic Aeshnidae with a focus on the West Palaearctic and some remarks on its genera worldwide (Aeshnidae, Odonata). *Diversity.* 2023;15(9):950. doi 10.3390/d15090950

Solovyev V.I., Bogdanova V.S., Dubatolov V.V., Kosterin O.E. Range of a Palearctic uraniid moth *Eversmannia exornata* (Lepidoptera: Uraniidae: Epipleminae) was split in the Holocene, as evaluated using histone H1 and COI genes with reference to the Beringian disjunction in the genus *Oreta* (Lepidoptera: Drepanidae). *Org Divers Evol.* 2015;15(2):285-300. doi 10.1007/s13127-014-0195-1

Solovyev V.I., Dubatolov V.V., Vavilova V.Y., Kosterin O.E. Estimating range disjunction time of the Palaearctic Admirals (*Limenitis* L.) with *COI* and histone H1 genes. *Org Divers Evol.* 2022;22:975-1002. doi 10.1007/s13127-022-00565-9

Staden R., Judge D.P., Bonfield J.K. Managing sequencing projects in the GAP4 environment. In: Krawetz S.A., Womble D.D. (Eds). Introduction to Bioinformatics. Human Press Inc., 2023;327-224. doi 10.1007/978-1-59259-335-4_20

Stein G.S., Stein J.L., Marzluff W.F. (Eds). Histone Genes: Structure, Organization, and Regulation. New York: Willey, 1984

Tamura K., Stecher G., Peterson D., Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30:2725-2729. doi 10.1093/molbev/mst197

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

Received March 19, 2025. Revised April 24, 2025. Accepted April 24, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):883-895

doi 10.18699/vjgb-25-95

Complete plastome sequences of *Lonicera* L. species: implications for phylogeny and comparative analysis

S.S. Almerekova [1], M.M. Yermagambetova [1], D.Y. Yerbolatov [1], M.Y. Ishmuratova [2], Y.K. Turuspekov [1, 2]

[1] Institute of Plant Biology and Biotechnology, Almaty, Kazakhstan
[2] Karaganda Buketov University, Karaganda, Kazakhstan
yerlant@yahoo.com

Abstract. *Lonicera* L. is one of the largest and economically significant genera in the family Caprifoliaceae Juss., with a controversial taxonomy. To contribute to its molecular taxonomy, we sequenced the plastomes of *Lonicera* species: *Lonicera caerulea* (two subspecies), *L. tatarica*, and *L. micrantha* – using next-generation sequencing technology and conducted a comparative analysis. Plastome sizes ranged from 153,985 bp in *L. micrantha* to 164,000 bp in *L. caerulea* subsp. *pallasii*, each containing 130 genes, including 85 protein-coding, 37 tRNA, and 8 rRNA genes. Five protein-coding (*rps7*, *rps12*, *ndhB*, *ycf2*, and *ycf15*), 7 tRNA (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), and 4 rRNA (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*) genes were duplicated. Comparative analysis of *Lonicera* plastome boundaries revealed structural variations in *L. caerulea* subsp. *altaica* and *L. caerulea* subsp. *pallasii*, particularly in *ndhA* gene distribution. Three highly variable, two intergenic (*ycf1-trnN-GUU* and *trnN-GUU-ndhF*) and one genic (*accD*) region were identified. A total of 641 simple sequence repeats were detected in four plastomes. Phylogenetic analyses grouped *Lonicera* samples into two clades corresponding to subgenera *Periclymenum* and *Chamaecerasus*. In this study, the plastid genomes of two subspecies of *L. caerulea* and species *L. micrantha* were sequenced for the first time. The maximum likelihood tree derived from complete plastid genome sequences proved to be the most informative, showing a topology consistent with previous studies. The nucleotide sequences of variable regions (*accD-ycf1-ndhF-trnN-GUU*) demonstrate high potential for use in DNA barcoding and may serve as valuable molecular markers for species phylogenetic studies within the genus *Lonicera*.

Key words: *Lonicera*; Kazakhstan; next-generation sequencing; variable regions; DNA-barcoding markers; simple sequence repeats

For citation: Almerekova S.S., Yermagambetova M.M., Yerbolatov D.Y., Ishmuratova M.Y., Turuspekov Y.K. Complete plastome sequences of *Lonicera* L. species: implications for phylogeny and comparative analysis. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(6):883-895. doi 10.18699/vjgb-25-95

Funding. This research has been funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR21882166).

Полные последовательности пластомов видов *Lonicera* L.: значение для филогении и сравнительный анализ

Ш.С. Альмерекова [1], М.М. Ермагамбетова [1], Д.Е. Ерболатов [1], М.Ю. Ишмуратова [2], Е.К. Туруспеков [1, 2]

[1] Институт биологии и биотехнологии растений, Алматы, Казахстан
[2] Карагандинский университет им. академика Е.А. Букетова, Караганда, Казахстан
yerlant@yahoo.com

Аннотация. Род *Lonicera* L. – один из крупнейших и экономически значимых в семействе Caprifoliaceae Juss., таксономия которого остается спорной. С целью внесения вклада в молекулярную таксономию данного рода мы секвенировали пластомы видов *Lonicera* – *L. caerulea* (два подвида), *L. tatarica* и *L. micrantha* – с использованием технологии секвенирования нового поколения и провели сравнительный анализ. Размеры пластомов варьировали от 153 985 п. н. у *L. micrantha* до 164 000 п. н. у *L. caerulea* subsp. *pallasii*; каждый пластом содержал 130 генов, включая 85 генов, кодирующих белок, 37 тРНК и 8 рРНК генов. Пять белок-кодирующих (*rps7*, *rps12*, *ndhB*, *ycf2* и *ycf15*), 7 тРНК (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG* и *trnV-GAC*) и 4 рРНК (*rrn4.5*, *rrn5*, *rrn16* и *rrn23*) гена были дуплицированы. Сравнительный анализ границ пластомов у видов *Lonicera* выявил структурные вариации у *L. caerulea* subsp. *altaica* и *L. caerulea* subsp. *pallasii*, особенно в распределении гена *ndhA*. Были идентифицированы три высоковариабельные области: две межгенные (*ycf1-trnN-GUU* и *trnN-GUU-ndhF*) и одна генная (*accD*). В четырех пластомах обнаружено всего 641 простая повторяющаяся последовательность (SSR). Филогенетический анализ сгруппировал образцы *Lonicera* в два клада, соответствующих подродам *Periclymenum* и *Chamaecerasus*. В настоящем исследовании впервые секвенированы пластидные геномы двух

подвидов *L. caerulea* и вида *L. micrantha*. Древо, построенное методом maximum likelihood на основе полных последовательностей пластомов, оказалось наиболее информативным и демонстрировало топологию, согласующуюся с предыдущими исследованиями. Нуклеотидные последовательности вариабельных регионов (*accD-ycf1-ndhF-trnN-GUU*) обладают высоким потенциалом для использования в ДНК-баркодировании и могут служить ценными молекулярными маркерами для филогенетических исследований видов внутри рода *Lonicera*.

Ключевые слова: *Lonicera*; Казахстан; секвенирование нового поколения; вариабельные регионы; маркеры ДНК-баркодирования; простые повторяющиеся последовательности

Introduction

Lonicera L. is the largest genus in the family Caprifoliaceae Juss., comprising approximately 140 species (Wang G.Q. et al., 2024), which are widely distributed across North America, Europe, Asia, and Africa (Donoghue et al., 2001; Wen, 2001). *Lonicera* species exhibit diverse constituents, including saponins, flavonoids, iridoids, phenolic acids, alkaloidal glycosides, etc. (Lin et al., 2008; Ali et al., 2013; Yang Q.R. et al., 2016; Ni, 2017). Moreover, it exhibits a range of biological activities, including antioxidant, anti-inflammatory, antiviral, anti-hepatoma, and hepatoprotective effects (Yoo et al., 2008; Park et al., 2012; Kong et al., 2017; Ge et al., 2018; Liu M. et al., 2020). Besides their biological activities, species of the *Lonicera* genus also hold significant ornamental value and are widely used in landscaping (Hayes, Peterson, 2020; Varlashchenko et al., 2021).

In Kazakhstan, the genus is represented by 22 species (Abdulina, 1999), two of which are listed in the Red Book of Kazakhstan (Baitulin, Sitpayeva, 2014). The species *Lonicera caerulea* subsp. *altaica* (Pall.) Gladkova, *L. caerulea* subsp. *pallasii* (Ledeb.) Browicz, *L. tatarica* L., and *L. micrantha* Trautv. ex Regel are widely distributed across Kazakhstan. According to the Plants of the World Online (https://powo.science.kew.org/) *L. caerulea* subsp. *altaica* is native to a vast range extending from Eastern Europe to Siberia and Mongolia, while *L. caerulea* subsp. *pallasii* is found in Northern and Eastern Europe, Siberia, and Central Asia. *L. tatarica* occurs naturally from Eastern Europe to Central Siberia and northeastern China; in contrast, *L. micrantha* is native to Kazakhstan. These species play a crucial role in the region's floral biodiversity and are of particular ecological and conservation significance. Additionally, they possess medicinal properties and have been traditionally used in folk medicine for various therapeutic applications (Golubev et al., 2022; Boyarskikh, Kostikova, 2023; Taldybay et al., 2024). In Kazakhstan, *Lonicera* species have been studied using botanical (Ametov et al., 2016; Vdovina et al., 2024), phytochemical (Kushnarenko et al., 2016), and biochemical (Vdovina, 2019) assessments.

The phylogenetic relationships within *Lonicera* remain incompletely resolved, presenting ongoing systematic challenges and requiring revisions to its classification (Wang G.Q. et al., 2024). Over time, various classification systems for *Lonicera* have been proposed (Maximowicz, 1877; Rehder, 1903; Nakai, 1938; Hara, 1983). According to A. Rehder (1903), *Lonicera* is divided into two subgenera, *Chamaecerasus* (L.) Rehder and *Periclymenum* (Mill.) Rehder; within the subgenus *Chamaecerasus*, it is further classified into four sections: *Isoxylosteum* Rehder, *Isika* DC., *Coeloxylosteum* Rehder, and *Nintooa* DC. H. Hara (1983) proposed a classification, which was based on the study by C.J. Maximowicz (1877), dividing *Lonicera* into subgenera *Lonicera* and *Caprifolium* (Mill.) Dippel, with further subdivision of subgenus *Lonicera* into four sections (*Isika* (Anderson) Rehder, *Caeruleae* (Rehder) Nakai, *Lonicera* and *Nintooa* (Sweet) Maxim) and five subsections (*Purpurascentes*, *Monanthae*, *Isika*, *Bracteatae*, and *Rhodanthae*). Later, P.S. Hsu et al. (1988) classified *Lonicera* into subgenera *Chamaecerasus* and *Lonicera*; further, subgenus *Chamaecerasus* was divided into four sections (*Coeloxylosteum*, *Isika*, *Isoxylosteum*, and *Nintooa*).

With advancements in molecular genetic technologies, numerous studies have focused on the phylogenetics of *Lonicera*. For instance, M. Nakaji et al. (2015) investigated the phylogenetic relationships among 23 Japanese *Lonicera* species using nucleotide sequences of five plastid non-coding regions (*rpoB-trnC*, *atpB-rbcL*, *trnS-trnG*, *petN-psbM*, and *psbM-trnD*). The findings support the fundamental validity of the classification by H. Hara (1983) of higher taxonomic groups for Japanese *Lonicera* species. M. Srivastav et al. (2023) conducted a phylogenetic analysis using restriction site-associated DNA sequencing (RADSeq). The RADSeq-based phylogenetic tree revealed that the *Coeloxylosteum*, *Isoxylosteum*, and *Nintooa* sections within subgenus *Chamaecerasus* were monophyletic, whereas the *Isika* section was found to be paraphyletic. Using nuclear ribosomal DNA cistron and plastid genome data, X.L. Yang et al. (2024) confirmed the paraphyly of section *Isika* and the monophyly of sections *Coeloxylosteum*, *Isoxylosteum*, and *Nintooa* within subgenus *Chamaecerasus*, aligning with the findings of M. Srivastav et al. (2023). All of the above-mentioned studies have contributed to the classification of the genus *Lonicera*. However, due to widespread hybridization and introgression, the precise taxonomy of *Lonicera* remains unresolved (Wang H.X. et al., 2020).

The plastid is a vital organelle for photosynthesis in plants and possesses its own genome (Howe et al., 2003). The plastome is uniparentally inherited and highly conserved in gene content and organization (Howe et al., 2003). It ranges in size from approximately 120 to 160 kb and exhibits a quadripartite structure consisting of two identical inverted repeats (IR) and two single-copy regions: a large single-copy (LSC) region and a small single-copy (SSC) region (Palmer et al., 1988; Ruhlman, Jansen, 2014).

Advancements in high-throughput sequencing technologies have greatly facilitated plastid genome research, making it more accessible and enabling comprehensive genomic analyses. To date, only a few studies have been conducted on the comparative analysis of *Lonicera* plastid genomes. For example, seven plastid genomes (*L. ferdinandi*, *L. hispida*, *L. nervosa*, *L. fragrantissima* var. *lancifolia*, *L. stephanocarpa*, *L. tragophylla*, and *L. japonica*) (Liu M.L. et al., 2018) and three plastid genomes (*L. japonica*, *L. similis*, and *L. acumi-*

nata) (Yang C. et al., 2023) have been comparatively analyzed. Recent studies have demonstrated that nucleotide sequences of the plastid genome can provide valuable insights for phylogenomic analysis (Luo et al., 2021; Zhao et al., 2023) taxonomic classification (Li Q., 2022; Oyuntsetseg et al., 2024), and species identification, utilizing plastome sequences as a super barcode (Chen X. et al., 2018; Zhang Z. et al., 2019). Moreover, plastid genome nucleotide sequences serve as a valuable resource for identifying species-specific genetic markers, such as DNA barcoding (Hong et al., 2022; Tang et al., 2022; Almerekova et al., 2024), microsatellite (Zhu M. et al., 2021; Yermagambetova et al., 2023), and single nucleotide polymorphism (SNP) markers (Dong et al., 2021). Therefore, we believe that sequencing and comparative analysis of *Lonicera* plastomes can contribute insights into the taxonomic classification and phylogenetic relationships of the genus.

In this study, we sequenced the plastid genomes of *Lonicera* species (*L. caerulea* subsp. *altaica*, *L. caerulea* subsp. *pallasii*, *L. tatarica*, and *L. micrantha*) using Illumina Next Generation Sequencing technology. Among them, *L. caerulea* subsp. *altaica*, *L.caerulea* subsp. *pallasii*, and *L. micrantha* have been sequenced for the first time to date. We conducted a plastome-based analysis to characterize the plastomes of the selected *Lonicera* species. Our analysis included comparative plastome assessments with previously sequenced *Lonicera* species from GenBank, identification of potential molecular markers valuable for DNA barcoding and population genetics, and evaluation of the taxonomic positions of the studied *Lonicera* species.

Materials and methods

Plant leaf material collection and DNA extraction. Fresh leaf samples were collected from the eastern and central parts of Kazakhstan. Detailed information on the collection sites is provided in Table 1. The leaves were dried in silica gel and subsequently used for DNA extraction. Genomic DNA was extracted from the dried *Lonicera* leaves using the cetyltrimethylammonium bromide (CTAB) method (Doyle J.J., Doyle J.L., 1987). The quality and concentration of the extracted DNA were assessed using a NanoDrop™ One spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

Plastid genome sequencing, assembly, and annotation. DNA samples that passed Quality Control (QC) analysis were used for subsequent library preparation. The libraries were constructed using the TruSeq Nano DNA Kit (Illumina Inc., San Diego, CA, USA). Plastid genome sequencing was performed on the Illumina NovaSeq 6000 (Illumina Inc.) platform at Macrogen Inc. (Seoul, Republic of Korea). The quality of raw sequencing data was assessed using FastQC 0.11.7 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc, accessed on 02 December 2024). The adapter sequences were removed from the raw reads using Trimmomatic 0.38 (Bolger et al., 2014). *De novo* assembly was conducted using NOVOplasty (Dierckxsens et al., 2017), and plastome annotation was performed using the Plastid Genome Annotator (PGA) (Qu et al., 2019). Gene maps of the annotated plastid genomes of *L. caerulea* subsp. *altaica*, *L.caerulea* subsp. *pallasii*, *L. tatarica*, and *L. micrantha* were drawn using the OrganellarGenomeDRAW tool 1.3.1 (OGDRAW) (Lohse et al., 2007). The newly sequenced plastomes of these species have been deposited in GenBank under the accession numbers PV026015-PV026018.

Table 1. Information on the collection sites of studied *Lonicera* species in Kazakhstan

Species	Collection sites
L. caerulea subsp. *altaica*	Kazakhstan, East Kazakhstan region, Western Altai, foothills of the Ivanovsky ridge
L. caerulea subsp. *pallasii*	Kazakhstan, Karaganda region, Karkaraly district, Karkaraly Mountains
L. tatarica	Kazakhstan, Karaganda region, Bukhar-Zhyrau district
L. micrantha	Kazakhstan, Karaganda region, Karkaraly district, Karkaraly Mountains

Comparative plastome analysis. Comparative plastome analysis of the studied *Lonicera* species was conducted using mVISTA (Frazer et al., 2004) in Shuffle-LAGAN mode, with the plastid genomes of *L. caerulea* (OQ784224) and *L. tatarica* (OQ784187) serving as references. Additionally, the junction sites of the four *Lonicera* plastomes were examined using the IRscope online tool (Amiryousefi et al., 2018), utilizing the same reference genomes, *L. caerulea* (OQ784224) and *L. tatarica* (OQ784187).

Nucleotide variability analysis. The complete plastid genome sequences of the *Lonicera* species were aligned using Geneious Prime® 2025.0.3 (https://www.geneious.com, accessed on 10 February 2025). The aligned sequences were then analyzed for nucleotide variability (*Pi*) using a sliding window approach in DnaSP v6 (Rozas et al., 2003). The sliding window analysis was performed with a window length of 600 bp and a step size of 200 bp.

Simple sequence repeats analysis and comparative genome analysis. Simple sequence repeats (SSRs) in the nucleotide sequences of the four studied *Lonicera* plastomes were identified using MISA software (Beier et al., 2017). The detection thresholds were set as follows: eight repeats for mononucleotide SSRs, four repeats for di- and trinucleotide SSRs, and three repeats for tetra-, penta-, and hexanucleotide SSRs.

Phylogenetic analysis. Phylogenetic analysis was conducted using alignments of complete plastid genome sequences, protein-coding gene sequences, and variable region gene sequences from *L. caerulea* subsp. *altaica*, *L. caerulea* subsp. *pallasii*, *L. tatarica*, and *L. micrantha*, along with GenBank samples, including outgroup species (*Heptacodium miconioides* and *Triosteum himalayanum*). A total of 24 complete plastid genomes were selected to construct phylogenetic trees in order to determine the phylogenetic placement of the studied species within the genus *Lonicera*. The sequence alignment of the complete plastid genomes was conducted using Geneious Prime® 2025.0.3 (https://www.geneious.com, accessed on 12 February 2025). Phylogenetic relationships were inferred using the maximum likelihood (ML) and Bayesian inference (BI) methods. Maximum likelihood trees were generated using IQ-TREE 2.2.2.6 (Nguyen et al., 2015). The software was also used to determine the optimal tree-building model,

identified as GTR+F+I+R2 for complete plastid genome and variable region genes data, and as TVM+F+I+R3 for protein-coding genes data, which were then applied to reconstruct the ML phylogenetic tree. BI phylogenetic trees were reconstructed using MrBayes 3.2.7 (Ronquist et al., 2012). The resulting phylogenetic trees were visualized using FigTree (Rambaut, 2009). The network analysis was performed in SplitsTree4 (Huson, Bryant, 2006) with the Neighbor-Net algorithm.

Results

General features of the four *Lonicera* plastomes

Illumina sequencing generated paired-end reads with an average length of 150 bp for the four *Lonicera* plastomes. The lengths of the plastid genomes of *L. caerulea* subsp. *altaica*, *L. caerulea* subsp. *pallasii*, *L. tatarica*, and *L. micrantha* were 163,889; 164,000; 154,587, and 153,985 bp, respectively (Fig. 1).

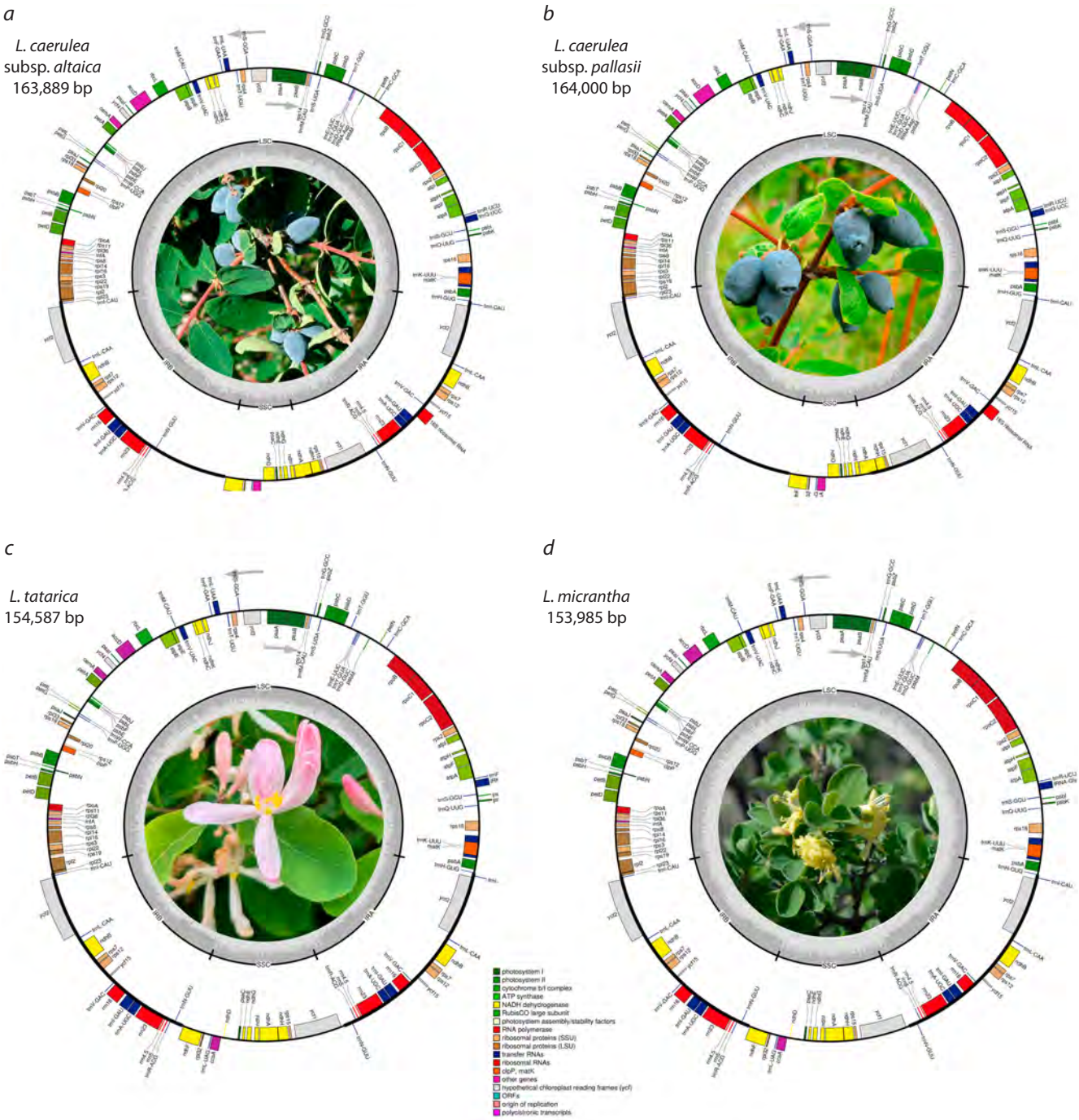


Fig. 1. Plastid genome maps of *L. caerulea* subsp. *altaica* (*a*), *L. caerulea* subsp. *pallasii* (*b*), *L. tatarica* (*c*) and *L. micrantha* (*d*) species.

Genes positioned outside the outer circle are transcribed in a counterclockwise direction, while those inside the circle are transcribed in a clockwise direction. The inner circle represents GC and AT content, with darker gray indicating GC content and lighter gray representing AT content. Genes are color-coded according to their functional categories. The plastid genome map displays a large single-copy (LSC) region, small single-copy (SSC) region, and inverted repeat regions (IRA and IRB).

Table 2. General characteristics of the plastid genomes of the studied *Lonicera* species

Genome features	*L. caerulea* subsp. *altaica*	*L. caerulea* subsp. *pallasii*	*L. tatarica*	*L. micrantha*
GenBank numbers	PV026015	PV026016	PV026017	PV026018
Genome size (bp)	163,889	164,000	154,587	153,985
LSC	88,119	88,813	88,185	88,040
SSC	10,172	10,169	18,750	18,589
IRA	32,799	32,509	23,826	23,678
IRB	32,799	32,509	23,826	23,678
Number of total genes (unique)	130 (115)	130 (115)	130 (115)	130 (115)
Protein genes	85 (80)	85 (80)	85 (80)	85 (80)
tRNA genes	37 (30)	37 (30)	37 (30)	37 (30)
rRNA genes	8 (4)	8 (4)	8 (4)	8 (4)
GC content (%)	38.05	38.12	38.42	38.40
in LSC	36.95	36.92	36.88	36.83
in SSC	32.90	32.84	32.94	33.06
in IRA	40.34	40.59	43.44	43.43
in IRB	40.34	40.59	43.44	43.43

The plastid genome structure consisted of a large single-copy (LSC) region, ranging from 88,040 bp in *L. micrantha* to 88,813 bp in *L. caerulea* subsp. *pallasii*, a small single-copy (SSC) region varying from 10,172 bp in *L. caerulea* subsp. *altaica* to 18,750 bp in *L. tatarica*, and an inverted repeat (IR) region, spanning from 47,356 bp in *L. micrantha* to 65,598 bp in *L. caerulea* subsp. *altaica*. The two inverted repeat regions were designated as IRA and IRB. The total GC content of the four *Lonicera* plastid genomes was relatively consistent, ranging from 38.05 % in *L. caerulea* subsp. *altaica* to 38.42 % in *L. tatarica* plastome. The IR regions exhibited higher GC content (40.34–43.44 %) compared to the single-copy regions, with the LSC region ranging from 36.83 to 36.95 % and the SSC region from 32.84 to 33.06 % (Table 2).

The four assembled plastid genomes of *Lonicera* exhibited identical gene content, intron numbers, and gene order. The plastid genomes of the studied *Lonicera* species comprised 130 genes, including 85 protein-coding genes, 37 tRNA genes, and eight rRNA genes. Among them, five protein-coding genes (*rps7*, *rps12*, *ndhB*, *ycf2*, and *ycf15*), seven tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), and four rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*) were duplicated within the IR regions of the four *Lonicera* plastid genomes. A total of 17 genes contained introns, of which 16 genes (*rps12*, *rps16*, *rpl2*, *rpl16*, *rpoC1*, *atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, and *trnV-UAC*) had a single intron, while *ycf3* was the only gene containing two introns. The *rps12* gene exhibited trans-splicing, with its 5′ end located in the LSC region, while its 3′ end was positioned in the IR regions (Fig. 1, Table 3).

Comparative analysis of the four *Lonicera* plastomes

A comparative analysis of the complete plastid genomes of six *Lonicera* species was conducted using mVISTA, with *L. caerulea* (OQ784224) and *L. tatarica* (OQ784187) as reference genomes. The alignment revealed high sequence conservation across the plastomes, with most variations occurring in non-coding regions. Among the coding regions, *accD* exhibited the highest level of divergence. The IR regions were the most conserved, while the LSC and SSC regions displayed higher levels of sequence divergence (Fig. 2).

Inverted repeat expansion and contraction

A comparative analysis of the LSC/IRB/SSC/IRA boundary regions was conducted in the plastomes of *Lonicera* species (*L. caerulea* subsp. *altaica*, *L. caerulea* subsp. *pallasii*, *L. tatarica*, and *L. micrantha*), using *L. caerulea* (OQ784224) and *L. tatarica* (OQ784187) from GenBank as reference sequences. There were structural differences in LSC/IRB/SSC/IRA boundaries of *Lonicera* plastomes. The length of the IR regions ranged from 23,678 to 32,799 bp in four studied *Lonicera* plastomes with some expansion. A notable difference was found in *L. caerulea* subsp. *altaica* and *L. caerulea* subsp. *pallasii* plastomes, where the gene *ndhA*, which crossed over the IRA/SSC boundaries, was similar to those in GenBank (*L. caerulea*). The *ycf1* gene's distance from the IRA region was 246 and 268 bp in *L. tatarica*, and *L. micrantha*, respectively. At the IRB/SSC border, the *ndhF* gene was fully present within the SSC region in all *Lonicera* plastomes, extending into the IRB region with lengths ranging from 41 to 84 bp (Fig. 3).

Nucleotide diversity analysis

To assess nucleotide diversity values, the four *Lonicera* complete plastid genomes in this study were aligned. The aligned nucleotide sequences were then analyzed to calculate the nucleotide diversity of the plastid genome using DnaSP. The results revealed that the *Pi* values in the four *Lonicera* plastomes ranged from 0 to 0.15222. Three highly variable regions were identified: two intergenic regions (*ycf1-trnN-GUU* and *trnN-GUU-ndhF*) and one genic region (*accD*). Among these, the *accD* gene region exhibited the highest *Pi* value (0.15222), followed by the *ycf1-trnN-GUU* region

Table 3. Gene composition and functional categorization of the *L. caerulea* subsp. *altaica*, *L. caerulea* subsp. *pallasii*, *L. tatarica*, and *L. micrantha* plastid genomes

Category	Group of genes	Names of genes
Self-replication	Ribosomal RNA	*rrn4.5* (2), *rrn5* (2), *rrn16* (2), *rrn23* (2)
	Transfer RNA	*trnA-UGC** (2), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC**, *trnH-GUG*, *trnI-CAU* (2), *trnI-GAU** (2), *trnK-UUU**, *trnL-CAA* (2), *trnL-UAA**, *trnL-UAG*, *trnM-CAU*, *trnN-GUU* (2), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG* (2), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC* (2), *trnV-UAC**, *trnW-CCA*, *trnY-GUA*
	Small subunit of ribosome	*rps2*, *rps3*, *rps4*, *rps7* (2), *rps8*, *rps11*, *rps12** (2), *rps14*, *rps15*, *rps16**, *rps18*, *rps19*
	Large subunit of ribosome	*rpl2**, *rpl14*, *rpl16**, *rpl20*, *rpl22*, *rpl23*, *rpl32*, *rpl33*, *rpl36*
	RNA polymerase	*rpoA*, *rpoB*, *rpoC1**, *rpoC2*
	Translation initiation factor	*infA*
Photosynthesis	ATP synthase	*atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI*
	NADH dehydrogenase	*ndhA**, *ndhB** (2), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
	Subunits of cytochrome	*petA*, *petB**, *petD**, *petG*, *petL*, *petN*
	Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
	Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
	Rubisco	*rbcL*
Other genes	Maturase	*matK*
	Protease	*clpP*
	Envelope membrane protein	*cemA*
	Subunit of acetyl-CoA-carboxylase	*accD*
	C-type cytochrome synthesis gene	*ccsA*
Genes of unknown function	Conserved hypothetical chloroplast ORF	*ycf1*, *ycf2* (2), *ycf3***, *ycf4*, *ycf15* (2)

* Genes containing a single intron, genes containing two introns; (2) – duplicated genes.

(0.10250) and *trnN-GUU-ndhF* (0.09722). Notably, the *accD* region with the highest nucleotide diversity was concentrated in the LSC region of the plastid genome (Fig. 4).

Repeat sequence analysis

This study identified 163, 163, 158, and 157 SSRs in the plastid genomes of *Lonicera caerulea* subsp. *altaica*, *L. caerulea* subsp. *pallasii*, *L. tatarica*, and *L. micrantha*, respectively, resulting in a total of 641 SSRs. Five types of SSRs were identified, including mono-, di-, tri-, tetra-, and hexa-nucleotide repeats. Most of the identified SSR markers were located within the intergenic regions of the plastid genome's LSC region. Detailed information is provided in Supplementary Table S1[1]. Mononucleotide repeats were the most abundant SSR motifs, comprising approximately 72.70 % of the total SSRs, followed by dinucleotide repeats (18.72 %) and tetranucleotide repeats (5.93 %). The most abundant SSR motifs were mononucleotide repeats, which accounted for approximately 72.70 % of the total SSRs, followed by dinucleotide (18.72 %) and tetranucleotide (5.93 %) repeats. Most of the mononucleotide repeats consisted of A/T (451) rather than C/G (15), while the majority of dinucleotide repeats were composed of AT/AT (75) rather than AG/CT (45). Trinucleotide (1.25 %) and hexanucleotide (1.40 %) repeats were rare across the studied plastid genomes but were present in all four plastomes. Pentanucleotide repeats were not found in any of the studied plastomes (Table 4).

Phylogenetic analysis

The ML method was used to reconstruct the phylogenetic trees based on nucleotide sequences of complete plastid genomes (Fig. 5*a*), protein-coding genes (Fig. 5*b*), and variable region genes (Fig. 5*c*). *H. miconioides* and *T. himalayanum* were used as outgroups, while 22 *Lonicera* samples were included as ingroups. The ML trees showed that the analyzed *Lonicera* species were grouped into two clades: the *Periclymenum* subgenus clade and the *Chamaecerasus* subgenus clade. There were five subclades in the ML phylogenetic trees (Fig. 5*a–c*), which represented the sections *Eucarpifolia* and *Phenianthi* (Subclade I) within subgenus *Periclymenum*, section *Isika* (Subclades II and III), sections *Isika* and *Coeloxylosteum* (Subclades IV), and section *Nintooa* (Subclade V) within subgenus *Chamaecerasus*. The phylogenetic tree reveals that *L. tatarica* forms a subclade (IV) consisting of species from the *Coeloxylosteum* section, clustering with the *L. tatarica* (MK970584) sequence from GenBank and indicating a close relationship with *L. maackii* (MN256451) from GenBank. Furthermore, *L. caerulea* subsp. *altaica* and *L. caerulea*

[1] Supplementary Table S1 and Figs S1–S3 are available at: https://vavilovj-icg.ru/download/pict-2025-29/appx29.xlsx

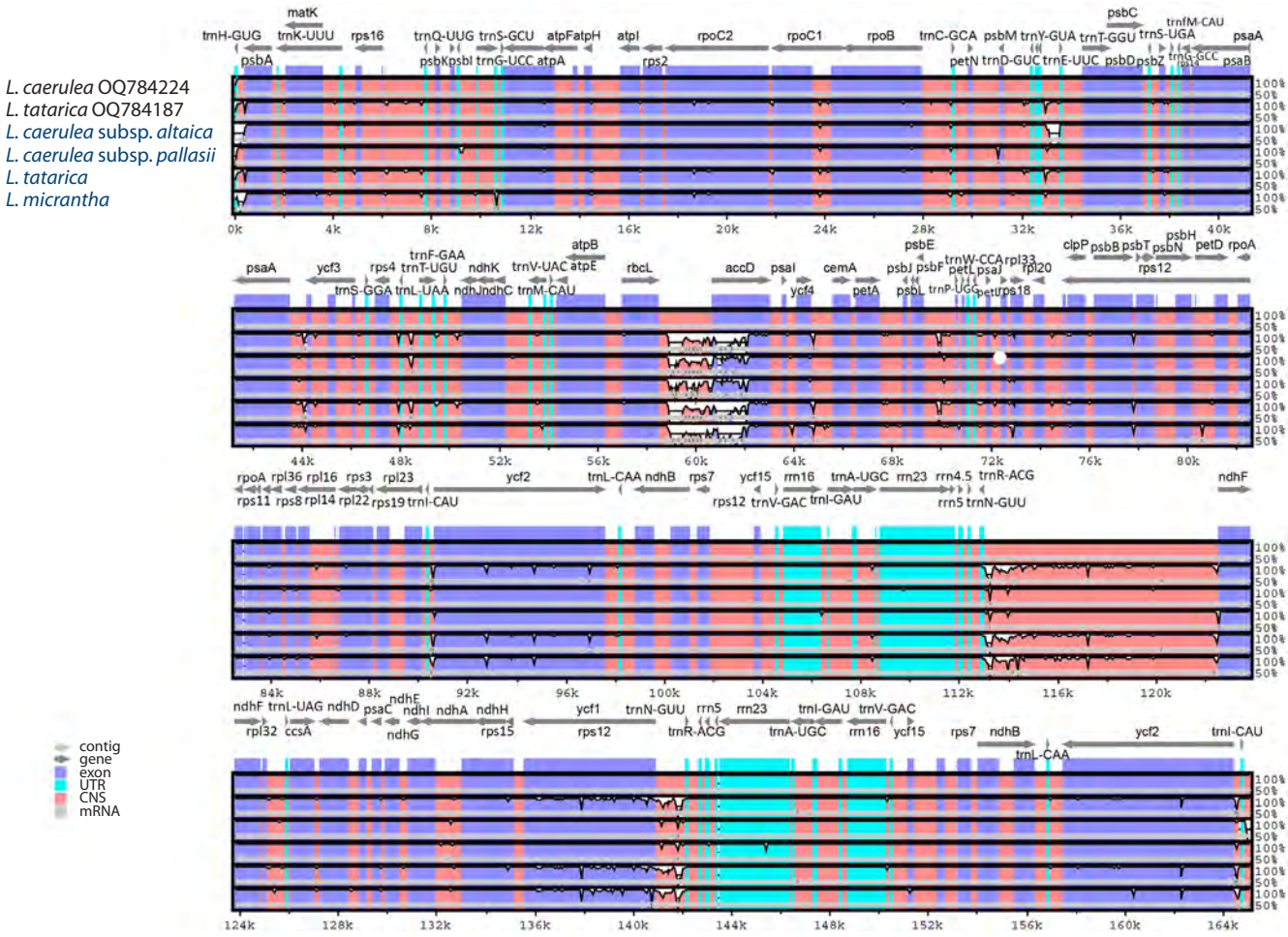


Fig. 2. Comparison of complete plastid genomes of six *Lonicera* samples using mVISTA, with *L. caerulea* (OQ784224) and *L. tatarica* (OQ784187) as reference genomes.

Gray arrows above the alignment indicate gene locations, while different colors distinguish coding and non-coding regions. The horizontal axis represents plastome coordinates, and the vertical scale depicts sequence identity percentages, ranging from 50 to 100 %.

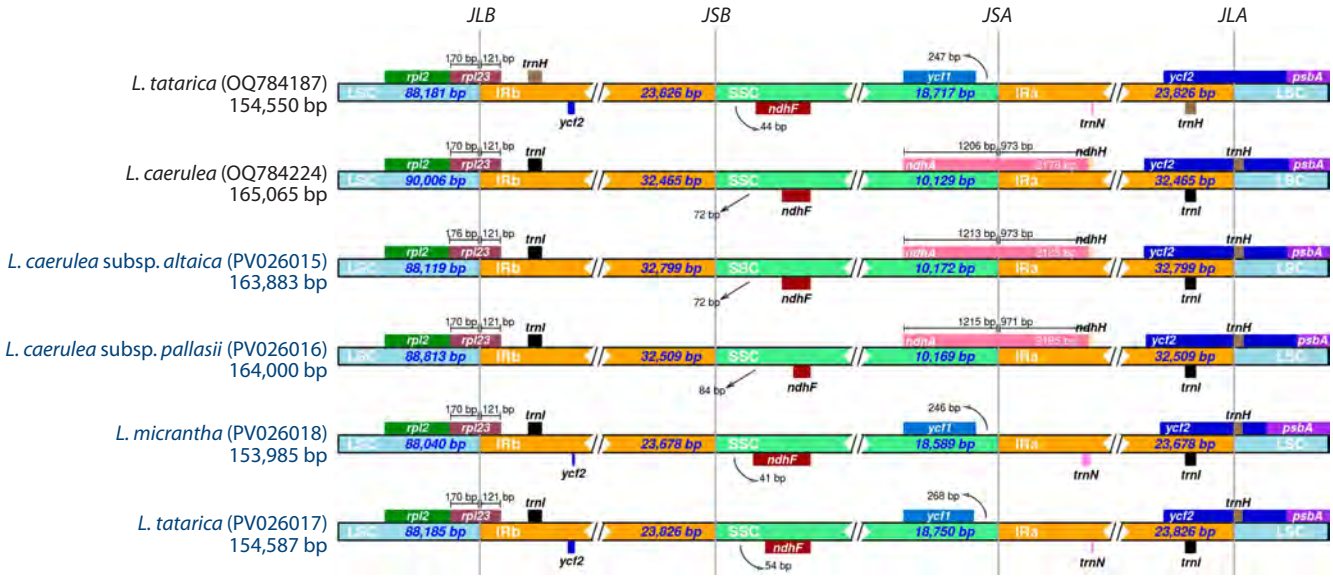


Fig. 3. Comparison of the junctions between the LSC, IR, and SSC regions in *Lonicera* plastomes. Species highlighted in blue were analyzed in this study.

JLB represents the junction between the LSC and IRB regions, JSB marks the boundary between the IRB and SSC regions; JSA indicates the junction between the SSC and IRA regions, and JLA denotes the boundary between the IRA and LSC regions.

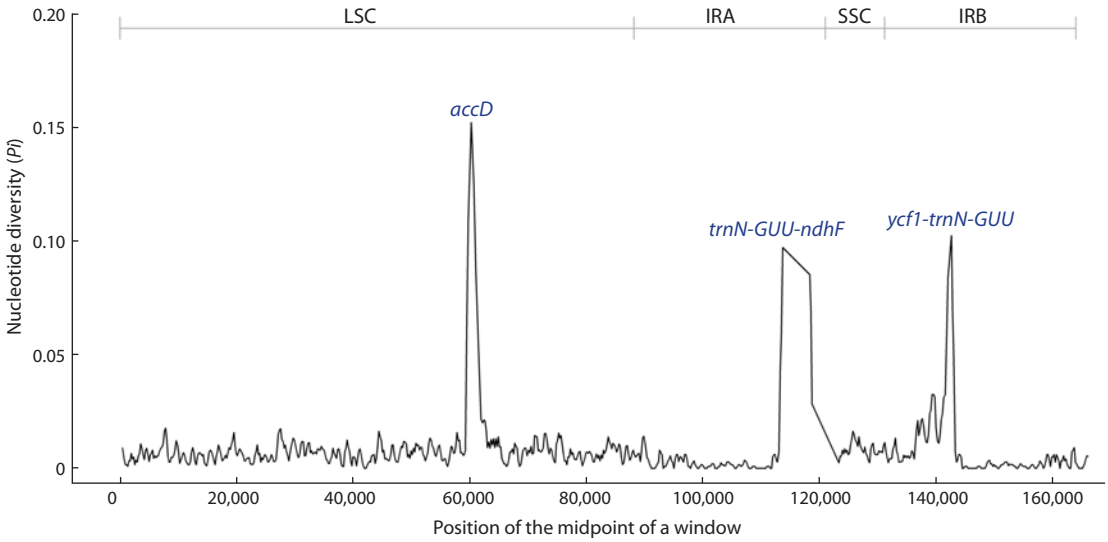


Fig. 4. Nucleotide diversity of the four *Lonicera* plastid genomes using sliding window analysis (window length – 600 bp, step size – 200 bp).

The *X*-axis represents the midpoint position of each window, while the *Y*-axis denotes the nucleotide diversity (*Pi*) value for each window.

Table 4. The number and types of identified simple sequence repeats in the four *Lonicera* plastid genomes

Type	Repeat unit	*L. caerulea* subsp. *altaica*	*L. caerulea* subsp. *pallasii*	*L. tatarica*	*L. micrantha*	Total	%
Mono-	A/T	118	116	109	108	451	72.70
	C/G	4	4	5	2	15	
Di-	AG/CT	10	10	14	11	45	18.72
	AT/AT	19	20	17	19	75	
Tri-	AAC/GTT	1	1	0	1	3	1.25
	AAG/CTT	1	1	1	1	4	
	AAT/ATT	0	0	1	0	1	
Tetra-	AAAG/CTTT	2	2	1	2	7	5.93
	AAAT/ATTT	2	3	3	5	13	
	AATC/ATTG	1	1	1	1	4	
	AATT/AATT	1	1	1	1	4	
	AGAT/ATCT	2	2	3	3	10	
Hexa-	AAAATG/ATTTTC	2	2	0	2	6	1.40
	AATGAT/ATCATT	0	0	0	1	1	
	AAACAT/ATGTTT	0	0	1	0	1	
	AAGGGT/ACCCTT	0	0	1	0	1	
Total		163	163	158	157	641	100

subsp. *pallasii* samples analyzed in this study clustered in one subclade (III) with the *L. caerulea* (OQ784224) and *L. caerulea* subsp. *edulis* (OP345475) sequences from GenBank. Also, *L. micrantha* is positioned within the *Chamaecerasus* subgenus in subclade II and clusters closely with species from the *Isika* section, particularly *L. tangutica* (MZ962399) and *L. microphylla* (OP936076) from GenBank. Most of the described subclades exhibit strong bootstrap support (100 %) at the corresponding nodes, except for subclade II in the ML phylogenetic tree based on complete plastid genome data (Fig. 5*a*), which has moderate support (53 %), indicating high confidence in their overall phylogenetic relationships.

The BI phylogenetic trees were constructed using the same set of samples based on nucleotide sequences from complete plastid genomes (Fig. S1), protein-coding genes (Fig. S2), and variable region genes (Fig. S3). The resulting trees consistently divided the analyzed *Lonicera* species into two distinct clades corresponding to the subgenera *Periclymenum* and *Chamaecerasus*. The topologies of the BI phylogenetic trees were largely congruent with those obtained using ML methods.

To further investigate the relationships and potential reticulation within *Lonicera* species, we constructed a SplitsTree phylogenetic network (Fig. 5*d*) based on complete plastid genome sequences from 22 *Lonicera* and two outgroup samples.

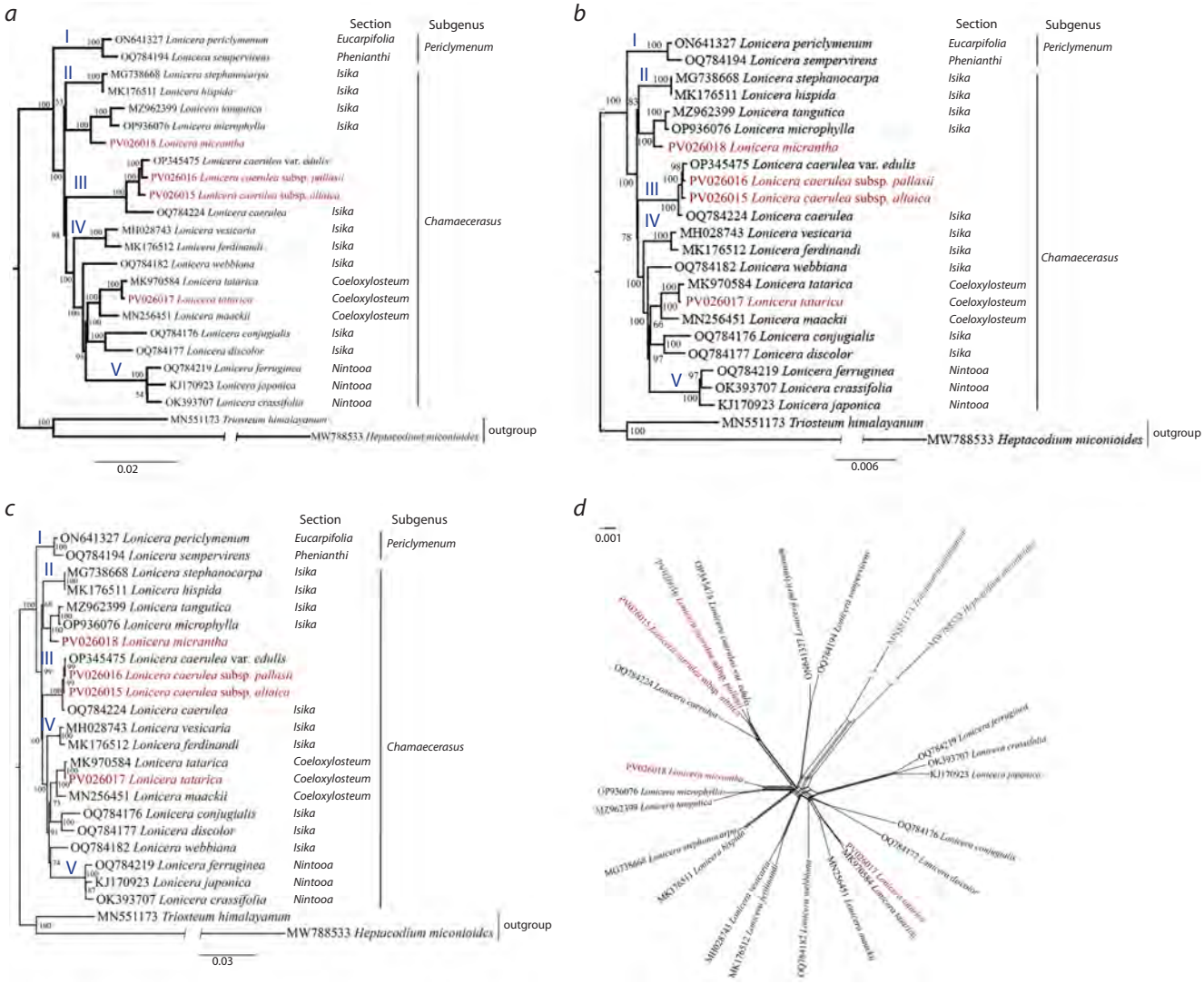


Fig. 5. Maximum likelihood phylogenetic tree of the genus *Lonicera* inferred from nucleotide sequences of the complete plastid genome (*a*), protein-coding genes (*b*), and variable region genes (*c*). Numbers at the nodes of the phylogenetic trees represent bootstrap support values.

Roman numerals (I–V) at the nodes of the phylogenetic trees denote subclade numbers. Splitstree neighbour-net network of 22 *Lonicera* and 2 outgroup plastid genomes (*d*). Species highlighted in red were sequenced in this study.

In this analysis, relationships were generally congruent with those in ML and BI phylogenetic trees. The results of the phylogenetic network coincided with the ML dendrogram (Fig. 5*a–c*). The network indicates that *L. micrantha* has evolved significantly earlier than *L. microphylla* and *L. tangutica*. Also, three subspecies of *L. caerulea* (*altaica*, *pallasii*, and *edulis*) seem to be hybrid forms of the species (Fig. 5*d*).

Discussion

In the present study, the complete plastid genomes of the *L. caerulea* subsp. *altaica*, *L. caerulea* subsp. *pallasii*, *L. tatarica*, and *L. micrantha* were sequenced using next-generation sequencing technology. These genomes were then compared with those of other *Lonicera* species to enhance our understanding of the molecular taxonomy of the genus.

The plastid genomes of the studied *Lonicera* exhibited the typical circular structure found in angiosperms (Palmer et al., 1988; Ruhlman, Jansen, 2014), consisting of an LSC region, an SSC region, and two IR regions (Fig. 1). Our annotation identified a total of 130 genes, including 115 unique genes, consisting of 85 (80 unique) protein-coding genes, 37 (30 unique) tRNA genes, and eight (4 unique) rRNA genes (Table 3). Previous studies have reported slightly different numbers of annotated protein-coding genes, with 82 and 83 genes identified in earlier analyses (He et al., 2017; Liu M.L. et al., 2018; Yang C. et al., 2023). The discrepancies in gene annotation primarily arise from differences in the *ycf15* gene, which are lost in *L. japonica*, *L. ferdinandi*, *L. hispida*, *L. nervosa*, *L. fragrantissima* var. *lancifolia*, *L. stephanocarpa*, *L. tragophylla*, *L. acuminata*, and *L. similis* plastomes. Additionally, we identified the *trnM-CAU* gene, which was not annotated in previous *Lonicera* plastome studies (Frazer et al., 2004). These variations underscore the importance of annotation accuracy, indicating that further comparative analyses are necessary to refine gene identification within the genus.

The genome sizes varied among species, ranging from 153,985 bp in *L. micrantha* to 164,000 bp in *L. caerulea* subsp. *pallasii*. Notably, the plastid genome sizes of *L. cae-*

rulea subsp. *altaica* (163,889 bp) and *L. caerulea* subsp. *pallasii* (164,000 bp) were larger than those of the other two (*L. tatarica*, and *L. micrantha*) studied species (Table 2). The variations in the total length of plastid genomes are typically associated with the expansion and contraction of IR regions (Zhang X.F. et al., 2021). In this study, the IR regions of *L. caerulea* subsp. *altaica* (32,799 bp) and *L. caerulea* subsp. *pallasii* (32,509 bp) were found to be longer than the SSC and LSC regions, contributing to their relatively larger plastome sizes (Fig. 3). These results align with the previously reported plastid genome size of *L. caerulea* (165,065 bp) (Yang X.L. et al., 2024), suggesting that the specific plastid genome lengths observed in the studied *L. caerulea* subspecies may be a common evolutionary characteristic with the *L. caerulea* species.

Molecular markers are essential tools in modern biological research, playing a crucial role in unraveling genetic diversity, phylogenetic relationships, and population dynamics (Wang X.R., Szmidt, 2001; Al-Hadeithi, Jasim, 2021). Among them, DNA barcoding markers offer an efficient approach to species identification by targeting short, conserved regions of the genome (Chac, Thinh, 2023). These markers have made one of the most significant contributions to advancing our understanding of evolutionary processes, establishing DNA barcoding as a core methodology in plant taxonomy (Purty, Chatterjee, 2016; Zhu S. et al., 2022). The highly variable regions in nucleotide sequences of the plastid genome can be used as potential specific DNA barcoding markers for specific plant genera. Using mVISTA (Fig. 2) and sliding window analysis (Fig. 4), we identified three highly variable regions in this study: two intergenic regions (*ycf1-trnN-GUU* and *trnN-GUU-ndhF*) and one genic region (*accD*). These regions show promise as DNA barcoding markers for the phylogenetic analysis of *Lonicera* species. Notably, the *trnN-GUU-ndhF* region has been reported as particularly useful for developing molecular markers in *Lonicera* species (Liu M.L. et al., 2018; Yang C. et al., 2023).

Our study found that the *accD* gene region is the most variable, a finding consistent with previous studies in Asteraceae (Kim et al., 2020) and Fabaceae (Zhang T. et al., 2024). Two intergenic regions (*ycf1-trnN-GUU* and *trnN-GUU-ndhF*) identified in this study were also reported in other plant species. For example, *ycf1-trnN-GUU* was a highly variable region in the plastid genomes of *Parasenecio* (Liu X. et al., 2023) and *Medicago* (Jiao et al., 2023) species. Z. Cao et al. (2023) and W. Xing et al. (2024) reported that the *trnN-GUU-ndhF* intergenic region is hypervariable in the plastid genomes of *Neocinnamomum* taxa and *Pinellia ternata*, respectively. This study identified two highly variable intergenic regions and one genic region as promising candidates for DNA barcoding markers in future research. Nonetheless, further studies are needed to assess the effectiveness of these divergent markers in the phylogenetic analysis of *Lonicera* species.

Another important class of molecular markers is the simple sequence repeat (SSR) markers, which are widely recognized for their value in plant population genetics in assessing genetic diversity, population structure, and evolutionary relationships (Chen F. et al., 2015; Yermagambetova et al., 2024). In our study, we identified a total of 641 SSR markers across the plastid genomes analyzed, with individual counts ranging from 158 in *L. tatarica* to 163 in both *L. caerulea* subsp. *altaica* and *L. caerulea* subsp. *pallasii* plastid genomes (Table 4). Notably, the majority of these SSRs were located in the intergenic regions of the LSC region, a distribution pattern that aligns well with previous findings on angiosperm plastomes (Xia C. et al., 2022; Nyamgerel et al., 2024).

Our results reveal that mononucleotide repeats are the most prevalent SSR motifs across the four *Lonicera* plastomes analyzed, which are common for Caprifoliaceae representatives (Liu H. et al., 2022; Wang L. et al., 2024). Notably, the majority of the mononucleotide repeats were composed of A/T (451) rather than C/G (15), while dinucleotide repeats were predominantly composed of AT/AT (75) as opposed to AG/CT (45). This distribution is consistent with patterns observed in plastid genomes of many other plant species (Li X.Q. et al., 2019; Souza et al., 2019). Numerous studies have demonstrated that SSR markers derived from plastid genome sequences are effective for assessing genetic diversity in plant populations (Jo et al., 2022; Lācis et al., 2022; Guo et al., 2025). The SSR markers identified in our study hold potential for population genetic analyses within the genus *Lonicera*. However, further validation is required to confirm their efficacy and reliability in elucidating the genetic structure of *Lonicera* species populations.

Plastid genome nucleotide sequences have become a powerful tool in phylogenetic studies of different plant genera (Wu et al., 2021; Xia Q. et al., 2023). Their conserved structure, uniparental inheritance, and relatively slow mutation rate make them ideal for resolving evolutionary relationships across diverse plant lineages (Chen J. et al., 2022; Feng et al., 2024). With the development of next-generation sequencing technologies, the rapid and cost-effective sequencing of entire plastid genomes has become increasingly accessible, enhancing their utility in plant taxonomy by providing greater phylogenetic resolution and a deeper understanding of plant evolutionary history. This study utilized complete plastome sequences, protein-coding gene sequences, and variable region gene sequences for the phylogenetic analysis of the studied *Lonicera* species (*L. caerulea* subsp. *altaica*, *L. caerulea* subsp. *pallasii*, *L. tatarica*, and *L. micrantha*), along with publicly available sequences from GenBank, to contribute to a better understanding of phylogenetic relationships within the genus. The ML trees based on the sequences of the complete plastid genome (Fig. 5*a*), protein-coding genes (Fig. 5*b*), and variable region genes (Fig. 5*c*) of the 22 *Lonicera* samples and two outrgoup samples (*H. miconioides* and *T. himalayanum*) was reconstructed.

The phylogenetic analyses in this study revealed that the *Lonicera* species were grouped into two major clades, corresponding to the subgenera *Periclymenum* and *Chamaecerasus*, which is consistent with previous phylogenetic studies (Srivastav et al., 2023; Yang X.L. et al., 2024). Furthermore, the larger clade representing subgenus *Chamaecerasus* was further divided into four distinct subclades corresponding to sections *Isika* (Subclades II and III), sections *Isika* and *Coeloxylosteum* (Subclades IV), and section *Nintooa* (Subclade V). Within subgenus *Chamaecerasus*, species are subclustering into four subclades and align with recognized sectional classifications (Srivastav et al., 2023).

The placement of *L. micrantha*, *L. caerulea* subsp. *altaica* and *L. caerulea* subsp. *pallasii*, which had not been previously

assigned to a section, suggests their belonging to the *Isika* section based on their close clustering with other members of this section. Additionally, the finding supports the assumption that *L. caerulea* subsp. *altaica* and *L. caerulea* subsp. *pallasii* share a common evolutionary history with *L. caerulea*, which is also supported by the plastid genome structure of these species. Furthermore, *L. tatarica* forms a subclade (IV) with *L. tatarica* (MK970584) and *L. maackii* (MN256451) from GenBank, grouping within the *Coeloxylosteum* section. Notably, these samples are positioned between the species of section *Isika*, suggesting a possible evolutionary relationship between these two sections.

The phylogenetic analysis aimed to clarify the phylogenetic positions of the studied *Lonicera* species from Kazakhstan using plastid genome nucleotide sequences, including three newly sequenced ones (*L. caerulea* subsp. *altaica*, *L. caerulea* subsp. *pallasii*, and *L. micrantha*) in this study. The genomic data obtained in this study provide valuable resources for future phylogenetic research, contributing to an understanding of evolutionary relationships within the genus *Lonicera* and supporting further taxonomic revisions. Based on the comparison of phylogenetic trees reconstructed using different datasets and methods, we conclude that the maximum likelihood tree derived from complete plastid genome sequences was the most informative, and its topology is consistent with those reported in previous studies (Srivastav et al., 2023; Yang X.L. et al., 2024). Furthermore, the nucleotide sequences of variable regions such as *accD-ycf1-ndhF-trnN-GUU* also demonstrate high potential for use in DNA barcoding, and may serve as valuable molecular markers for species phylogenetic studies within the genus *Lonicera*.

Conclusion

The complete plastid genomes of *L. caerulea* subsp. *altaica*, *L. caerulea* subsp. *pallasii*, *L. tatarica*, and *L. micrantha* exhibited the typical circular structure with four distinct regions. Structural variations were observed in the plastomes of *L. caerulea* subsp. *altaica* and *L. caerulea* subsp. *pallasii*, particularly in genome sizes, which were larger than in the other two species (*L. tatarica* and *L. micrantha*) due to an extended IR region. This finding aligns with previous studies on *L. caerulea* plastomes, further supporting their shared evolutionary history. The nucleotide sequences of variable regions such as *accD-ycf1-ndhF-trnN-GUU* demonstrate high potential for use in DNA barcoding, and may serve as valuable molecular markers for species phylogenetic studies within the genus *Lonicera*. Further studies are required to assess the effectiveness of the identified simple sequence repeats.

References

Abdulina S.A. Checklist of Vascular Plants of Kazakhstan. Almaty, 1999 (in Russian)

Al-Hadeithi Z.S.M., Jasim S.A. Study of plant genetic variation through molecular markers: an overview. *J Pharm Res Int*. 2021;33:464-473. doi 10.9734/jpri/2021/v33i45B32828

Ali I., Khan D., Ali F., Bibi H., Malik A. Phytochemical, antioxidant and antifungal studies on the constituents of *Lonicera quinqueocularis*. *J Chem Soc Pak*. 2013;35(1):139-143. doi 10.5897/JMPR2013.5245

Almerekova S., Yermagambetova M., Ivashchenko A., Abugalieva S., Turuspekov Y. Assessment of complete plastid genome sequences of *Tulipa alberti* Regel and *Tulipa greigii* Regel species from Kazakhstan. *Genes*. 2024;15:1447. doi 10.3390/genes15111447

Ametov A.A., Mukhitdinov N.M., Abidkulova K.T., Karasholakova L.N., Ydyrys A. Characteristics of plant communities with Lonicera iliensis Pojark. in the middle flow of the River Ili. *KazNU Bulletin. Biol Series*. 2016;4(69):12-21 (in Russian)

Amiryousefi A., Hyvönen J., Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics*. 2018;34:3030-3031. doi 10.1093/bioinformatics/bty220

Baitulin I.O., Sitpayeva G.T. (Eds) Red Book of Kazakhstan: Plants. Astana, 2014 (in Russian)

Beier S., Thiel T., Munch T., Scholz U., Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33(16):2583-2585. doi 10.1093/bioinformatics/btx198

Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120. doi 10.1093/bioinformatics/btu170

Boyarskikh I.G., Kostikova V.A. Changes in the individual and group composition of polyphenols in leaves of *Lonicera caerulea* subsp. *altaica* and *Spiraea chamaedryfolia* as related to chemical element content in soil and plants on ultra-alkaline parent rock material. *Rastitelnye Resursy*. 2023;59(2):164-179. doi 10.31857/S0033994623020048 (in Russian)

Cao Z., Yang L., Xin Y., Xu W., Li Q., Zhang H., Tu Y., Song Y., Xin P. Comparative and phylogenetic analysis of complete chloroplast genomes from seven *Neocinnamomum* taxa (Lauraceae). *Front Plant Sci*. 2023;14:1205051. doi 10.3389/fpls.2023.1205051

Chac L.D., Thinh B.B. Species identification through DNA barcoding and its applications: a review. *Biol Bull*. 2023;50:1143-1156. doi 10.1134/s106235902360229x

Chen F., Liu H., Yao Q., Fang P., Lv F. Genetic variations and evolutionary relationships among radishes (*Raphanus sativus* L.) with different flesh colors based on red pigment content, karyotype and simple sequence repeat analysis. *Afr J Biotechnol*. 2015;16:3270-3281. doi 10.5897/AJB2015.14911

Chen J., Xie D., He X., Yang Y., Li X. Comparative analysis of the complete chloroplast genomes in *Allium* section *Bromatorrhiza* species (Amaryllidaceae): phylogenetic relationship and adaptive evolution. *Genes*. 2022;13(7):1279. doi 10.3390/genes13071279

Chen X., Zhou J., Cui Y., Wang Y., Duan B., Yao H. Identification of *Ligularia* herbs using the complete chloroplast genome as a superbarcode. *Front Pharmacol*. 2018;9:695. doi 10.3389/fphar.2018.00695

Dierckxsens N., Mardulyn P., Smits G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 2017;45(4):e18. doi 10.1093/nar/gkw955

Dong S., Ying Z., Yu S., Wang Q., Liao G., Ge Y., Cheng R. Complete chloroplast genome of *Stephania tetrandra* (Menispermaceae) from Zhejiang Province: insights into molecular structures, comparative genome analysis, mutational hotspots and phylogenetic relationships. *BMC Genomics*. 2021;22(1):880. doi 10.1186/s12864-021-08193-x

Donoghue M.J., Bell C.D., Li J. Phylogenetic patterns in Northern Hemisphere plant geography. *Int J Plant Sci*. 2001;162(S6):S41-S52. doi 10.1086/323278

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

Feng Z., Zheng Y., Jiang Y., Pei J., Huang L. Phylogenetic relationships, selective pressure and molecular markers development of six species in subfamily Polygonoideae based on complete chloroplast genomes. *Sci Rep*. 2024;14(1):9783. doi 10.1038/s41598-024-58934-7

Frazer K.A., Pachter L., Poliakov A., Rubin E.M., Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32:W273-W279. doi 10.1093/nar/gkh458

Ge L., Li J., Wan H., Zhang K., Wu W., Zou X., Wu S., Zhou B., Tian J., Zeng X. Novel flavonoids from *Lonicera japonica* flower buds and validation of their anti-hepatoma and hepatoprotective activity *in vitro* studies. *Ind Crops Prod*. 2018;125:114-122. doi 10.1016/j.indcrop.2018.08.073

Golubev D., Zemskaya N., Shevchenko O., Shaposhnikov M., Kukuman D., Patov S., Punegov V., Moskalev A. Honeysuckle extract (*Lonicera pallasii* L.) exerts antioxidant properties and extends the lifespan and healthspan of *Drosophila melanogaster*. *Biogerontology*. 2022;23(2):215-235. doi 10.1007/s10522-022-09954-1

Guo Q., Xue X., Wang D., Zhang L., Liu W., Wang E., Cui X., Hou X. Genetic diversity and population genetic structure of *Paeonia suffruticosa* by chloroplast DNA simple sequence repeats (CpSSRs). *Hortic Plant J*. 2025;11(1):367-376. doi 10.1016/j.hpj.2023.10.006

Hara H. A revision of Caprifoliaceae of Japan with Reference to Allied Plants in other Districts and the Adoxaceae. Ginkgoana. Tokyo: Academia Scientific Book, 1983

Hayes D.J., Peterson B.J. Growth of *Lonicera caerulea* across fertility and moisture conditions: comparisons with *Lonicera villosa* and invasive congeners. *HortScience*. 2020;55(2):149-155. doi 10.21273/HORTSCI14318-19

He L., Qian J., Li X., Sun Z., Xu X., Chen S. Complete chloroplast genome of medicinal plant *Lonicera japonica*: genome rearrangement, intron gain and loss, and implications for phylogenetic studies. *Molecules*. 2017;22(2):249. doi 10.3390/molecules22020249

Hong Z., He W., Liu X., Tembrock L.R., Wu Z., Xu D. Comparative analyses of 35 complete chloroplast genomes from the genus *Dalbergia* (Fabaceae) and the identification of DNA barcodes for tracking illegal logging and counterfeit rosewood. *Forests*. 2022;13(4):626. doi 10.3390/f13040626

Howe C.J., Barbrook A.C., Koumandou V.L., Nisbet R.E.R., Symington H.A., Wightman T.F. Evolution of the chloroplast genome. *Philos Trans R Soc Lond B Biol Sci*. 2003;358(1429):99-106. doi 10.1098/rstb.2002.1176

Hsu P.S., Hu C.C., Wang H.J. Flora Reipublicae Popularis Sinicae. Vol. 72. Science Press, 1988 (in Chinese)

Huson D.H., Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2006;23(2):254-267. doi 10.1093/molbev/msj030

Jiao Y.X., He X.F., Song R., Wang X.M., Zhang H., Aili R., Chao Y.H., Shen Y.H., Yu L.X., Zhang T.J., Jia S.G. Recent structural variations in the *Medicago* chloroplast genomes and their horizontal transfer into nuclear chromosomes. *J Syst Evol*. 2023;61(4):627-642. doi 10.1111/jse.12900

Jo I.H., Han S., Shim D., Ryu H., Hyun T.K., Lee Y. Complete chloroplast genome of the inverted repeat-lacking species *Vicia bungei* and development of polymorphic simple sequence repeat markers. *Front Plant Sci*. 2022;13:891783. doi 10.3389/fpls.2022.891783

Kim G.B., Lim C.E., Kim J.S., Kim K., Lee J.H., Yu H.J., Mun J.H. Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: insights into evolutionary divergence and phylogenomic implications. *BMC Genomics*. 2020;21(1):415. doi 10.1186/s12864-020-06812-7

Kong D., Li Y., Bai M., Deng Y., Liang G., Wu H. A comparative study of the dynamic accumulation of polyphenol components and the changes in their antioxidant activities in diploid and tetraploid *Lonicera japonica*. *Plant Physiol Biochem*. 2017;112:87-96. doi 10.1016/j.plaphy.2016.12.027

Kushnarenko S.V., Karasholakova L.N., Ozek G., Abidkulova K.T., Mukhitdinov N.M., Baser K.H.C., Ozek T. Investigation of essential oils from three natural populations of *Lonicera iliensis*. *Chem Nat Compd*. 2016;52:751-753. doi 10.1007/s10600-016-1765-6

Li Q. The complete chloroplast genomes of *Primula obconica* provide insight that neither species nor natural section represent monophyletic taxa in *Primula* (Primulaceae). *Genes*. 2022;13(4):567. doi 10.3390/genes13040567

Li X.Q., Zuo Y.J., Zhu X.X., Liao S., Ma J.S. Complete chloroplast genomes and comparative analysis of sequence evolution among seven *Aristolochia* (Aristolochiaceae) medicinal species. *Int J Mol Sci*. 2019;20(5):1045. doi 10.3390/ijms20051045

Lin L.M., Zhang X.G., Zhu J.J., Gao H.M., Wang Z.M., Wang W.H. Two new triterpenoid saponins from the flowers and buds of *Lonicera japonica*. *J Asian Nat Prod Res*. 2008;10(10):925-929. doi 10.1080/10286020802217366

Liu H., Liu W., Ahmad I., Xiao Q., Li X., Zhang D., Fang J., Zhang G., Xu B., Gao Q., Chen S. Complete chloroplast genome sequence of *Triosteum sinuatum*, insights into comparative chloroplast genomics, divergence time estimation and phylogenetic relationships among Dipsacales. *Genes*. 2022;13(5):933. doi 10.3390/genes13050933

Liu M.L., Fan W.B., Wang N., Dong P.B., Zhang T.T., Yue M., Li Z.H. Evolutionary analysis of plastid genomes of seven *Lonicera* L. species: implications for sequence divergence and phylogenetic relationships. *Int J Mol Sci*. 2018;19(12):4039. doi 10.3390/ijms19124039

Liu M., Yu Q., Yi Y., Xiao H., Putra D.F., Ke K., Zhang Q., Li P. Antiviral activities of *Lonicera japonica* Thunb. components against grouper iridovirus *in vitro* and *in vivo*. *Aquaculture*. 2020;519:734882. doi 10.1016/j.aquaculture.2019.734882

Liu X., Luo J., Zhang M., Wang Q., Liu J., Wu D., Fu Z. Phylogenomic analysis of two species of *Parasenecio* and comparative analysis within tribe Senecioneae (Asteraceae). *Diversity*. 2023;15:563. doi 10.3390/d15040563

Lohse M., Drechsel O., Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet*. 2007;52:267-274. doi 10.1007/s00294-007-0161-y

Luo C., Huang W., Sun H., Yer H., Li X., Li Y., Yan B., Wang Q., Wen Y., Huang M., Huang H. Comparative chloroplast genome analysis of Impatiens species (Balsaminaceae) in the Karst area of China: insights into genome evolution and phylogenomic implications. *BMC Genomics*. 2021;22(1):571. doi 10.1186/s12864-021-07807-8

Lācis G., Kārkliņa K., Bartulsons T., Stalažs A., Jundzis M., Baļķe I., Ruņģis D., Strautiņa S. Genetic structure of a *Ribes* genetic resource collection: inter- and intra-specific diversity revealed by chloroplast DNA simple sequence repeats (CpSSRs). *Sci Hortic*. 2022;304:111285. doi 10.1016/j.scienta.2022.111285

Maximowicz C.J. Diagnoses Plantarum Novarum Asiaticarum. Petropoli, Imperialis Academiae Scientiarum, 1877. doi 10.5962/bhl.title.46308

Nakai T. A new classification of the genus *Lonicera* in the Japanese Empire, together with the diagnoses of new species and new varieties. *J Jpn Bot*. 1938;14:359-375

Nakaji M., Tanaka N., Sugawara T. A molecular phylogenetic study of *Lonicera* L. (Caprifoliaceae) in Japan based on chloroplast DNA sequences. *Acta Phytotaxon Geobot*. 2015;66(3):137-151. doi 10.18942/apg.KJ00010115701

Nguyen L.T., Schmidt H.A., Von Haeseler A., Minh B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268-274. doi 10.1093/molbev/msu300

Ni F.Y. Chemical constituents from flower buds of *Lonicera japonica*. *Chin Tradit Herb Drugs*. 2017;48(18):3689-3692. doi 10.7501/j.issn.0253-2670.2017.18.004

Nyamgerel N., Baasanmunkh S., Oyuntsetseg B., Tsegmed Z., Bayarmaa G., Lazkov G., Pyak E., Gil H.Y., Park I., Choi H.J. Comparative plastome analysis and taxonomic classification of snow lotus species (*Saussurea*, Asteraceae) in Central Asia and Southern Siberia. *Funct Integr Genomics*. 2024;24(2):42. doi 10.1007/s10142-024-01309-y

Oyuntsetseg D., Nyamgerel N., Baasanmunkh S., Oyuntsetseg B., Urgamal M., Yoon J.W., Bayarmaa G.A., Choi H.J. The complete chloroplast genome and phylogenetic results support the species position of *Swertia banzragczii* and *Swertia marginata* (Gentianaceae) in Mongolia. *Bot Stud*. 2024;65(1):11. doi 10.1186/s40529-024-00417-z

Palmer J.D., Jansen R.K., Michaels H.J., Chase M.W., Manhart J.R. Chloroplast DNA variation and plant phylogeny. *Ann Missouri Bot Gard*. 1988;75(4):1180-1206. doi 10.2307/2399279

Park H.S., Park K.I., Lee D.H., Kang S.R., Nagappan A., Kim J.A., Kim E.H., Lee W.S., Shin S.C., Hah Y.S., Kim G.S. Polyphenolic extract isolated from Korean *Lonicera japonica* Thunb. induces G2/M cell cycle arrest and apoptosis in HepG2 cells: involvement of PI3K/Akt and MAPKs. *Food Chem Toxicol*. 2012;50(7):2407-2416. doi 10.1016/j.fct.2012.04.034

Purty R.S., Chatterjee S. DNA barcoding: an effective technique in molecular taxonomy. *Austin J Biotechnol Bioeng*. 2016;3(1):1059

Qu X.J., Moore M.J., Li D.Z., Yi T.S. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 2019;15:50. doi 10.1186/s13007-019-0435-7

Rambaut A. FigTree, a graphical viewer of phylogenetic trees. 2009. Available: http://tree.bio.ed.ac.uk/software/figtree/

Rehder A. Synopsis of the genus *Lonicera*. *Mo Bot Gard Annu Rep*. 1903;27-232. doi 10.2307/2400049

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539-542. doi 10.1093/sysbio/sys029

Rozas J., Sánchez-DelBarrio J.C., Messeguer X., Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. 2003;19(18):2496-2497. doi 10.1093/bioinformatics/btg359

Ruhlman T.A., Jansen R.K. The plastid genomes of flowering plants. *Methods Mol Biol*. 2014;1132:3-38. doi 10.1007/978-1-62703-995-6_1

Souza U.J.B., Nunes R., Targueta C.P., Diniz-Filho J.A.F., Telles M.P.C. The complete chloroplast genome of *Stryphnodendron adstringens* (Leguminosae – Caesalpinioideae): comparative analysis with related mimosoid species. *Sci Rep*. 2019;9(1):14206. doi 10.1038/s41598-019-50620-3

Srivastav M., Clement W.L., Landrein S., Zhang J., Howarth D.G., Donoghue M.J. A phylogenomic analysis of *Lonicera* and its bearing on the evolution of organ fusion. *Am J Bot*. 2023.110(4):e16143. doi 10.1002/ajb2.16143

Taldybay A., Aidarbayeva D., Kurmantayeva A., Mussaev K., Amanbekova D., Joltukova B. Medicinal plants in the flora of Zhetysu Alatau, Zhetysu region, Kazakhstan. *Casp J Environ Sci*. 2024; 22(3):567-579. doi 10.22124/CJES.2024.7831

Tang D., Lin Y., Wei F., Quan C., Wei K., Wei Y., Cai Z., Kashif M.H., Miao J. Characteristics and comparative analysis of *Mesona chinensis* Benth chloroplast genome reveals DNA barcode regions for species identification. *Funct Integr Genomics*. 2022;22:467-479. doi 10.1007/s10142-022-00846-8

Varlashchenko L., Balabak A., Mamchur V., Polishchuk V. Application of introduced representatives of *Lonicera pileata* Oliv. in landscaping of the Right-Bank Forest-Steppe of Ukraine. *Grassroots J Nat Res*. 2021;4(3):34-41. doi 10.33002/nr2581.6853.040304

Vdovina T.A. Biochemical evaluation of fruits of promising forms of Altai honeysuckle (*Lonicera altaica* Pall.), introduced in the conditions of the Astana Botanical Garden. *Probl Bot South Sib Mongol*. 2019;18(1):556-560. doi 10.14258/pbssm.2019117 (in Russin)

Vdovina T.A., Lagus O.A., Isakova E.A., Vinokurov A.A. State of coenopopulations of wild berry plants in the territory of Kazakhstan Altai. *Bull Karaganda Univ Biol Med Geogr Ser*. 2024;11629(4): 129-134. doi 10.31489/2024BMG4/129-134

Wang G.Q., Morales-Briones D.F., Landis J.B., Wang H.X., Wang H.F. Progress in molecular systematics of Caprifoliaceae. *Taxon*. 2024; 74(1):5-12. doi 10.1002/tax.13279

Wang H.X., Liu H., Moore M.J., Landrein S., Liu B., Zhu Z.X., Wang H.F. Plastid phylogenomic insights into the evolution of the Caprifoliaceae *s.l.* (Dipsacales). *Mol Phylogenet Evol*. 2020;142: 106641. doi 10.1016/j.ympev.2019.106641

Wang L., Li F., Zhao K., Yang J., Sun H., Cui X., Dong W., Li E., Wang N. Comparative plastomes sheds light on phylogeny of *Weigela*. *Front Plant Sci*. 2024;15:1487725. doi 10.3389/fpls.2024.1487725

Wang X.R., Szmidt A.E. Molecular markers in population genetics of forest trees. *Scand J For Res*. 2001;16(3):199-220. doi 10.1080/02827580118146

Wen J. Evolution of Eastern Asian–Eastern North American biogeographic disjunctions: a few additional issues. *Int J Plant Sci*. 2001; 162(S6):S117-S122. doi org/10.1086/322940

Wu L., Cui Y., Wang Q., Xu Z., Wang Y., Lin Y., Song J., Yao H. Identification and phylogenetic analysis of five *Crataegus* species (Rosaceae) based on complete chloroplast genomes. *Planta*. 2021; 254(1):14. doi 10.1007/s00425-021-03667-4

Xia C., Wang M., Guan Y., Li Y., Li J. Comparative analysis of complete chloroplast genome of ethnodrug *Aconitum episcopale* and insight into its phylogenetic relationships. *Sci Rep*. 2022;12:9439. doi 1038/s41598-022-13524-3

Xia Q., Zhang H., Lv D., El-Kassaby Y.A., Li W. Insights into phylogenetic relationships in *Pinus* inferred from a comparative analysis of complete chloroplast genomes. *BMC Genomics*. 2023;24:346. doi 10.1186/s12864-023-09439-6

Xing W., Yu W., Kong Y., Ren X., Zhu L., Li Q., Yang Y., Cheng Y., Wang H. Intraspecific chloroplast genome genetic polymorphism of *Pinellia ternata* (Xi Junecry) and its revelation of a single origin in phylogeny. *Genes*. 2024;15(12):1638. doi 10.3390/genes15121638

Yang C., Zhang N., Wu S., Jiang C., Xie L., Yang F., Yu Z. A comparative analysis of the chloroplast genomes of three *Lonicera* medicinal plants. *Genes*. 2023;14(3):548. doi 10.3390/genes14030548

Yang Q.R., Zhao Y.Y., Hao J.B., Li W.D. Research progress on chemical constituents and their differences between Lonicerae japonicae flos and Lonicerae flos. *Zhongguo Zhong Yao Za Zhi*. 2016;41(7): 1204-1211 (in Chinese) doi 10.4268/cjcmm20160708

Yang X.L., Sun Q.H., Morales-Briones D.F., Landis J.B., Chen D.J., Wang H.X., Wen J., Wang H.F. New insights into infrageneric relationships of *Lonicera* (Caprifoliaceae) as revealed by nuclear ribosomal DNA cistron data and plastid phylogenomics. *J Syst Evol*. 2024;62(3):333-357. doi 10.1111/jse.13014

Yermagambetova M., Abugalieva S., Turuspekov Y., Almerekova S. Illumina sequencing data of the complete chloroplast genome of rare species *Juniperus seravschanica* (Cupressaceae) from Kazakhstan. *Data Brief*. 2023;46:108866. doi 10.1016/j.dib.2022.108866

Yermagambetova M., Almerekova S., Ivashchenko A., Turuspekov Y., Abugalieva S. Genetic diversity of *Tulipa alberti* and *T. greigii* populations from Kazakhstan based on application of expressed sequence tag simple sequence repeat markers. *Plants*. 2024;13(18): 2667. doi 10.3390/plants13182667

Yoo H.J., Kang H.J., Song Y.S., Park E.H., Lim C.J. Anti-angiogenic, antinociceptive and anti-inflammatory activities of *Lonicera japonica* extract. *J Pharm Pharmacol*. 2008;60(6):779-786. doi 10.1211/jpp.60.6.0014

Zhang T., Li M., Zhu X., Li S., Guo M., Guo C., Shu Y. Comparative chloroplast genome analysis provided adaptive evolution insights in *Medicago ruthenica*. *Int J Mol Sci*. 2024;25:8689. doi 10.3390/ijms25168689

Zhang X.F., Landis J.B., Wang H.X., Zhu Z.X., Wang H.F. Comparative analysis of chloroplast genome structure and molecular dating in Myrtales. *BMC Plant Biol*. 2021;21:219. doi 10.1186/s12870-021-02985-9

Zhang Z., Zhang Y., Song M., Guan Y., Ma X. Species identification of *Dracaena* using the complete chloroplast genome as a super-barcode. *Front Pharmacol*. 2019;10:1441. doi 10.3389/fphar.2019.01441

Zhao S.Y., Muchuku J.K., Liang H.Y., Wang Q.F. A complete chloroplast genome of a traditional Chinese medicine herb, *Rubia podantha*, and phylogenomics of Rubiaceae. *Physiol Mol Biol Plants*. 2023;29:843-853. doi 10.1007/s12298-023-01302-y

Zhu M., Feng P., Ping J., Li J., Su Y., Wang T. Phylogenetic significance of the characteristics of simple sequence repeats at the genus level based on the complete chloroplast genome sequences of Cyatheaceae. *Ecol Evol*. 2021;11(20):14327-14340. doi 10.1002/ece3.8151

Zhu S., Liu Q., Qiu S., Dai J., Gao X. DNA barcoding: an efficient technology to authenticate plant species of traditional Chinese medicine and recent advances. *Chin Med*. 2022;17:112. doi 10.1186/s13020-022-00655-y

Conflict of interest. The authors declare no conflict of interest.
Received July 7, 2025. Revised July 29, 2025. Accepted July 29, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(6):896-904

doi 10.18699/vjgb-25-96

A higher far-red intensity promotes the transition to flowering in triticale grown under speed breeding conditions

A.O. Blinkov [1] ✉, V.M. Nagamova [1], Y.V. Minkova [1], N.Yu. Svistunova [1], S. Radzeniece [1], A.A. Kocheshkova [1], N.N. Sleptsov [1, 2], A.V. Freymans[3], V.V. Panchenko[4], A.G. Chernook [1], G.I. Karlov [1], M.G. Divashuk [1]

[1] All-Russian Research Institute of Agricultural Biotechnology, Moscow, Russia
[2] Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, Russia
[3] LLC "Climbiotech", Moscow, Russia
[4] P.P. Lukyanenko National Grain Centre, Krasnodar, Russia
✉ aoblinkov@gmail.com

Abstract. It typically takes 12 to 15 years to develop a new promising variety. One of the ways to reduce this time is through speed breeding. This method allows for up to six consecutive generations of spring cereals in a single year. Although far-red light is often overlooked in speed breeding protocols, it serves as a potent inducer of accelerated flowering in various plant species. In this study, we explored the advantages of far-red light as a means to optimize the speed breeding of spring triticale. Experimental plants were cultivated under three conditions with different red to far-red ratios at 660 nm (R – red) and 730 nm (FR – far red): 1) 3.75 (R > FR); 2) 0.8 (R = FR) and 3) 0.3 (R < FR). We found that the onset of triticale flowering occurred significantly earlier at the lowest red to far-red light ratio (R/FR 0.3). On average, plants bloomed 2.6 and 4.1 days earlier in a mineral wool and a soil mixture at R/FR 0.3, respectively, than those grown at R/FR 3.75. A negative effect of higher-intensity far-red light on the reproductive system of triticale was observed. Additionally, seeds obtained from plants grown under higher-intensity far-red light showed significantly lower germination energy and capacity. No differences were found in the regenerative capacity of isolated embryos *in vitro* obtained from plants grown under the different spectral compositions. Our results demonstrate that the accelerated triticale development requires not only the involvement of far-red light, but also a specific red to far-red light ratio close to 0.3. A modified speed breeding protocol relying on this ratio enabled flowering to commence as early as 33.9 ± 1.2 days after sowing. The same triticale variety grown under field conditions in the Krasnodar region and in traditional laboratory growing conditions with a photoperiod of 18/6 h day/night flowered 25 to 29 days later than those cultivated under the speed breeding conditions.

Key words: far-red light; red light; speed breeding; triticale

For citation: Blinkov A.O., Nagamova V.M., Minkova Y.V., Svistunova N.Yu., Radzeniece S., Kocheshkova A.A., Sleptsov N.N., Freymans A.V., Panchenko V.V., Chernook A.G., Karlov G.I., Divashuk M.G. A higher far-red intensity promotes the transition to flowering in triticale grown under speed breeding conditions. *Vavilovskii Zhurnal Genetiki i Selektsii* = *Vavilov J Genet Breed*. 2025;29(6):896-904. doi 10.18699/vjgb-25-96

Funding. The research was funded by the Ministry of Science and Higher Education of the Russian Federation, state task number FGUM-2024-0002.

Увеличение доли дальнего красного света сокращает вегетационный период тритикале в условиях спидбридинга

А.О. Блинков [1] ✉, В.М. Нагамова [1], Я.В. Минькова [1], Н.Ю. Свистунова [1], С. Радзениеце [1], А.А. Кочешкова [1], Н.Н. Слепцов [1, 2], А.В. Фрейманс[3], В.В. Панченко[4], А.Г. Черноок [1], Г.И. Карлов [1], М.Г. Дивашук [1]

[1] Всероссийский научно-исследовательский институт сельскохозяйственной биотехнологии, Москва, Россия
[2] Российский государственный аграрный университет – МСХА им. К.А. Тимирязева, Москва, Россия
[3] ООО «Климбиотех», Москва, Россия
[4] Национальный центр зерна им. П.П. Лукьяненко, Краснодар, Россия
✉ aoblinkov@gmail.com

Аннотация. Работа по созданию нового перспективного сорта занимает в среднем 12–15 лет. Одним из возможных решений проблемы сокращения длительности селекционного процесса становится технология спидбридинг (speed breeding). Метод, направленный на сокращение вегетационного периода, позволяет получать до шести последовательных поколений яровых злаков за один год. К сожалению, в протоколах спидбридинга уделено мало внимания дальнему красному свету – широко известному индуктору быстрого перехода к цветению. В нашей работе мы оценили возможность использования дальнего красного света для оптимизации спидбридинга яровой

тритикале. Экспериментальные растения выращивали в трех вариантах освещения, различающихся соотношением уровней излучения в области 660 нм (К – красный) и 730 нм (ДК – дальний красный): 1) К/ДК 3.75 (К > ДК); 2) К/ДК 0.8 (К = ДК) и 3) К/ДК 0.3 (К < ДК). В результате установлено, что начало цветения тритикале наступало значительно раньше при самом низком соотношении красного к дальнему красному свету (К/ДК 0.3). В среднем при К/ДК 0.3 растения, вегетирующие на минеральной вате и почвенной смеси, зацветали соответственно на 2.6 и 4.1 суток быстрее, чем при варианте К/ДК 3.75. Статистически значимой разницы по продолжительности периода от посева до цветения между вариантами К/ДК 3.75 и К/ДК 0.8 не выявлено. Показано негативное влияние увеличенной доли дальнего красного света на репродуктивную систему тритикале. У семян, сформировавшихся при К/ДК 0.3, наблюдалась значительно меньшая энергия прорастания и всхожесть. Различий в регенерационных способностях изолированных *in vitro* зародышей, полученных от тритикале, выросшей под светом с разным спектральным составом, не обнаружено. Полученные нами результаты демонстрируют, что для сокращения времени от посева до цветения тритикале важно не только наличие дальнего красного света, но и его соотношение с красным, а именно использование состава, близкого к соотношению К/ДК 0.3. Модифицированный по спектральному составу света протокол спидбридинга позволил инициировать цветение уже на 33.9 ± 1.2 сутки с момента посева. Аналогичный сорт тритикале в полевых условиях Краснодарского края и классических лабораторных условиях выращивания с фотопериодом 18/6 ч день/ночь зацветал на 25–29 суток позже, чем в условиях спидбридинга.
Ключевые слова: дальний красный свет; красный свет; спидбридинг; тритикале

Introduction

Breeders and geneticists have always sought to obtain homozygous cereal lines with specified traits more rapidly, which has led to the adoption of approaches such as shuttle breeding (Mergoum et al., 2009), the production of doubled haploids (Timonova et al., 2022), the use of embryo culture (Liu et al., 2016) and molecular markers (Fedyaeva et al., 2023). However, these methods are not always accessible to specific laboratories or breeding centers, may require highly qualified personnel, and some of them do not result in the desired reduction in the time required to develop pure lines.

In recent years, speed breeding – a method based on reducing the generation time of plants to approximately two months – has been gaining popularity (Ghosh et al., 2018; Watson et al., 2018). By reducing generation time, speed breeding enables the production of up to six successive generations of spring cereals within 12 months, allowing the development of pure lines in a single year. The essence of speed breeding lies in the utilization of physical factors that reduce the time from sowing to flowering, decrease the duration of the generative stage of development, overcome post-harvest seed dormancy, and thereby minimize the time required to grow one generation. This technology is simple, low-cost, and enables work with genotypes adapted to various natural and climatic zones, enabling it to be actively integrated into diverse breeding and research programs (Hickey et al., 2017; Li et al., 2019; Vikas et al., 2021).

To reduce the time from sowing to flowering in cereals, prolonged photoperiod, a spectral composition of light including the visible light radiation range of 400–700 nm, light intensity of 450–500 μmol/($m^2 \cdot s$) (Watson et al., 2018), root restriction (Zheng et al., 2023), strict temperature control (Ficht et al., 2023), elevated CO_2 concentrations, and removal of tillering shoots (Tanaka et al., 2016) are employed. To shorten the maturation period, forced drying of immature seeds followed by overcoming their post-harvest dormancy (Marenkova et al., 2024) or embryo culture (Zheng et al., 2023) is used. However, there are a number of parameters, the role of which in reducing the generation time of plants is not entirely clear. One of them is the presence of far-red light during the growing period.

Far-red (FR) light (730 nm) is considered a strong inducer of photomorphogenesis and, depending on its ratio to red (R) light (660 nm), differentially affects seed germination, stem elongation, leaf blade growth, tillering, and the reduction of time from sowing to flowering (Rajcan et al., 2004; Ugarte et al., 2010; Kegge et al., 2015; Demotes-Mainard et al., 2016). Light radiation at these wavelengths and their ratio to each other (commonly described as R/FR) serve as a specific signal for plants, which is perceived by the family of phytochrome photoreceptors. In monocots, phytochromes are represented by three receptors: *PhyA*, *PhyB*, and *PhyC* (Demotes-Mainard et al., 2016; Kippes et al., 2020). Far-red (FR) light can exist in lower (R/FR > 1), higher (R/FR < 1), or equal (R/FR = 1) ratios relative to red light. Daylight contains approximately equal proportions of red and far-red light (1.0–1.3). This ratio decreases to around 0.6 during sunrises and sunsets. A low red-to-far-red light ratio is also observed under leaf and forest canopies, which is due to the active absorption of red light by photosynthetic pigments and the reflection of far-red light from leaves. In such cases, a low R/FR ratio serves as an indicator of the proximity of competing neighbors and triggers the shade avoidance syndrome. This syndrome manifests as enhanced elongation growth, reorientation of leaves toward regions of unattenuated daylight, and accelerated flowering, thereby improving plant survival (Demotes-Mainard et al., 2016; Smith, 2000).

In laboratory conditions, the greatest influence on the growth and development of cereals is exerted by a ratio where far-red light predominates over red light (R/FR < 1). Under light with such a spectral composition, a significant reduction in the time from sowing to flowering and a decrease in tillering shoot growth are observed (Davis, Simmons, 1994; Ugarte et al., 2010; Toyota et al., 2014; Lei et al., 2022). However, despite a number of positive opportunities for speed breeding that far-red light may provide, increasing its amount in the spectral composition of light contributes to a decrease in fertile flowers and grain number per spike (Ugarte et al., 2010; Dreccer et al., 2022).

In the protocols of speed breeding for cereal crops, the utilization of far-red light has received limited attention: in

the graphs of the spectral composition of light presented in research studies, one can observe both its complete absence (Watson et al., 2018; Ficht et al., 2023) and various ratios to red light, in which the latter strongly predominates (Ghosh et al., 2018; Watson et al., 2018; Cha et al., 2022). Only in a small number of studies has far-red light been incorporated in an equal ratio with red light (Zakieh et al., 2021).

There are a number of publications on the influence of far-red light on wheat (Toyota et al., 2014; Dreccer et al., 2022; Lei et al., 2024), barley (Deitzer et al., 1979; Davis, Simmons, 1994; Kegge et al., 2015), and other cereal crops (Rajcan et al., 2004; Markham et al., 2010; Huber et al., 2024). Regarding triticale, little attention has been paid to this topic, and practically no similar studies have been conducted for this crop (Kalituho et al., 1997). A similar situation exists with speed breeding studies for this crop: in open access, only a few studies can be found for spring (Cha et al., 2021) and winter (Zheng et al., 2023) triticale. Therefore, the objectives of this work are to evaluate the influence of far-red light and its ratio to red light under speed breeding conditions on the time from sowing to flowering, main agronomic traits, and reproductive system of triticale.

Materials and methods

Plant material and growing conditions. The object of the study was the spring triticale (× *Triticosecale* Wittm.) variety Dublet (Danko Hodowla Roślin, Poland). Dublet is one of the earliest-ripening among spring triticale varieties (Losert et al., 2016), so the obtained data can be used as an indicator of the minimum generation cycle duration under speed breeding conditions in triticale. As a doubled haploid (Arseniuk, 2019), the Dublet variety exhibits high uniformity in both the onset of developmental phases and morphological traits. Moreover, this variety is widely distributed in Europe (Lekontzeva et al., 2019; Faccini et al., 2023; Radivon, Zhukovsky, 2023) and is known to every specialist working with this crop.

The seeds treated with the fungicide Maxim (Syngenta, France) were preliminarily germinated on water-moistened filter paper in darkness at a temperature of +25 °C. After twenty-four hours, only sprouted seeds were transferred to the substrate. Trays with 110 mL cell volumes were used for cultivation. Two substrate variants were employed: 1) a soil mixture consisting of peat, chernozem, sand, and vermiculite in a ratio of 5:3:1:1 (50 g of moistened mixture per tray cell); 2) mineral wool cubes measuring 50 × 45 × 45 mm (one cube per tray cell). One sprouted seed was placed in each tray cell at 1 cm depth. The growth chamber was maintained at a constant temperature of +25–26 °C and an air humidity of 35–45 %.

For the first two weeks, plants in the soil mixture were watered as needed, and fertilization was performed once a week with Tripart fertilizer (General Hydroponics Europe, France) according to the manufacturer's instructions. Two weeks after sowing, the plants were transferred to watering with fertilizer three times a week. Mineral wool cubes were irrigated with fertilizer daily. Foliar feeding with Siliplant (Nest-M, Russia) was performed once a week in accordance with the manufacturer's recommendations. Treatments for diseases and pests were carried out as necessary. Tiller removal was performed during the plants growth. The photoperiod was maintained at 22/2 hours day/night according to (Watson et al., 2018). Adjustable multichromatic PWM-dimmable LED lamps (PrometheusVNIISB by Gorshkoff, Russia) were used as light sources (chip emitters: 460, 660, 735 nm, white4000K (EPIstar, China); multi-channel pulse-width modulation controller (BKD, Russia); total power of 800 watts).

As control conditions, triticale was cultivated in a Fitotron SGC 120 climatic chamber (Weiss Technik, Netherlands) with fluorescent lamps under a photoperiod of 18/6 h day/night, a light intensity of 285 μmol/($m^2 \cdot s$) at shelf level, a temperature of +22 °C, and air humidity of 65 % round the clock. Sowing and plant care were similar to those described above. As an additional control, data from long-term field trials of the P.P. Lukyanenko National Grain Center (Krasnodar region, Russia) were used. Agronomic practices and sowing dates were conventional for the region.

Influence of far-red light on the triticale growth stages and main agronomic traits. The degree of influence of far-red light on triticale was determined by growing plants under three lighting conditions differing in the ratio of radiation levels in the 660 nm region (R – red) and 730 nm (FR – far-red): 1) R/FR ratio = 3.75 (hereinafter referred to as R > FR) (Fig. 1*a*); 2) R/FR ratio = 0.8 (hereinafter referred to as R = FR) (Fig. 1*b*); 3) R/FR ratio = 0.3 (hereinafter referred to as R < FR) (Fig. 1*c*).

The light intensity in all variants was set to 330 μmol/($m^2 \cdot s$) at shelf level. Far-red light was introduced one week after seed germination. Lighting parameters were adjusted and verified using a PG200N spectrometer (United Power Research Technology Corp., Taiwan).

The onset of growing stages was assessed individually for each plant according to (Zadoks et al., 1974). The onset of the heading stage was defined as the day when the spike fully emerged from the flag leaf sheath (phase Z5.9). The onset of the flowering stage was defined as the day when the first anthers appeared on the spikes (phase Z6.1).

To evaluate the influence of far-red light on triticale, an analysis of the main agronomic traits of all experimental plants was conducted based on the following parameters: plant height (cm), spike length (cm), vegetative weight of the dried spike and culm (g), number of spikelets (pcs.) and grains (pcs.) per spike, number of grains per spikelet (pcs.), and weight of 1,000 grains (g).

Effect of far-red light on seed viability. The evaluation of the influence of far-red light on seed viability indicators was conducted using two methods: 1) by culturing immature embryos; and 2) by germinating seeds on filter paper. In the first method, embryo isolation was performed on the 15th day after flowering. Caryopses were sterilized in a 50 % solution of the commercial agent "Belizna", followed by three washes with sterile distilled water. Embryo isolation was carried out under an Olympus SZ61 stereoscopic microscope (Olympus, Japan). Cultivation was performed in Petri dishes containing agar-solidified Murashige and Skoog medium (Murashige, Skoog, 1962). The cultivation lasted for 10 days under a photoperiod of 22/2 h day/night, a light intensity of 80 μmol/($m^2 \cdot s$), and a temperature of +24 °C.

In the second method, starting from the 17th day after flowering, the amount of watering was gradually reduced until it was completely discontinued on the day of spike cutting, which occurred on the 20th day after flowering. The cut spikes

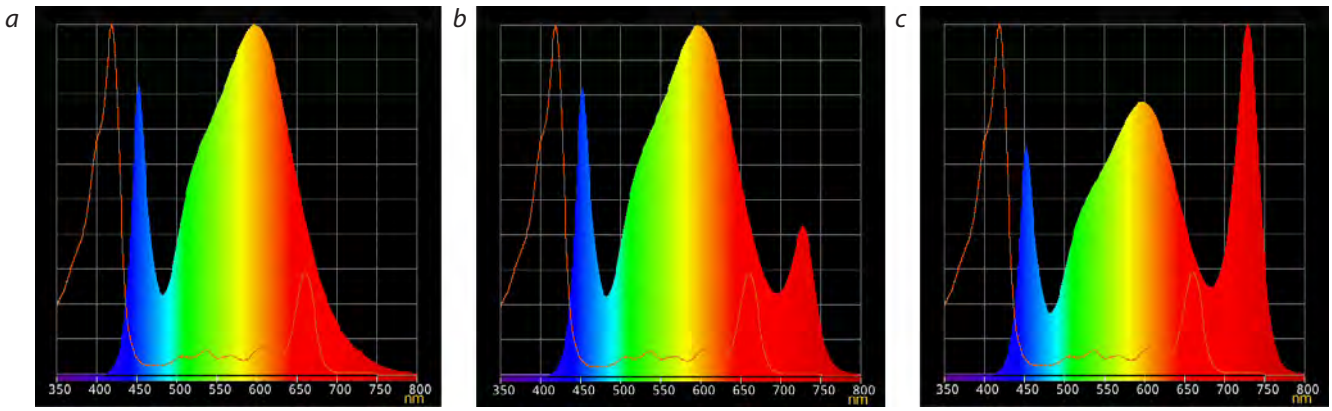


Fig. 1. Spectral composition of light used in the experiment: *a* – R > FR, R/FR ratio = 3.75; *b* – R = FR, R/FR ratio = 0.8; *c* – R < FR, R/FR ratio = 0.3.

were placed in paper bags, which were subjected to forced drying at a temperature of +28 °C for 7–10 days, depending on the drying rate. After drying, the spikes were threshed, and the seeds were stored in paper bags at room temperature for one week. Next, the seeds were placed in Petri dishes on filter paper moistened with a 0.5 mg/L solution of gibberellic acid (Sigma-Aldrich, USA) and incubated under cold pretreatment conditions (+4 °C, darkness, three days), followed by germination in darkness at +25 °C. Germination energy was assessed on the third day, and germination capacity was evaluated on the seventh day after placing the Petri dishes with seeds at +25 °C.

Statistical analysis. To evaluate the degree of influence of the spectral composition of light on the vegetative period of triticale, a twofold replication was used for each variant, with 10 plants in each replication. In total, 120 plants were analyzed. The number of days from sowing to flowering of each individual plant was assessed. Under field conditions, the number of days from sowing to mass flowering was evaluated.

To assess the regeneration capacity and viability of isolated embryos, a fourfold replication was employed, with 10 isolated embryos in each replication. The evaluation of germination energy and seed germination capacity was performed in fourfold replication. Each replication contained 50 seeds.

Statistical processing was performed using the R programming language (version 4.3.2). The influence of the spectral composition of light on various indicators of triticale plants was assessed using one-factor analysis of variance (ANOVA), followed by multiple comparisons of mean values using Tukey's test to determine significant differences between plant groups.

Results

Effect of far-red light on triticale growth stages

As a result of the conducted experiment, the one-factor analysis of variance revealed a statistically significant reduction in the time from sowing to the onset of flowering in plants grown under the light with the spectral composition of R/FR = 0.3 compared to other lighting variants ($p < 0.05$). This trend was observed in both substrate variants. Plants under the light with the spectral composition of R/FR = 0.3 flowered on average 2.6 and 4.1 days faster when using mineral wool and soil mixture, respectively, than those under the light with the spectral composition of R/FR = 3.75. No statistically significant difference in the duration of the period from sowing to flowering was found between the R/FR = 3.75 and R/FR = 0.8 variants ($p > 0.05$) (Table 1).

Table 1. Mean values ± 95 % confidence interval for heading and flowering dates in plants of the Dublet variety grown under three lighting conditions with different spectral compositions

Substrate	Spectral composition (R/FR)	Duration from sowing to heading, days	Duration from sowing to flowering, days
Mineral wool	3.75	33.5 ± 1.0a[1]	37.3 ± 1.1a
	0.8	34.0 ± 1.6a	37.3 ± 1.9a
	0.3	**31.1 ± 0.8b**	**34.7 ± 0.9b**
Soil mixture	3.75	34.1 ± 0.9a	38.0 ± 1.0a
	0.8	33.7 ± 1.1a	36.9 ± 1.2a
	0.3	**30.5 ± 1.2b**	**33.9 ± 1.2b**

Note. [1] Values followed by the same letter do not differ significantly ($p > 0.05$) according to Tukey's test. Bold type indicates values that are significantly different from the other variants ($p < 0.05$).

Table 2. Mean values ± 95 % confidence interval for the main agronomic traits of the Dublet variety grown under three lighting conditions with different spectral compositions

Traits	Spectral composition (R/FR)		
	3.75	0.8	0.3
Mineral wool			
Plant height, cm	57.9 ± 1.8a[1]	60.5 ± 2.2a	59.1 ± 1.6a
Culm dry weight, g	0.34 ± 0.08a	0.37 ± 0.04a	0.36 ± 0.02a
Spike length, cm	5.5 ± 0.3ab	5.7 ± 0.4b	**5.2 ± 0.1a**
Spike vegetative weight, g	1.24 ± 0.15ab	1.40 ± 0.18b	**1.15 ± 0.13a**
1,000-grain weight, g	32.6 ± 2.0a	36.3 ± 4.9a	**41.9 ± 2.0b**
Number of grains per spike, pcs.	30.1 ± 4.1ab	32.1 ± 4.3b	**24.7 ± 2.3a**
Number of spikelets per spike, pcs.	14.2 ± 1.2a	13.7 ± 0.7a	**12.2 ± 0.3b**
Number of grains per spikelet, pcs.	2.1 ± 0.2a	2.3 ± 0.2a	2.0 ± 0.2a
Soil mixture			
Plant height, cm	60.8 ± 3.0a	60.5 ± 1.6a	58.4 ± 2.2a
Culm dry weight, g	0.41 ± 0.03a	0.46 ± 0.03a	0.44 ± 0.03a
Spike length, cm	6.1 ± 0.3a	6.1 ± 0.3a	**5.6 ± 0.2b**
Spike vegetative weight, g	1.44 ± 0.11ab	1.52 ± 0.08b	**1.34 ± 0.08a**
1,000-grain weight, g	35.7 ± 2.9a	**40.4 ± 1.7b**	**41.4 ± 2.3b**
Number of grains per spike, pcs.	33.1 ± 2.6a	30.6 ± 2.1a	**26.1 ± 2.6b**
Number of spikelets per spike, pcs.	15.8 ± 1.1a	14.8 ± 0.7ab	**13.9 ± 0.5b**
Number of grains per spikelet, pcs.	2.1 ± 0.1a	2.1 ± 0.1a	**1.9 ± 0.1b**

Note. [1] Values followed by the same letter do not differ significantly ($p > 0.05$) according to Tukey's test. Bold type indicates values that are significantly different from the other variants ($p < 0.05$).

Effect of far-red light on main agronomic traits of triticale

No significant differences in vegetative weight or straw height were observed among triticale plants grown under light with different spectral compositions ($p > 0.05$). A significant influence of the high amount of far-red light on spike productivity was evident (Table 2). When triticale was cultivated under R/FR = 0.3, plants on both substrate variants formed shorter spikes with fewer spikelets, leading to a reduction in the vegetative weight of the spike and the number of grains per spike ($p < 0.05$). An increased amount of far-red light resulted in fewer grains per spikelet, but only in plants grown on the soil mixture. Despite this, a statistically significant increase in 1,000-grain weight was observed in plants grown under R/FR = 0.3 on both substrate variants ($p < 0.05$). In the majority of cases, no statistically significant difference was detected between the R/FR = 3.75 and R/FR = 0.8 variants in terms of productivity indicators.

Effect of far-red light on seed viability and germination

A statistically significant decrease in germination energy and capacity was observed in seeds obtained from plants grown under an increased amount of far-red light ($p < 0.05$). In isolated embryos *in vitro* derived from plants cultivated under light with different spectral compositions, no statistically significant differences in regeneration frequency were detected ($p > 0.05$) (Table 3). Already on the third day after the start of cultivation, regardless of the lighting conditions of the donor plants, the embryos developed coleoptiles and roots, and by the tenth day of cultivation, the embryos exhibited one fully formed leaf and a well-developed root system.

Control plants growing

Plants under all control conditions exhibited a prolonged germination–flowering period. According to long-term cultivation data in the Krasnodar region, the anthesis of triticale variety Dublet occurred on days 60–64 when sown in early March and on days 50–52 when sown in early April. Under the conditions of a climate-controlled chamber with a photoperiod of 18/6 h day/night, triticale reached the flowering stage 62.5 ± 2.0 and 59.2 ± 2.6 days after sowing on mineral wool and on soil mixture, respectively.

Discussion

Speed breeding has demonstrated its popularity across various fields of genetics, breeding, and biotechnology (Ghosh et al., 2018). Modifications of established protocols are being implemented, including simplifying their organization, transi-

Table 3. Germination energy and capacity, as well as regeneration frequency of isolated embryos obtained from plants grown under three lighting conditions with different spectral compositions

Substrate	Spectral composition (R/FR)	Germination energy, %	Germination capacity, %	Regeneration of isolated embryos, %
Mineral wool	3.75	97.1 ± 4.0a[1]	96.5 ± 4.0a	97.5 ± 3.7a
	0.8	85.5 ± 8.5a	96.8 ± 2.5a	100 ± 0a
	0.3	**44.1 ± 21.9b**	**77.7 ± 12.7b**	100 ± 0a
Soil mixture	3.75	93.3 ± 4.5a	98.9 ± 1.6a	100 ± 0a
	0.8	**51.3 ± 15.8b**	**81.4 ± 11.5b**	100 ± 0a
	0.3	**68.5 ± 17.3b**	89.0 ± 10.8ab	100 ± 0a

Note. [1]Values followed by the same letter do not differ significantly ($p > 0.05$) according to Tukey's test. Bold type indicates values that are significantly different from the other variants ($p < 0.05$).

tioning to high-throughput capacity, incorporating molecular genetics methods, and integrating them into the breeding process (Kigoni et al., 2023; Marenkova et al., 2024). At present, speed breeding protocols have been successfully tested in numerous cereal species (Watson et al., 2018; Cha et al., 2021). Despite active work in this area, most published studies on cereal speed breeding have not adequately addressed one of the strongest inducers of shortening the sowing-to-flowering period – far-red light. However, under speed breeding conditions, its efficacy has been demonstrated for crops such as rapeseed (Song et al., 2022), amaranth (Jähne et al., 2020), and pepper (Choi et al., 2023).

The shortening of the vegetative period is one of the primary manifestations of shade avoidance syndrome, initiated by far-red light in the photoperiodic regulation of flowering. Light with an increased amount of far-red light is perceived by leaves and activates phytochrome photoreceptors, primarily *PhyA* and *PhyB*. Phytochromes trigger the expression of the central flowering regulator gene *CONSTANT* (*CO*), which, in turn, induces *FLOWERING LOCUS T* (*FT*) – the florigen in the vascular bundles of leaves. The FT protein moves from the leaves to the shoot apical meristem and, together with the FD protein (product of the *FLOWERING LOCUS D* gene (*FD*)), initiates the activity of genes such as *SUPPRESSOR OF OVEREXPRESSION OF CO1* (*SOC1*) and *APETALA1* (*AP1*), which determine the development of floral meristems (Demotes-Mainard et al., 2016; Sheerin, Hiltbrunner, 2017; Lebedeva et al., 2020).

To evaluate the effect of far-red light under speed breeding conditions, we conducted an experiment involving the cultivation of spring triticale on two types of substrates and under three lighting variants differing in spectral composition, characterized by varying ratios of red to far-red light.

The experiments conducted by us demonstrated a significant influence of an increased amount of far-red light (R < FR, R/FR = 0.3) on the onset of the flowering phase in triticale. Plants exposed to the light with the spectral composition of R/FR = 0.3 flowered 33.9 ± 1.2 and 34.7 ± 0.9 days after sowing when grown on soil mixture and mineral wool cubes, respectively, which is 4.1 and 2.6 days faster than under the light with R/FR = 3.75. No statistically significant difference in the duration of the vegetative period of triticale was observed under light spectra with R > FR and R = FR. The obtained results indicate that in order to shorten the sowing-to-flowering period in triticale, not only the presence of far-red light but also its ratio to red light is critical, specifically the use of a composition close to R/FR = 0.3 (Fig. 2). Our findings align with those of several other studies reporting similar results in cereals (Deitzer et al., 1979; Davis, Simmons, 1994; Toyota et al., 2014).

Despite the difference in flowering onset timing, amounting to 2.6 and 4.1 days, far-red light can be considered a valuable addition for creating conditions aimed at shortening the vegetative period of plants. This is because if the sowing–flowering period can be reduced by 3–4 days in one generation, the cumulative effect when sequentially growing six generations (a number that typically facilitates the production of a pure line) could reach up to 20 days.

Currently, only a limited number of studies are dedicated to speed breeding of spring and winter triticale, demonstrating the high responsiveness of this crop to factors influencing the shortening of the vegetative period (Cha et al., 2021; Zheng et al., 2023). It has been shown that for spring triticale, the average time from sowing to heading ranges between 33–42 days depending on the genotype (Cha et al., 2021), whereas spring bread wheat under speed breeding conditions flowers, depending on the genotype, between 35.7 and 75 days after sowing (Ghosh et al., 2018; Watson et al., 2018; Cha et al., 2020). Our work confirms the significant impact of the speed breeding method on reducing the vegetative period in triticale: the spectrally modified protocol enabled the initiation of flowering as early as 33.9 ± 1.2 days after sowing. The same triticale cultivar under field conditions in the Krasnodar region and classical laboratory cultivation conditions flowered 25–29 days later than under speed breeding conditions.

The results of the evaluation of triticale yield structure did not reveal significant changes in the height of the studied plants when grown under the light with a spectral composition containing an increased amount of far-red light, although numerous studies report stem elongation in cereals under far-red light (Kegge et al., 2015; Lei et al., 2022). Shade avoidance syndrome, which causes shoot elongation, has also been absent in other studies where far-red light was used as a supplement to shorter wavelengths (400–680 nm) (Huber et al., 2024). This suggests that the use of increased amounts of far-red light in the optical radiation spectrum under triticale speed breeding conditions does not lead to such an inconvenient factor in practice as the formation of tall plants and their lodging.

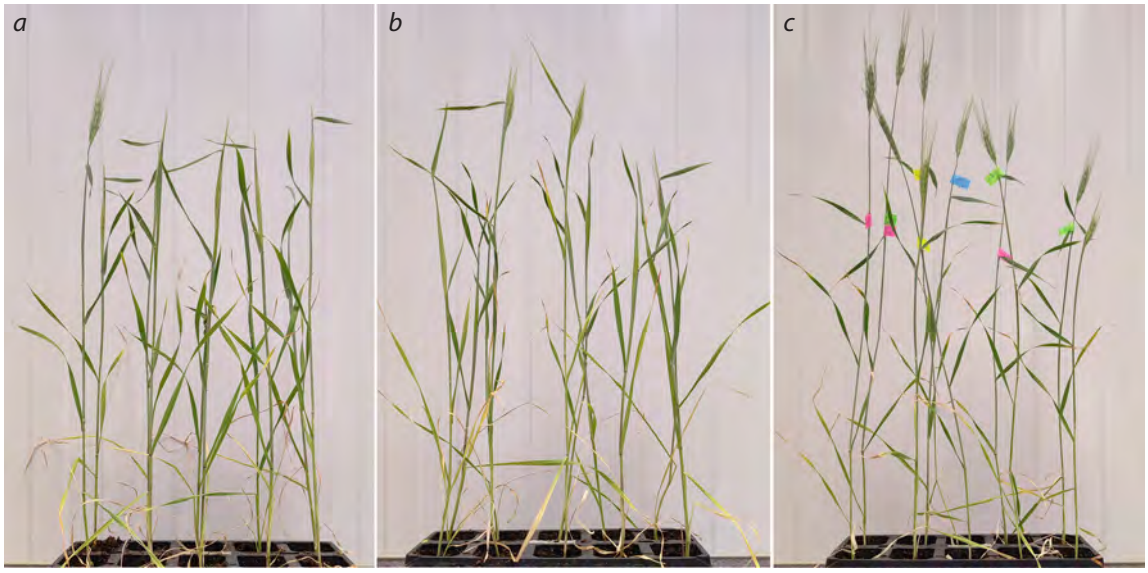


Fig. 2. Plants on the 30th day after sowing (all plants were sown on the same day), cultivated under three lighting conditions with different spectral compositions, and at the stage of: *a*, *b* – the beginning of heading, R/FR = 3.75 (*a*), R/FR = 0.8 (*b*); *c* – full heading, R/FR = 0.3.

The data obtained by us demonstrated a strong influence of far-red light on the productivity of triticale spikes. In plants of the Dublet variety grown under the light with the spectral composition of R/FR = 0.3, a shorter spike with fewer spikelets was formed, leading to a significant reduction in the vegetative mass of the spike and the number of grains per spike. Similar results have been reported in bread wheat, where an increased amount of far-red light reduces the number of fertile flowers and the number of grains per spike (Ugarte et al., 2010; Dreccer et al., 2022), which is likely associated with the inhibitory effect of far-red light on plant nitrogen assimilation (Lei et al., 2024).

Despite the negative influence of far-red light on spike productivity components, the 1,000-grain weight of plants grown under an increased amount of far-red light in the spectral compositions was significantly higher. This may be associated with the Emerson effect, which involves enhanced photosynthetic efficiency when far-red light is used in combination with shorter wavelengths (400–680 nm) (Huber et al., 2024).

Additionally, during our study, a statistically significant negative impact of far-red light on germination energy and capacity was detected. The germination capacity of seeds obtained from plants grown under the light with the spectral composition of R/FR = 0.3 ranged from 77.7 ± 12.7 to 89.0 ± 10.8 % depending on the growth substrate. At the same time, the regeneration frequency of isolated embryos *in vitro* was equally high for all seeds, regardless of the lighting conditions under which the donor plants were cultivated. Given that in cereals, the speed breeding system is compatible with the single-seed descent method (Alahmad et al., 2018; Watson et al., 2018), where one seed per spike is selected for each subsequent generation to preserve genetic diversity and prevent the expansion of cultivation areas, the use of high amounts of far-red light in the spectral compositions will not become a limiting factor when cultivating plants using this method. However, it should be noted that when the primary goal of plant cultivation is propagation and obtaining seeds with good germination capacity, it is necessary to reduce the amounts of far-red light in the optical radiation spectrum to a level where R/FR > 1.

Conclusion

Our study demonstrated that under speed breeding conditions, the use of the highest amount of far-red light in the spectral composition (R/FR = 0.3), compared to the spectrum where the R/FR ratio is 3.75, resulted in a statistically significant reduction in time from sowing to flowering by 2.6 and 4.1 days for plants grown in mineral wool and soil mixture, respectively. No statistically significant difference in the duration from sowing to flowering was detected between the R/FR = 3.75 and R/FR = 0.8 variants. The speed breeding protocol with a modified light spectrum induced flowering as early as 33.9 ± 1.2 days after sowing. The same triticale variety flowered 25–29 days later under field conditions in the Krasnodar region and conventional laboratory cultivation with a photoperiod of 18/6 h day/night compared to modified speed breeding conditions. No statistically significant increase in plant height was observed when using the highest amount of far-red light in the spectral composition. A negative influence of far-red light on spike parameters (length, vegetative weight, number of spikelets and grains per spike) as well as germination energy and capacity was detected. It can be reasonably assumed that increasing the amount of far-red light in the optical radiation spectrum (R/FR = 0.3) could serve as a beneficial addition to speed breeding conditions not only for triticale but also for other cereals.

References

Alahmad S., Dinglasan E., Leung K.M., Riaz A., Derbal N., Voss-Fels K.P., Able C.J., Bassi F.M., Christopher C.J., Hickey L.T. Speed breeding for multiple quantitative traits in durum wheat. *Plant Methods*. 2018;14:36. doi 10.1186/s13007-018-0302-y

Arseniuk E. Recent developments in triticale breeding research and production – an Overview. *Ekin J Crop Breed Genet*. 2019;5(2):68-73

Cha J.K., Lee J.H., Lee S.M., Ko J.M., Shin D.J. Heading date and growth character of Korean wheat cultivars by controlling photoperiod for rapid generation advancement. *Korean J Breed Sci.* 2020; 52(1):20-24. doi 10.9787/KJBS.2020.52.1.20

Cha J.K., Park M.R., Shin D., Kwon Y., Lee S.M., Ko J.M., Kim K.M., Lee J.H. Growth characteristics of triticale under long-day photoperiod for rapid generation advancement. *Korean J Breed Sci.* 2021; 53(3):200-205. doi 10.9787/KJBS.2021.53.3.200

Cha J.K., O'Connor K., Alahmad S., Lee J.H., Dinglasan E., Park H., Lee S.M., … Kim K.M., Ko J.M., Hickey L.T., Shin D., Dixon L.E. Speed vernalization to accelerate generation advance in winter cereal crops. *Mol Plant.* 2022;15(8):1300-1309. doi 10.1016/j.molp.2022.06.012

Choi H., Back S., Kim G.W., Lee K., Venkatesh J., Lee H.B., Kwon J.K., Kang B.C. Development of a speed breeding protocol with flowering gene investigation in pepper (*Capsicum annuum*). *Front Plant Sci.* 2023;14:1151765. doi 10.3389/fpls.2023.1151765

Davis M.H., Simmons S.R. Far-red light reflected from neighbouring vegetation promotes shoot elongation and accelerates flowering in spring barley plants. *Plant Cell Environ.* 1994;17(7):829-836. doi 10.1111/j.1365-3040.1994.tb00177.x

Deitzer G.F., Hayes R., Jabben M. Kinetics and time dependence of the effect of far-red light on the photoperiodic induction of flowering in Wintex barley. *Plant Physiol.* 1979;64(6):1015-1021. doi 10.1104/pp.64.6.1015

Demotes-Mainard S., Peron T., Corot A., Bertheloot J., Gourrierec J.L., Pelleschi-Travier S., Crespel L., Morel P., Huche-Thelier L., Boumaza R., Vian A., Guerin V., Leduc N., Sakr S. Plant responses to red and far-red lights, applications in horticulture. *Envir Exp Bot.* 2016; 121:4-21. doi 10.1016/j.envexpbot.2015.05.010

Dreccer M.F., Zwart A.B., Schmidt R.C., Condon A.G., Awasi M.A., Grant T.J., Galle A., Bourot S., Frohberg C. Wheat yield potential can be maximized by increasing red to far-red light conditions at critical developmental stages. *Plant Cell Environ.* 2022;45(9):2652-2670. doi 10.1111/pce.14390

Faccini N., Morcia C., Terzi V., Rizza F., Badeck F.W. Triticale in Italy. *Biology*. 2023;2(10):1308. doi 10.3390/biology12101308

Fedyaeva A.V., Salina E.A., Shumny V.K. Pre-harvest sprouting in soft winter wheat (*Triticum aestivum* L.) and evaluation methods. *Russ J Genet*. 2023;59(1):1-11. doi 10.1134/S1022795423010052

Ficht A., Bruch A., Rajcan I., Pozniak C., Lyons E.M. Evaluation of the impact of photoperiod and light intensity on decreasing days to maturity in winter wheat. *Crop Sci.* 2023;63(2):812-821. doi 10.1002/csc2.20886

Ghosh S., Watson A., Gonzalez-Navarro O.E., Ramirez-Gonzalez R.H., Yanes L., Mendoza-Suárez M., Simmonds J., … Domoney C., Uauy C., Hazard B., Wulff B.B.H., Hickey L.T. Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. *Nat Protoc.* 2018;13(12):2944-2963. doi 10.1038/s41596-018-0072-z

Hickey L.T., German S.E., Pereyra S.A., Diaz J.E., Ziems L.A., Fowler R.A., Platz G.J., Franckowiak J.D., Dieters M.J. Speed breeding for multiple disease resistance in barley. *Euphytica.* 2017;213:64. doi 10.1007/s10681-016-1803-2

Huber M., de Boer H.J., Romanowski A., van Veen H., Buti S., Kahlon P.S., van der Meijden J., Koch J., Pierik R. Far-red light enrichment affects gene expression and architecture as well as growth and photosynthesis in rice. *Plant Cell Environ.* 2024;47(8):2936-2953. doi 10.1111/pce.14909

Jähne F., Hahn V., Würschum T., Leiser W.L. Speed breeding short-day crops by LED-controlled light schemes. *Theor Appl Genet.* 2020; 133(8):2335-2342. doi 10.1007/s00122-020-03601-4

Kalituho L.N., Chaika M.T., Kabashnikova L.F., Makarov V.N., Khirpach V.A. On the phytochrome mediated action of brassinosteroids. *Proc Plant Growth Regul Soc Am.* 1997;24:140-145

Kegge W., Ninkovic V., Glinwood R., Welschen R.A.M., Voesenek L.A.C.J., Pierik R. Red: far-red light conditions affect the emission of volatile organic compounds from barley (*Hordeum vulgare*), leading to altered biomass allocation in neighbouring plants. *Ann Bot.* 2015;115(6):961-970. doi 10.1093/aob/mcv036

Kigoni M., Choi M., Arbelaez J.D. 'Single-Seed-SpeedBulks': a protocol that combines 'speed breeding' with a cost-efficient modified single-seed descent method for rapid-generation-advancement in oat (*Avena sativa* L.). *Plant Methods.* 2023;19(1):92. doi 10.1186/s13007-023-01067-1

Kippes N., VanGessel C., Hamilton J., Akpinar A., Budak H., Dubcovsky J., Pearce S. Effect of *phyB* and *phyC* loss-of-function mutations on the wheat transcriptome under short and long day photoperiods. *BMC Plant Biol.* 2020;20:297. doi 10.1186/s12870-020-02506-0

Lebedeva M.A., Dodueva I.E., Gancheva M.S., Tvorogova V.E., Kuznetsova K.A., Lutova L.A. The evolutionary aspects of flowering control: Florigens and Anti-florigens. *Russ J Genet*. 2020;56(11): 1323-1344. doi 10.1134/S102279542011006X

Lei K., Tan Q., Zhu L., Xu L., Yang S., Hu J., Gao L., Hou P., Shao Yu., Jiang D., Cao W., Dai T., Tian Z. Low red/far-red ratio can induce cytokinin degradation resulting in the inhibition of tillering in wheat (*Triticum aestivum* L.). *Front Plant Sci.* 2022;13:971003. doi 10.3389/fpls.2022.971003

Lei K., Hu H., Chang M., Sun C., Ullah A., Yu J., Dong C., Gao Q., Jiang D., Cao W., Tian Z., Dai T. A low red/far-red ratio restricts nitrogen assimilation by inhibiting nitrate reductase associated with downregulated *TaNR1.2* and upregulated *TaPIL5* in wheat (*Triticum aestivum* L.). *Plant Physiol Biochem.* 2024;206:107850. doi 10.1016/j.plaphy.2023.107850

Lekontzeva T.A., Yufereva N. I., Statzenko E.S. Assessment of initial material (base line) for creation of spring triticale varieties in the climate of Volgo-Vyatka region. *Dal'nevostochnyy Agrarnyy Vestnik = The Far East Agrarian Herald*. 2019;2(50):45-52. doi 10.24411/1999-6837-2019-2-45-52 (in Russian)

Li H., Zhou Y., Xin W., Wei Y., Zhang J., Guo L. Wheat breeding in northern China: achievements and technical advances. *Crop J*. 2019; 7(6):718-729. doi 10.1016/j.cj.2019.09.003

Liu H., Zwer P., Wang H., Liu C., Lu Z., Wang Y., Yan G. A fast generation cycling system for oat and triticale breeding. *Plant Breed.* 2016;135(5):574-579. doi 10.1111/pbr.12408

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

Marenkova A.G., Blinkov A.O., Radzeniece S., Kocheshkova A.A., Karlov G.I., Divashuk M.G. Testing and modification of the protocol for accelerated growth of malting barley under speed breeding conditions. *Nanobiotechnol Rep*. 2024;19(5):808-814. doi 10.1134/S2635167624601955

Markham M.Y., Stoltenberg D.E. Corn morphology, mass, and grain yield as affected by early-season red: Far-red light environments. *Crop Sci.* 2010;50(1):273-280. doi 10.2135/cropsci2008.10.0614

Mergoum M., Singh P.K., Pena R.J., Lozano-del Río A.J., Cooper K.V., Salmon D.F., Gómez Macpherson H. Triticale: A "New" Crop with Old Challenges. In: Carena M. (Eds). Cereals. Handbook of Plant Breeding. Vol. 3. Springer, 2009;267-287. doi 10.1007/978-0-387-72297-9

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

Radivon V.A., Zhukovsky A.G. Analysis of the susceptibility of spring triticale varieties to diseases for 2012–2022. *Zashchita Rasteniy = Plant Protection.* 2023;1(47):128-135 (in Russian)

Rajcan I., Chandler K.J., Swanton C.J. Red-far-red ratio of reflected light: a hypothesis of why early-season weed control is important in corn. *Weed Sci.* 2004;52(5):774-778. doi 10.1614/WS-03-158R

Sheerin D.J., Hiltbrunner A. Molecular mechanisms and ecological function of far-red light signalling. *Plant Cell Environ*. 2017;40(11): 2509-2529. doi 10.1111/pce.12915

Smith H. Phytochromes and light signal perception by plants – an emerging synthesis. *Nature*. 2000;407(6804):585-591. doi 10.1038/35036500

Song Y., Duan X., Wang P., Li X., Yuan X., Wang Z., Li X., Yuan X., Wang Z., Wan L., Yang G., Hong D. Comprehensive speed breeding: a high-throughput and rapid generation system for long-day crops. *Plant Biotechnol J*. 2022;20(1):13-15. doi 10.1111/pbi.13726

Tanaka J., Hayashi T., Iwata H. A practical, rapid generation-advancement system for rice breeding using simplified biotron breeding system. *Breed Sci*. 2016;66(4):542-551. doi 10.1270/jsbbs.15038

Timonova E.M., Adonina I.G., Salina E.A. The influence of combinations of alien translocations on *in vitro* androgenesis in spring common wheat (*Triticum aestivum* L.). *Trudy po Prikladnoy Botanike, Genetike i Selektsii = Proceedings on Applied Botany, Genetics, and Breeding*. 2022;183(1):127-134. doi 10.30901/2227-8834-2022-1-127-134 (in Russian)

Toyota M., Tatewaki N., Morokuma M., Kusutani A. Tillering responses to high red/far-red ratio of four Japanese wheat cultivars. *Plant Prod Sci*. 2014;17(2):124-130. doi 10.1626/pps.17.124

Ugarte C.C., Trupkin S.A., Ghiglione H., Slafer G., Casal J.J. Low red/far-red ratios delay spike and stem growth in wheat. *J Exp Bot*. 2010;61(11):3151-3162. doi 10.1093/jxb/erq140

Vikas V.K., Sivasamy M., Jayaprakash P., Vinod K.K., Geetha M., Nisha R., Peter J. Customized speed breeding as a potential tool to advance generation in wheat. *Indian J Genet Plant Breed*. 2021;81(2):199-207. doi 10.31742/IJGPB.81.2.3

Watson A., Ghosh S., Williams M.J., Cuddy W.S., Simmonds J., Rey M.D., Hatta M.A.M., … Uauy C., Boden S.A., Park R.F., Wulff B.B.H., Hickey L.T. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat Plants*. 2018;4(1):23-29. doi 10.1038/s41477-017-0083-8

Zadoks J.C., Chang T.T., Konzak C.F. A decimal code for the growth stages of cereals. *Weed Res*. 1974;14(6):415-421. doi 10.1111/j.1365-3180.1974.tb01084.x

Zakieh M., Gaikpa D.S., Leiva Sandoval F., Alamrani M., Henriksson T., Odilbekov F., Chawade A. Characterizing winter wheat germplasm for fusarium head blight resistance under accelerated growth conditions. *Front Plant Sci*. 2021;12:705006. doi 10.3389/fpls.2021.705006

Zheng Z., Gao S., Wang H., Liu C. Shortening generation times for winter cereals by vernalizing seedlings from young embryos at 10 degree Celsius. *Plant Breed*. 2023;142(2):202-210. doi 10.1111/pbr.13074

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

Received November 5, 2024. Revised May 13, 2025. Accepted May 14, 2025.

Прием статей через электронную редакцию на сайте http://vavilov.elpub.ru/index.php/jour
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» до 2011 г. выходил под названием «Информационный вестник ВОГиС»/ "The Herald of Vavilov Society for Geneticists and Breeding Scientists".

Сетевое издание «Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» – реестровая запись СМИ Эл № ФС77-85772, зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 14 августа 2023 г.

Издание включено ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Russian Science Citation Index, Российский индекс научного цитирования, ВИНИТИ, Web of Science CC, Scopus, PubMed Central, DOAJ, ROAD, Ulrich's Periodicals Directory, Google Scholar.

Открытый доступ к полным текстам:
русскоязычная версия – на сайте https://vavilovj-icg.ru/
и платформе Научной электронной библиотеки, elibrary.ru/title_about.asp?id=32440

англоязычная версия – на сайте vavilov.elpub.ru/index.php/jour
и платформе PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/journals/3805/

✉ email: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.
Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218.
Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.
Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.

Дата выхода в свет 30.09.2025. Формат 60 × 84 1/8. Уч.-изд. л. 23.9